Genetic diversity analysis of grain yield and quality traits in rice germplasm accessions (Oryza sativa L.) by principal component analysis

Vageeshvari, Prabha R Chaudhari and Deepak Gauraha

DOI: https://doi.org/10.22271/chemi.2020.v8.i6b.10755

Abstract
The Rice germplasm provide with ample of genetic diversity and a treasury of valuable genes. It is a rich pool of important genes that plant breeders can exploit for crop improvement. Principal components with eigen values greater than 1 and variance level greater than 4 level is perceived to be the main PC. In this study, eight components exhibited Eigen values of >1 and showed cumulative variation of 75.22%. The first PC (15.83%) and second PC (15.33%) showed for cumulative variation 31.36%. The PC1 recorded 15.83%, although PC2, PC3, PC4, PC5, PC6, PC7 and PC8 displayed 15.33%, 12.64%, 9.03%, 7.4%, 5.91%, 4.61% and 4.27% variance respectively between the accessions for the characters in analysis. Genotypes Saraiphool, MTU 1010, Zinc rice-2, Chaptimathalya, Bora, Korma, R-RKM-1, Kalajeera, Jauphool, Tiltaturi and Parwatkala were found to be promising and found to be genetically divergent. Hence, these genotypes could be utilized in the hybridization programme to produce transgressive segregants.

Keywords: Soil and foliar, nitrogen, boron and salicylic, quality, Konkan

Introduction
Rice (Oryza sativa L.) is the most important food crop of about 3 billion people, nearly half the world's population, depends on rice for survival and main crop to fight world's hunger and poverty. In India it is cultivated in an area of 45.54 million hectares with a production of 99.18 million tones and productivity of 2177kg ha-1(Anon 2012).[2] Rice gives a great wealth of material for genetical studies because of its wide ecological distribution and enormous discrepancies encountered for multitudinous morphological and physiological characteristic. The Rice germplasm provide with ample of genetic diversity and a treasury of valuable genes. It is a rich pool of important genes that plant breeders can exploit for crop improvement (Yadav et al.,2013) [14]. Traversing diversity in a land race collection is essential for identifying novel genes and hence improvement of the germplasm (Thomson et al., 2007) [13]. Rice germplasm is a heritage of mankind that has evolved through several millennia of cultivation and selection by our farming ancestors. In Chhattisgarh, over 20,000 rice cultivars have been and is commonly known as Rice Bowl of India. Some farmers’ varieties/land races/local varieties are still popular in some pockets of different parts of Chhattisgarh due to their special medicinal values which are not present in modern varieties. Medicinal values of traditional rice varieties in Chhattisgarh were available on farmer perceptions. Total thirteen rice germplasm accessions are available as medicinal rice genotypes and each having different properties to cure different health problems in human and animals. (Richhariya et al., 1978). Analysis of related to yield and their attributing traits and their compression with the increasing population is a important aspects by which a good result may be made to fulfil the current demand. Grain yield is a complex character and is controlled by many factors. Selection for desirable types should not only be restricted to grain yield alone but other components related to grain quality should also be considered. Any crop improvement program depends on the utilization of germplasm stock available in different research organizations institutes of the world. In different statistical analysis Principal Component Analysis (PCA) is a powerful tool in modern data analysis because it is a simple, non-parametric method for extracting relevant information from confusing data sets.
This technique was initially floated by Pearson (1901) \(^8\) and later developed by Hotelling (1933) \(^4\). With minimal effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden information that often underlie it (Shlens, 2009) \(^12\). 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, along which the variation in the data is maximal (Anderson, 1972 and Morrison, 1978) \(^1\). By using a few components, each sample can be represented by relatively few numbers instead of by values for thousands of variables (Ringer, 2008) \(^10\). Thus, the primary benefit of PCA arises from quantifying the importance of each dimension for describing the variability of a data set in more interpretable and more visualized dimensions through linear combinations of variables that accounts for most of the variation present in the original set of variables. Considering the importance of PCA an investigation was carried out on rice germplasm lines with an objective to identification of the traits responsible for the yield and quality and identified diverse genotypes for further breeding programme.

**Material and methods**

The experimental material comprised forty five indigenous rice germplasm (thirteen genotypes having medicinal value and other genotypes selected on the husk and grain colour basis) along with checks were grown in randomized complete block design with two replications at the Research cum Instructional Farm, Department of Genetics and Plant Breeding, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh during Kharif 2019. Crop was raised following recommended package of practices. Observations were recorded on five randomly tagged plants of each genotype per replication. Data were recorded on thirty eight quantitative and quality traits.

**Statistical analysis:** Principal Component analysis was also used to determined genetic variability for these traits. Genotypic means were used for the PCA with respect to each trait. Each data analysis was conducted by using SAS (Statistical Analysis System) version 9.2.

**Results and Discussion**

The analysis of variance indicated the existence of highly significant differences among genotypes for all the characters studied. This indicated presence of high variability among the genotypes, which provides ample scope for selection for different quantitative characters and quality traits for rice improvement. Significant variation in all the traits studied indicated the presence of high genetic diversity among all the genotypes of rice. PCA was performed on rice genotypes for quantitative and quality traits to identify the important components and genotypes identified for better characters related to rice improvement programmes. In this study, eight components exhibited Eigen values of >1 and showed cumulative variation of 75.22\% (Table 1). The first PC (15.83%) and second PC (15.33%) showed for cumulative variation 31.36\%. Principal components, Eigen values, percentage contribution of every variable to overall variance and major contributing characteristics for each major component is described in table 1.

The important characters of PC1 and PC2 were the quality traits, PC3 and PC 4 were quantitative traits, PC5, PC6, PC7and PC8 were combination of quantitative and quality traits describe in table 1. Characters that have both positive and negative impacts on the PCs can be said to be the key source of variability and have mainly contributed to the differentiation of the rice genotypes. High-variability characters are required to have high range of gene transfer during crop improvement program (Nachimuthu et al., 2014) \(^8\). Therefore, these traits can be used in the choice of diverse genotypes from specific principal component. The outcomes of the current study were consistent with the findings of Sao et al., (2019) \(^11\), Ojha et al., (2017) \(^7\) and Gaur et al., (2017) \(^3\). The distribution of rice genotypes based on the first and second PC exhibited the phenotypic variation among the population and explains how these widely dispersed along both the axes. PCA scatter diagram showing the dispersion of rice genotypes across PC1 and PC2 and also shown in Fig. 1. The genotype wise scatter diagram of the rice genotypes (scores) across the first two PC axes (Fig.2) revealed that genotypes Saraiphool (31), MTU 1010 (42), Zinc rice-2 (43), Chaptimathyala (44), Bora (14), Korma (17), R-RKM-1 (19), Kalajeera (20), Jauphol (21), Tiltaturi (22) and Parwatkala (23) were very divergent for the characters under analysis. The derived information of this research on rice genotypes would be very useful to select potentially breeding lines for future rice improvement programmes. A good hybridization breeding program can be initiated by the selection of genotypes viz., Saraiphool, MTU 1010, Zinc rice-2, Chaptimathyala, Bora, Korma, R-RKM-1, Kalajeera, Jauphol, Tiltaturi and Parwatkala were found to be promising and found to be genetically divergent. (Table 2).

| Table 1: Eigen value, percentage of variance and eigen vector of forty rice germplasm accessions for yield and grain quality traits |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Traits**      | **Principal component** |
|                 | PC1   | PC2   | PC3   | PC4   | PC5   | PC6   | PC7   | PC8   |
| Eigen values    | 5.54  | 5.43  | 4.42  | 3.16  | 2.59  | 2.06  | 1.61  | 1.49  |
| % of Variance   | 15.83 | 15.33 | 12.64 | 9.03  | 7.4   | 5.91  | 4.61  | 4.27  |
| Cumulative%     | 15.83 | 31.36 | 44    | 53.03 | 60.43 | 66.34 | 70.95 | 75.22 |
| **Component Matrix** |
| Days to 50% flowering | 0.08  | -0.12 | 0.17  | -0.33 | -0.09 | 0.08  | 0.29  | 0.11  |
| Plant height (cm) | 0.05  | -0.07 | 0.25  | -0.29 | 0.12  | 0.13  | -0.06 | -0.18 |
| Flage leaf length (cm) | 0.10  | -0.05 | 0.07  | -0.19 | 0.32  | 0.10  | 0.04  | -0.05 |
| Leaf length of blad (cm) | -0.03 | 0.02  | 0.29  | 0.08  | 0.24  | 0.14  | -0.29 | -0.09 |
| Leaf width of blad (cm) | 0.04  | -0.05 | 0.22  | 0.06  | -0.12 | 0.09  | -0.02 | 0.02  |
| Panicle length (cm) | 0.06  | -0.21 | 0.18  | -0.16 | -0.03 | 0.24  | -0.25 | -0.08 |
| Number of effective panicles per plant | -0.07 | -0.06 | 0.06  | 0.42  | 0.19  | 0.26  | 0.01  | 0.22  |
| Number of non-effective panicles/ plant | -0.12 | -0.09 | 0.13  | 0.19  | -0.25 | 0.17  | -0.08 | -0.04 |
| Total number of tiller per plant | -0.07 | -0.07 | 0.06  | 0.42  | 0.17  | 0.27  | 0.00  | 0.22  |
| Number of filled spikelets per panicle | 0.21  | -0.23 | 0.25  | -0.11 | 0.10  | 0.08  | 0.06  | 0.03  |
| Parameter                                                                 | Mean 1 | Mean 2 | Mean 3 | Mean 4 | Mean 5 | Mean 6 |
|--------------------------------------------------------------------------|--------|--------|--------|--------|--------|--------|
| Number of unfilled spikelets per panicle                                | 0.15   | -0.25  | 0.17   | -0.02  | -0.21  | -0.16  |
| Total number of grains per panicle                                       | 0.21   | -0.25  | 0.25   | -0.10  | 0.01   | 0.02   |
| Spikelet fertility%                                                      | -0.05  | 0.14   | -0.03  | -0.06  | 0.37   | 0.25   |
| 1000 grain weight (g)                                                    | 0.08   | 0.30   | 0.18   | -0.13  | 0.04   | -0.06  |
| Biological yield per plant (g)                                           | -0.02  | -0.08  | 0.13   | 0.15   | 0.37   | -0.23  |
| Grain yield per plant (g)                                                | 0.13   | 0.00   | 0.24   | 0.20   | 0.14   | -0.14  |
| Harvesting index (%)                                                     | 0.15   | 0.09   | 0.12   | 0.07   | -0.23  | 0.08   |
| Paddy length (cm)                                                        | 0.27   | 0.25   | 0.08   | 0.04   | -0.02  | 0.07   |
| Paddy width (cm)                                                         | -0.07  | 0.31   | 0.18   | -0.13  | 0.10   | -0.10  |
| Biological yield per plant (g)                                           | -0.02  | -0.08  | 0.13   | 0.15   | 0.37   | -0.23  |
| Grain yield per plant (g)                                                | 0.13   | 0.00   | 0.24   | 0.20   | 0.14   | -0.14  |
| Harvesting index (%)                                                     | 0.15   | 0.09   | 0.12   | 0.07   | -0.23  | 0.08   |
| Paddy length (cm)                                                        | 0.27   | 0.25   | 0.08   | 0.04   | -0.02  | 0.07   |
| Paddy width (cm)                                                         | -0.07  | 0.31   | 0.18   | -0.13  | 0.10   | -0.10  |
| Biological yield per plant (g)                                           | -0.02  | -0.08  | 0.13   | 0.15   | 0.37   | -0.23  |
| Grain yield per plant (g)                                                | 0.13   | 0.00   | 0.24   | 0.20   | 0.14   | -0.14  |
| Harvesting index (%)                                                     | 0.15   | 0.09   | 0.12   | 0.07   | -0.23  | 0.08   |
| Paddy length (cm)                                                        | 0.27   | 0.25   | 0.08   | 0.04   | -0.02  | 0.07   |
| Paddy width (cm)                                                         | -0.07  | 0.31   | 0.18   | -0.13  | 0.10   | -0.10  |
| Biological yield per plant (g)                                           | -0.02  | -0.08  | 0.13   | 0.15   | 0.37   | -0.23  |
| Grain yield per plant (g)                                                | 0.13   | 0.00   | 0.24   | 0.20   | 0.14   | -0.14  |
| Harvesting index (%)                                                     | 0.15   | 0.09   | 0.12   | 0.07   | -0.23  | 0.08   |
| Paddy length (cm)                                                        | 0.27   | 0.25   | 0.08   | 0.04   | -0.02  | 0.07   |
| Paddy width (cm)                                                         | -0.07  | 0.31   | 0.18   | -0.13  | 0.10   | -0.10  |

**Fig 1:** Two dimensional PCA diagram showing the dispersion of rice germplasm lines

**Fig 2:** Scree Test for Component Analysis
Table 2: List of selected accession in each principal component on the basis of top 10 PC score

| S. No. | PC1     | PC2       | PC3      | PC4   | PC5     | PC6     | PC7     | PC8     |
|--------|---------|-----------|----------|-------|---------|---------|---------|---------|
| 1      | R-RKM-1 | Davar     | Parwatkala | Bhejari | Suldhan | Soth    | Bora    | IR64    |
| 2      | Saraihpoo | Barhasal  | Raj banko | Kujjii | Karhani | Kujjii  | Jauphool | Madhuraj dhan-55 |
| 3      | Bora    | Saraihpoo | Aalcha   | Karhani | