Assessment of Genetic Diversity for Grain Yield and Quality Traits in Rice (*Oryzasativa*l.) Using Principal Component Analysis

Ch Santhi Priya, Y Suneetha, D Ratna Babu and V Srinivasa Rao

Department of Genetic and Plant Breeding, Agricultural College, Bapatla 522 101, Andhra Pradesh

ABSTRACT

The present experiment was conducted with forty long duration rice genotypes for estimation of genetic diversity by using Principal Component analysis.The first four principal components with eigen values more than one contributed 78.913 per cent towards the total variability. The analysis thus identified the maximum contributing variables *i.e.,* L/B ratio, days to maturity, days to 50% flowering, seed yield per plant, grains per panicle, panicle length, test weight, productive tillers per plant and grain yield per plant. The contribution of the main characters for variance was easily identified by the characters loaded on the PC_1 with high loading values.

Key words*: D^2, Grain Yield, PCA, Rice.*

Rice (*Oryza sativa* L., 2n = 2x= 24) is the principal staple cereal food and source of calories for more than half of the world's population. A large number of variables are often measured by plant breeders, some of which may not be of sufficient discriminatory power for germplasm evaluation, characterization, and management. In such case, Principal Component Analysis (PCA) may be used to reveal patterns and eliminate redundancy in data sets (Adams, 1995; Amy and Pritts, 1991) as morphological and physiological variations routinely occur in crop species. Multivariate statistical techniques which simultaneously analyze multiple measurements on each individual under investigation, are widely used in analysis of genetic diversity irrespective of the nature of parameters. The study was undertaken to run a classificatory analysis on the rice genotypes by means of PCA, and descriptive ANOVA statistic, which would enable us to classify the available germplasm into distinct groups on the basis of their genetic diversity.

MATERIAL AND METHODS

The present investigation was conducted at the Agriculture college, Bapatla, during *kharif*2015. The experiment was laid out in a randomized block design with forty genotypes and three replications. The observations were recorded on ten randomly selected plants from each treatment for days to 50% flowering, days to maturity, productive tillers per plant, plant height, panicle length, grains per panicle, test weight, kernel length, kernel breadth, L/B ratio and grain yield per plant. For days to 50% flowering and days to maturity were recorded on plot basis. Principal component analysis was carried according to procedure described by Banfield (1978). PCA can be performed on two types of data matrices *viz.,* variance-covariance matrix and correlation matrix. With characters of different scale, a correlation matrix standardizing the original data set is preferred. If the characters are of same scale, a variance-covariance matrix can be used. In the present study, PCA was performed on the correlation matrix of traits, thereby removing the effects of scale (Jackson, 1991).

RESULTS AND DISCUSSION

In principal component analysis on correlation matrix, the standardization of columns (here characters) created 11 new variables for 40 genotypes without changing their relative positions. These 11 new variables are the principal components (PC_1, PC_2 …….PC_{11}). Each principal component is a linear combination of the 11 attributes of data matrix.

Results obtained from PCA on the correlation matrix of the traits reduce the dimensionality of the data set by creating several significant principal components having eigen value more than one. The PCA scores for individual

genotypes were used for clustering the genotypes as suggested by Anderberg (1993). Principal components (eigen value greater than one), eigen values (Latent Root), per cent variability, cumulative per cent variability and component loading of different characters are presented in Table 1. In the present study, the first four principal components with eigen values more than one contributed 78.913 per cent towards the total variability. The principal component with eigen values less than one were considered as non- significant. It was therefore inferred that the essential features of data set had been represented in the first four principal components.

The first principal component (PC$_1$) contributed maximum towards variability (39.895%). The characters *viz.,* grain yield per plant (0.447), grains per panicle (0.366), panicle length (0.362), test weight (0.351), productive tillers per plant (0.341), plant height (0.312), days to maturity (0.308), days to 50% flowering (0.276), L/B ratio (0.126), kernel breadth (0.097) and kernel length (0.052) which were positively loaded. The second principal component (PC$_2$) contributed 18.746 per cent of total variance and it reflected significant positive loading of the characters *viz.,* L/B ratio (0.571), days to 50% flowering (0.111), productive tillers per plant (0.076), test weight (0.048), days to maturity (0.031), panicle length (0.017), grain yield per plant (0.005) and kernel breadth (-0.571), kernel length (-0.517), plant height (-0.242) and grains per panicle (-0.014) were negatively loaded. The results were earlier reported by Haque *et al.* (2015) and Ravikumar *et al.*(2015).

