Principal Component Analysis for Assessment of Genetic Diversity in Rainfed Shallow Lowland Rice (*Oryza sativa* L)

Prince Raj[1], Anand Kumar[1*], Satyendra[1], S. P. Singh[1], Mankesh Kumar[1], Ravi Ranjan Kumar[2], Bishun Deo Prasad[2] and Sanjay Kumar[1]

[1]*Department of Plant Breeding and Genetics, Bihar Agricultural University, Sabour (Bhagalpur)-813210, India.*
[2]*Department of Molecular Biology and Genetic Engineering, Bihar Agricultural University, Sabour (Bhagalpur)-813210, India.*

Original Research Article

ABSTRACT

The genetic diversity was estimated using seventy two genotypes of rice in a randomized block design with three replications at the rice research farm of Bihar Agricultural University, Sabour (Bhagalpur) during *Kharif*, 2019-20 to determine the contribution of fifteen quantitative traits to the total variability in rice using Principal component analysis. In the present investigation PCA was performed for fifteen quantitative traits of rice. All the 3PCs exhibited more than 1.0 Eigen value and showed about 95.00% variability. Therefore, these PCs were given due important for the further explanation. The PC1 showed 77.28 per cent variation of total variation followed by second to third components which accounted 15.65 and 2.05 per cent of total variation presented among the

*Corresponding author: E-mail: dranandpbgbausabour@gmail.com;

genotypes, respectively. PC1 contributed 77.28% of the total variation and correlated with total carbohydrate, generation of H_2O_2, days to 50% flowering, biological yield, number of fertile grains per panicle, panicle length and flag leaf area while PC2 explained an additional 15.65% of the total variation and dominated by total carbohydrate, days to 50% flowering, harvest index, biological yield, total number of spikelet's and plant height. PC III accounted 2.05 per cent of the total variability and correlated with the traits like days to 50% flowering, biological yield, total number of spikelet's, 1000-seed weight, plant height, harvest index, generation of H_2O_2 and panicle length had maximum positive contribution Since, a total of 95.00% of the total variation was contributed by PC1 and PC2, therefore, these two principal components can be allowed for simultaneous selection of yield contributing traits in desi chickpea. Genotype usually found in more PC, were CR3933-13-2-1-4-1-2-1, TTB1011-14-171-2-2-1-2-1, TTB1032-45-937-2-3-3-1-1, (Santepheap3/IR49830-7/RajendraMahsuri)-1-3-1, (BR11/IR8041OB)-2-1-1, (RajendraMhasuri/CN1039)-4-2-1, TTB1011-14-243-1-2-2-2-1, TTB1032-45-937-2-3-3-1-1, CR4138-3-1-1, CR4139-9-2-1, CR4139-9-2- and CR4128-9-1-1. Genotypes fall in a common principal component were observed to be the most important factor for seed yield. These genotypes may further be utilized in breeding programmes for improving seed yield and these genotypes can be considered an ideotype breeding material for selection of traits *viz.* more total number of seed per plant and 100-seed weight further utilization in precise breeding programme.

1. INTRODUCTION

Genetic diversity is the foundation of the genetic improvement of crops. The knowledge of the extent and pattern of diversity in the crop species is a prerequisite for any crop improvement as it helps breeders in deciding suitable breeding strategies for their future improvement. Yield and yield contributing parameters are the most widely targeted traits for rice improvement programme worldwide. Yield is a complex trait which is affected by several factors and environment, hence, a well-known technique known as principal component analysis was used to identify and minimize the number of traits for effective selection. PCA is a standard tool in modern data analysis because it is a simple, non-parametric method for extracting relevant information from confusing data sets. It involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. It reduces the dimensionality of the data, while retaining most of the variation in the data set. PCA accomplishes this reduction by identifying directions, called principal components. The first principal component accounts for as much of the variability in the data as possible and each succeeding component accounts for as much of the remaining variability as possible. For the choice of diverse parents in any hybridization programme, multivariate analysis (Principal component analysis) has been extensively used. Thus the primary benefit of PCA arises from quantifying the importance of each dimension for describing the variability of a data set. In the present study, we carried out a PCA to identify agronomic attributes whose selection would lead to improvement in grain yield of rice.

2. MATERIALS AND METHODS

The experimental material comprised of seventy two genotypes of rice in Kharif, 2019-20 at rice Research Farm, Bihar Agricultural University, Sabour (Bhagalpur). The experiment was laid in randomized complete block design with three replications during *Kharif,* 2019-20 with inclusion of the recommended packages and practices needed for a healthy crop. Data for fifteen quantitative traits were recorded viz. days to 50 % flowering, days to maturity, plant height (cm), number of effective tillers/hill, flag leaf area, panicle length, number of fertile grains per panicle, total number of spikelet's, fertility%, 1000-seed weight, biological yield, grain yield per plant, harvest index (%), total carbohydrate and generation of H_2O_2. The days to 50% flowering and days to maturity were accounted on a plot basis and plant height (cm), number of effective tillers/hill, flag leaf area, panicle length, number of fertile grains per panicle, total number of spikelet's, fertility%, 1000-seed weight, biological yield, grain yield per plant, harvest index (%) were documented from random sample of five plants in each plot. According to the Massay [1] and Jolliffie [2] PCA is a well-known method of dimension reduction that can be used to reduce a large set of variables to a small set that still

contains most of the information in the large set. Therefore, in present investigation an attempt were taken to evaluate the 72 rainfed shallow lowland rice genotypes on the basis of PCA before taking up hybridization programme for further rice improvement programme.

3. RESULTS AND DISCUSSION

Principal component analysis is a simple non-parametric method for extracting relevant information from confusing data sets. With minimum efforts, this provide a road map for how to reduce a complex data set to a lower dimension to sometimes hidden, simplified structures that often underlines it. PC is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components the number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is accounts for as much of the variability in the data as possible) and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the proceeding components. The resulting vectors are an uncorrelated orthogonal basis set. The principal components are orthogonal because they are the Eigen vectors of the covariance matrix, which is symmetric. In present investigation principal component analysis was performed for yield and yield contributing traits of rice.

Analysis of variance revealed significant differences among the genotypes for all the characters under this investigation (Table-1). Thus, it indicated considerable amount of genetic variability among seventy two rice genotypes. Principal component analysis was performed to reveal the pattern of data matrix for determination and identification of selection criteria. The result of PCA explained the genetic diversity among the rice genotypes. The current research, PCA was performed for six yield and yield attributing traits in rice genotypes. Eigen values of three principal component axes and percentage of variation accounting for them obtained from the principal component analysis are presented in Table 2. According to Brejda et al. [3], data were considered in each components with Eigen value >1 which determined at least 10% of the variation. The higher Eigen values were considered as best representative of system attributes in principal components. As per the calculation, Eigen root values calculated as 38605.9 for the first principal component 7820.64 and 1028.8 for second and third principal component respectively.

The Eigen root of first principal component was accounted approximately 77.28 per cent variation of total variation followed by second to third components which accounted 15.65 and 2.05 per cent of total variation presented among the genotypes, respectively. About 95.00 per cent variation was noted by the third principal components indicating that considerable diversity present among different characters and rest of the components are not much important.

Table 1. Analysis of variance for fifteen quantitative characters in rice

Sl. No.	Characters	Mean sum of square		
	d.f.	Replication (01)	Treatment (71)	Error (71)
1.	Days to 50% flowering	38.03	141.94**	3.52
2.	Days to maturity	32.11	142.31**	5.53
3.	Plant height (cm)	2.7	650.95**	53.05
4.	No. of effective tillers/hill	29.19	11.73**	1.93
5.	Flag leaf area (cm2)	136.71	88.17**	16.99
6.	Panicle length (cm)	2.5	13.18**	2.93
7.	No. of fertile grains/panicle	641.74	3959.77**	558.92
8.	Total number of spikelets	574.04	5522.99**	434.73
9.	Fertility (%)	1.47	217.39**	59.47
10.	1000-seed weight (g)	0.54	42.62**	1.16
11.	Biological yield (g)	584.79	3472.74**	121.39
12.	Grain yield/plant (g)	226.98	280.61**	42.29
13.	Harvest index (%)	0.42	591.76**	49.11
14.	Total carbohydrate	0.13	410.39**	0.28
15.	Generation of H_2O_2	68.48	1636.87**	5.51

**, * = Significant at 5% and 1% level of probability, respectively

Table 2. Eigen value, contribution of variability and Eigen vectors for the principal component axes in rice

Characters	PC I	PC II	PC III
Eigen value (Root)	38605.9	7820.64	1028.8
% Var. Exp.	77.286	15.6563	2.05958
Cum. Var. Exp.	77.286	92.9424	95.0019
Days to 50% flowering	0.0434	0.04956	0.69204
Days to maturity	-0.0097	-0.017	-0.17873
Plant height (cm)	-0.0026	0.00155	0.06859
No. of effective tillers/hill	-0.0053	-0.0052	-0.02345
Flag leaf area (cm2)	0.0013	-0.0026	-0.0880
Panicle length (cm)	0.0031	-0.0118	0.0067
No. of fertile grains/panicle	0.0031	-0.0174	-0.2488
Total no. of spikelet's	0.0202	0.00765	0.12699
Fertility (%)	-0.0039	-0.0157	-0.17215
1000-seed weight (g)	0.0076	-0.0759	0.10098
Biological yield (g)	0.0303	0.0108	0.35656
Grain yield/plant (g)	-0.0161	-0.0114	-0.47062
Harvest index (%)	-0.0131	0.02578	0.06639
Total carbohydrate	0.9138	0.39812	-0.0678
Generation of H_2O_2	0.4012	-0.9118	0.0240

The first principal component vector (PC I) observed 77.28 per cent of the total variability. It could be seen from the first vector that the variables like total carbohydrate (0.913), generation of H2O2 (0.401), days to 50% flowering (0.043), biological yield (0.030), number of fertile grains per panicle (0.003), panicle length (0.003) and flag leaf area (0.001) had the maximum positive contribution towards genetic divergence, while, the remaining characters like grain yield per plant (-0.016), harvest index (-0.031), days to maturity (-0.009), number of effective tillers per hill (-0.005), fertility % (-0.003) and plant height (-0.002) had registered negative contribution.

The second vector (PC II) accounted for 15.65 per cent of total variation present in the population. The major characters that contributed to the second component include traits like total carbohydrate (0.398), days to 50 % flowering (0.049), harvest index (0.025), biological yield (0.010), total number of spikelet's (0.007) and plant height (0.001) had showed maximum positive contribution however, generation of H_2O_2 (-0.911), 1000-seed weight (-0.075), number of fertile grains per panicle (-0.017), days to maturity (-0.17), grain yield per plant (-0.011), number of effective tillers per hill (-0.005) and flag leaf area (-0.002) had maximum negative contribution towards genetic diversity.

PC III accounted 2.05 per cent of the total variability. The traits like days to 50% flowering (0.692), biological yield (0.356), total number of spikelet's (0.126), 1000-seed weight (0.100), plant height (0.06), harvest index (0.066), generation of H_2O_2 (0.024) and panicle length (0.006) had maximum positive contribution, while grain yield per plant (- 0.470), number of fertile grains per panicle (-0.248), days to maturity (-0.178), fertility% (-0.172), flag leaf area (-0.088), total carbohydrate (-0.006) and number of effective tillers per hill (-0.023)) had registered the maximum negative contribution towards genetic divergence.

Cumulatively, these three principal components explained 95.00 % of the total variation in the population. Thus, important characters coming together in different principal components and contributing towards explaining the variability and have the tendency to stay together) and offer opportunity for utilizing in crop improvement programs. Thus the principal component analysis was helpful, in revealing the high level of genetic variation existing in the population and explains which characters contribute for genetic diversity among the genotypes in the population. This will make opportunity for further improvement of the cultivars, in breeding programmes by helping with selection of parents suitable for morphological traits, analyzed in this study.

Important characters coming together in different PCs have tendency to remain together, which may be kept into consideration during utilization of these characters in breeding programme to

bring about rapid improvement for yield and other associated traits. In this study number of phenotypic traits can be identified with the help of principal component analysis, which are responsible for the observed genotypic variation present within each component. Consequently, traits coming collectively in various principal components and contributing towards elucidation the variability and have the propensity to remain together this may be kept into consideration during utilization of these characters in breeding programme.

Thus, PCA revealed principal discriminatory characteristics such as total carbohydrate, generation of H_2O_2, days to 50% flowering, biological yield and total number of spikelet's in diverse PC_s responsible for the reported genotypic variation within a group of genotypes. In this experiment number of phenotypic characters can be recognized with the help of principal component analysis, responsible for the genotypic variation present within each component. Genotype usually found in more

PC, were CR3933-13-2-1-4-1-2-1, TTB1011-14-171-2-2-1-2-1, TTB1032-45-937-2-3-3-1-1, (Santepheap3/IR49830-7/RajendraMahsuri)-1-3-1, (BR11/IR8041OB)-2-1-1, (RajendraMhasuri/CN1039)-4-2-1, TTB1011-14-243-1-2-2-2-1, TTB1032-45-937-2-3-3-1-1, CR4138-3-1-1, CR4139-9-2-1, CR4139-9-2- and CR4128-9-1-1. These genotypes may be used in future breeding programmes for improving grain yield and component characters. These genotypes can be assumed to be an ideotype breeding material for selection of traits viz. biological yield, total number of spikelets, total carbohydrate and generation of H_2O_2 and further utilization in precise breeding programme of rainfed shallow lowland rice. These genotypes which are common in more than one PC indicated that selection of genotype from these PC_s is useful in crop improvement programme in future. These findings are also confirmation with Gana et al. (2013), Nachimuthu et al. [4], Mahendran et al. [5] Gour et al. [6], Sathish and Senapati (2017), Ojha et al. [7], Akliliu et al. [8], Amrita et al. [9], Iqbal et al. [10] .and Miladinovic et al. [11].
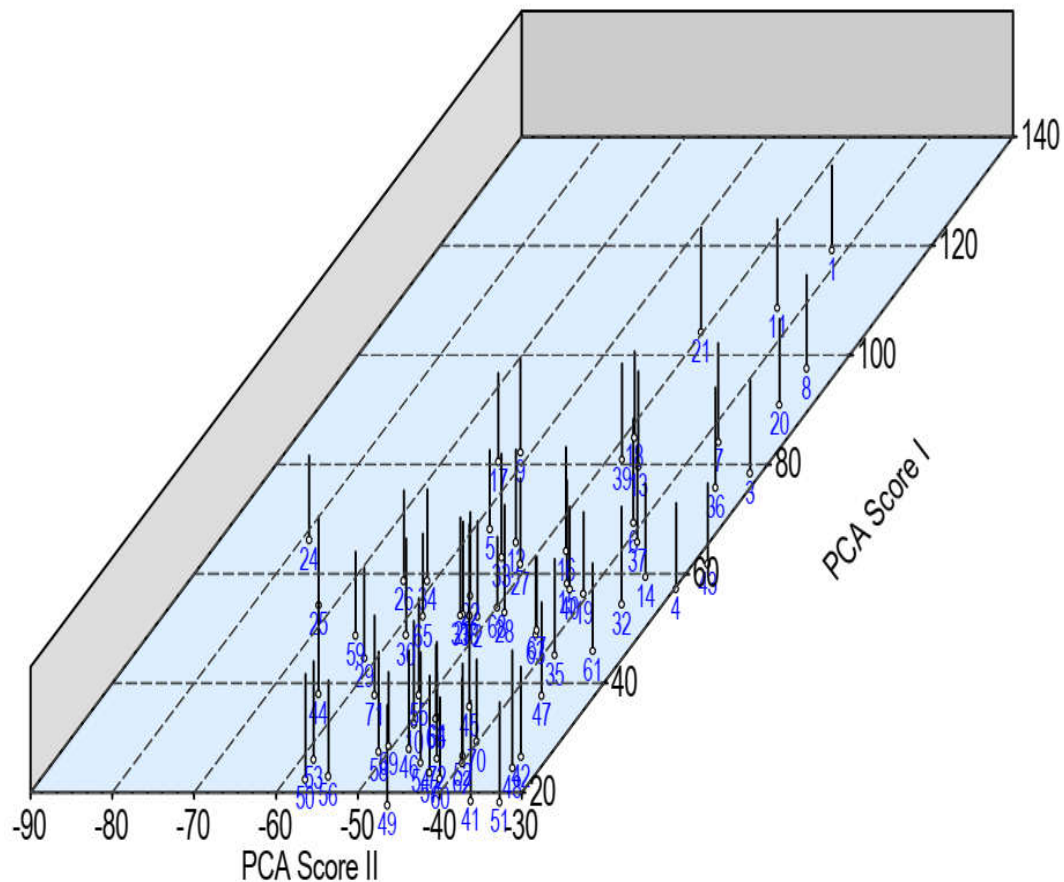

Fig. 1. Distribution and grouping of 72 rice genotypes across first two components based on PCA

4. CONCLUSIONS

The present research revealed that sufficient amounts of variability present in the rice genotypes. The morpho-grain value of the each trait measures the importance and contribution of each component. The results of PCA revealed that the three principal components explained 95.00 % of the total variations, thus suggesting that traits such as biological yield, total number of spikelet's, 1000-grain weight total carbohydrate and generation of H_2O_2 were the principal discriminatory characteristics. Therefore, the important characters coming collectively in various PCs and contributing towards explaining the variability and have the tendency to remain together this may be kept into consideration during utilization of these traits in breeding programme of rice.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Massey WF. Principal components regression in exploratory statistical research. J. Amer. Statist. Assoc. 1965;60:234-246.
2. Jolliffe I. Principal component analysis. John Wiley & Sons, Ltd; 2002.
3. Brejda JJ, Moorman TB, Karlen DL and Dao TH. Identification of regional soil quality factors and indicators. I. Central and Southern High- Plains. Soil Sci. Soc. Am. J. 2000;64: 2115-2124.
4. Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S. Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. Indian Journal of Science and Technology. 2014;7(10):1555-1562.
5. Mahendran R, Veerabadhiran P, Robin S and Raveendran M. Principal component analysis of rice germplasm accessions under high temperature stress. International Journal of Agricultural Science and Research. 2015;5(3):355-359.
6. Gour L, Maurya SB, Koutu GK, Singh SK, Shukla SS, Mishra DK. Characterization of rice (*Oryza sativa* L.) genotypes using principal component analysis including scree plot & rotated component matrix. International Journal of Chemical Studies. 2017;5(4):975-983.
7. Ojha GC, Sarawgi AK, Bhawana Sharma, Parikh M. Principal component analysis of morpho-physiological traits in rice germplasm accessions (*Oryza sativa* L.) under rainfed condition. International Journal of Chemical Studies. 2017; 5(5):1875-1878.
8. Akande SR. Multivariate analysis of the genetic diversity of pigeon pea germplasm from south-west Nigeria. Journal of Food Agriculture and Environment. 2007; 5(1):224.
9. Amrita B, Shrivastava A, Bisen R And Mishra S. Study of Principal Component Analyses for Yield Contributing Traits in Fixed Advanced Generations of Soybean (*Glycine max* (L.) Merrill). Soybean Research. 2014;44.
10. Iqbal Z, Arshad M, Ashraf M, Mahmood T and Waheed A. Evaluation of soybean [*Glycine max* (L.)Merrill] germplasm for some important morphological traits using multivariate analysis. Pakistan Journal of Botany. 2008;40(6):2323-2328.
11. Miladinovic Jegor, Hideki Kurosaki, Joe W Burton, Milica Hrustic, Dragana Miladinovic. The adaptability of short season soybean genotypes to varying longitudinal regions. European Journal of Agronomy. 2006;25:243-249.