The third principal component (PC$_3$) was characterized by 13.810 per cent contribution towards the total variability. The characters *viz.,* Days to maturity (0.519), days to 50% flowering (0.512), kernel length (0.317), L/B ratio (0.179), plant height (0.031), grains per panicle (0.005) were positively loaded. While, productive tillers per plant (-0.397), panicle length (-0.272), test weight (-0.311), kernel breadth (-0.081) and seed yield per plant (-0.001) were negatively loaded. The fourth principal component (PC$_4$) contributed 6.462 per cent towards the total variability and it reflected significant positive loading of the characters *viz.,* grains per panicle (0.624), kernel breadth (0.277), productive tillers per plant (0.075), test weight

(0.074), days to 50% flowering (0.069) and days to maturity (0.064). While, plant height (-0.656), panicle length (-0.248), L/B ratio (-0.091), kernel length (-0.088) and grain yield per plant (-0.078) were negatively loaded.The results were earlier reported by Haque *et al.* (2015) and Ravikumar *et al.*(2015).

The PCA scores for 40 rice genotypes in the first four principal components were computed. Principal component I, II, III and IV were considered as four axis and squared distance of each genotype from these four axis were calculated and presented in Table 2. The analysis thus identified the maximum contributing variables *i.e.,* L/B ratio, days to maturity, days to 50% flowering, grain yield per plant, grains per panicle, panicle length, test weight, productive tillers per plant and plant.It is important to study the variance as the relative contribution than the signs (indicative of direction) in principal component analysis.

Results of cluster analysis based on PCA scores were compared with the results of the principal component analysis on a visual aid as two dimensional and three dimensional scattered diagrams. The genotypes falling in same cluster were present closer to each other in scattered diagram. Two dimentional and three dimentional scatter diagrams were shown in Fig 1 and 2 respective, which are obtained using the first two PCA scores *i.e.,* PCA I and PCA II. The genotypes numbered 22, 5 and 25 *i.e.,* NLR 3390, MTU 2195-105-1-5 and MTU 2195-201-1-3 which are scattered away from other genotypes formed a separate cluster *i.e* cluster IV. Similarly genotypes numbered 12 (NLR 3385) and 17 (BPT 2658) were also scattered and formed two separate monogenotypic clusters while the genotypes which are closer to each are grouped in a single cluster.

But in choosing genotypes for hybridization programme we have to consider the inter-cluster distances. It is evident from the Table 2 that the maximum inter cluster distance was recorded between cluster IV and cluster VII (173.85). Based on the overall results obtained through PCA and agglomerative cluster methods,it is concluded that the genotypes BPT 2658 and NLR 3385 are showing better yield contributing characters are placed in clusters with maximum inter-cluster distance. Hence, these genotypes can be utilized in

Table 1. Eigen values, proportion of the total variance represented by first four principal components, cumulative per cent variance and component loading of different characters in rice (*Oryza sativa* L.).

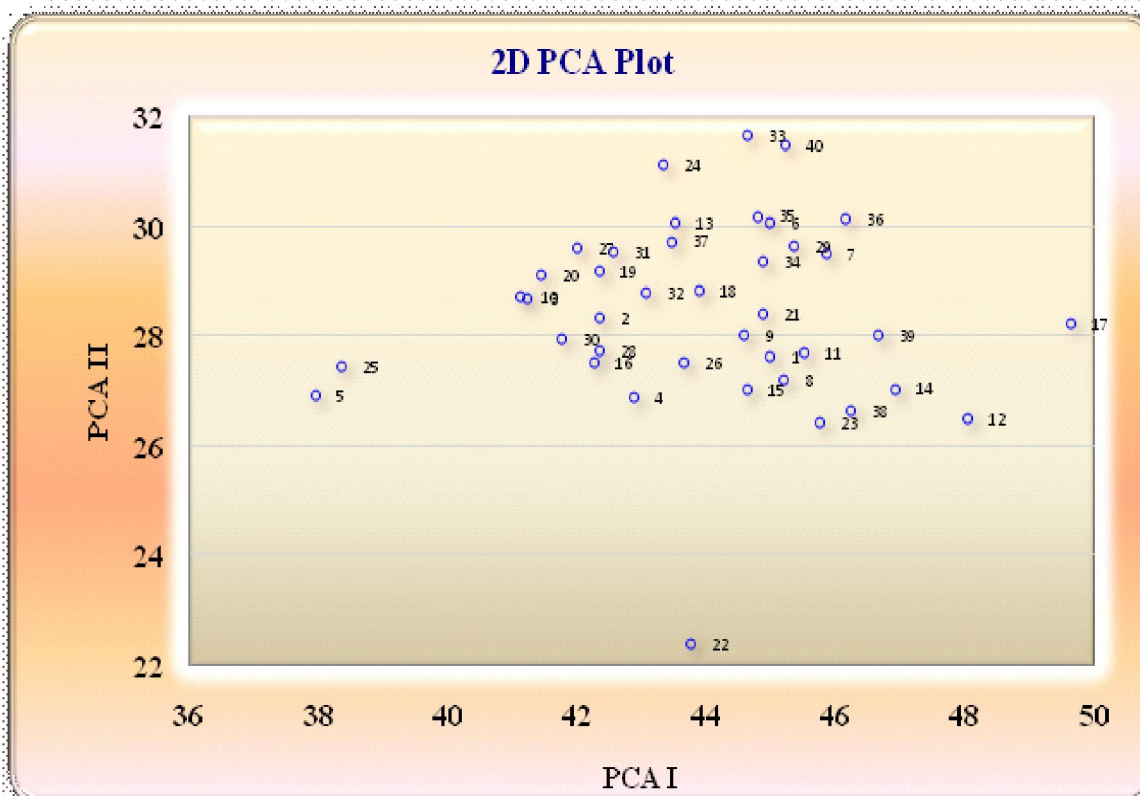|	PCA I	PCA II	PCA III	PCA IV
Eigen Value (Root)	4.388	2.062	1.519	0.711
% Var. Exp.	39.895	18.746	13.810	6.462
Cum. Var. Exp.	39.895	58.641	72.451	78.913
Days to 50% flowering	0.276	0.111	0.512	0.069
Days to maturity	0.308	0.031	0.519	0.064
Productive tillers per plant	0.341	0.076	-0.397	0.075
Plant height (cm)	0.312	-0.242	0.031	-0.656
Panicle length (cm)	0.362	0.017	-0.272	-0.248
Grains per panicle	0.366	-0.014	0.005	0.624
Test weight (g)	0.351	0.048	-0.311	0.074
Kernel length (mm)	0.052	-0.517	0.317	-0.088
Kernel breadth (mm)	0.097	-0.571	-0.081	0.277
L/ B ratio	0.126	0.571	0.179	-0.091
Grain yield per plant (g)	0.447	0.005	-0.001	-0.078

PC = Principal component

Fig. 1. Two dimensional graph showing relative position of 40 rice (*Oryza sativa* L.) genotypes based on PCA scores

Table 2. PCA scores of 40 rice (*Oryza sativa* L.) genotypes.

| S. No.	GENOTYPE	PCA I	PCA II	PCA III	PCA IV
1	RGL 7014	44.988	27.647	27.337	-12.402
2	BPT 2675	42.345	28.349	29.770	-12.305
3	BPT 5204	41.230	28.719	28.509	-12.763
4	MTU 2035-12-2-4-3	42.865	26.921	29.071	-13.319
5	MTU 2195-201-1-3	37.947	26.935	30.073	-11.950
6	BPT 2660	44.969	30.103	26.949	-12.272
7	NLR 3379	45.850	29.513	25.968	-12.103
8	MTU 2071-13-1-1	45.204	27.221	25.557	-11.381
9	MTU 335-19-1-1-1	44.591	28.045	27.271	-12.483
10	JMP 43	41.125	28.741	28.399	-12.715
11	MTU 2111-13-1-2	45.512	27.707	28.391	-13.025
12	NLR 3385	48.040	26.505	27.696	-13.124
13	BPT 2595	43.527	30.075	28.786	-12.344
14	MTU 2127-35-1-1-1	46.909	27.067	28.966	-11.419
15	NLR 3380	44.638	27.042	27.373	-11.887
16	MTU 2127-48-1-1-1	42.258	27.538	28.674	-13.731
17	BPT 2658	49.623	28.250	25.284	-13.236
18	BPT 2743	43.883	28.847	27.677	-12.328
19	NLR 5785-12-2-1	42.355	29.195	29.979	-13.124
20	JMP 24	41.442	29.150	28.320	-11.806
21	MTU 20601-1-1-1-1	44.868	28.418	27.184	-14.306
22	NLR 3390	43.756	22.425	30.854	-12.466
23	BPT 2661	45.749	26.457	27.502	-12.241
24	MTU 7029	43.332	31.163	30.611	-11.524
25	MTU 2195-105-1-5	38.359	27.472	29.322	-11.987
26	MTU 2022-8-3-3	43.661	27.554	27.897	-10.620
27	BPT 2650	42.000	29.617	28.389	-13.012
28	MTU 2244-47-15-6-77	42.341	27.760	29.372	-12.535
29	BPT 2757	45.364	29.675	26.955	-12.682
30	MTU PS 257-1-1-1	41.758	27.983	28.631	-13.519
31	BPT 2587	42.569	29.544	28.047	-12.108
32	MTU 2035-18-1-2-7	43.076	28.810	29.849	-11.944
33	BPT 2644	44.622	31.688	28.161	-14.197
34	BPT 2774	44.872	29.378	29.125	-14.671
35	BPT 2571	44.795	30.212	28.705	-12.491
36	BPT2758	46.145	30.161	27.325	-12.640
37	BPT 2593	43.476	29.754	29.762	-12.850
38	BPT 2763	46.242	26.653	26.903	-13.857
39	BPT 2741	46.650	28.041	28.807	-13.594
40	BPT 2772	45.208	31.503	27.567	-12.094
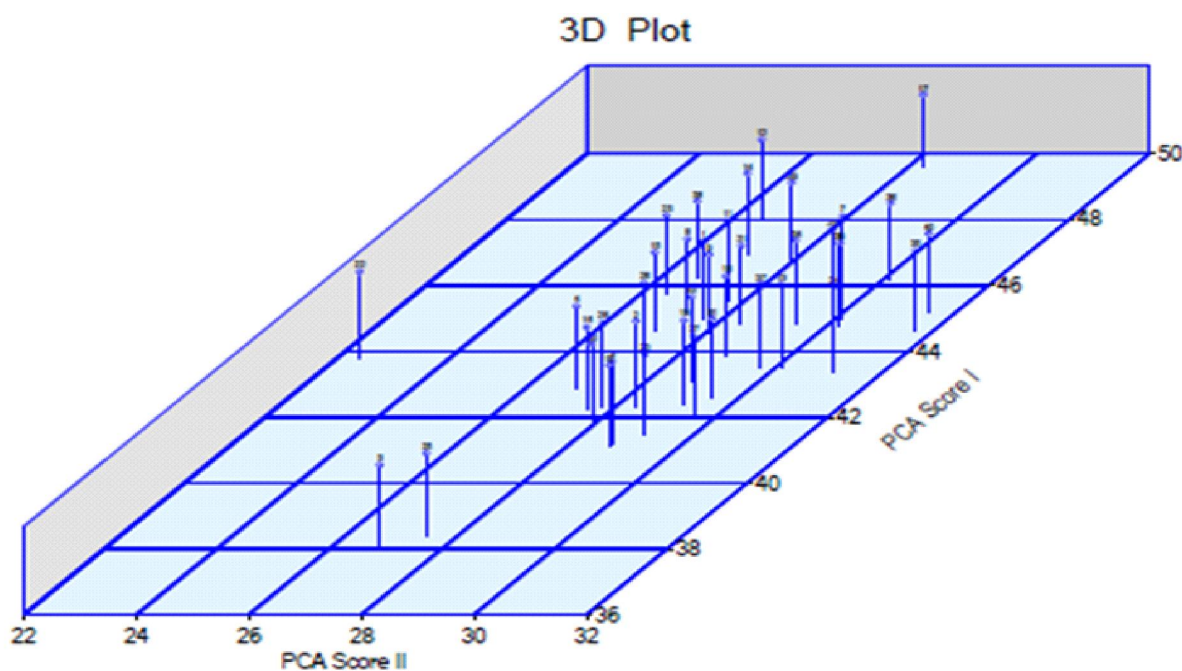
Fig. 2.   Three dimensional graph showing relative position of 40 rice (*Oryzasativa* L.) genotypes based on PCA scores

hybridization programmes to produce desirable recombinants.

LITERATURE CITED

Adams MW 1995 An estimate of homogeneity in crop plants with special reference to genetic vulnerability in dry season. Phseolus vulgaris.Ephytica, 26: 665-679.

Amy E L, Pritts M P 1991 Application of principal component analysis to horticultural research.Hortscience, 26(4): 334-338. Anonymous (2000).Economic survey (2000-2001).Government of Pakistan, Economic advisory wing, Finance Division, Islamabad, pp. 16-17.

Anderberg M R 1993 *Cluster analysis for application.* Academic Press, New York.

Banfield C F 1978 Principal component analysis for genstat.*Journal of Statistics and Computer Simulation*, 6: 211-222.

Haque A, Begum S N, Robin A H K and Hassan L 2015 Partitioning of Rice (*Oryzasativa L.*) Genotypes Based on Morphometric Diversity. *American Journal of Experimental Agriculture*, 7(4): 242-250.

Jackson J E 1991 *A User's Guide to Principal Components.* John Wiley and Sons Inc., New York.

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, Bharathalakshmi M and Reddy A V 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.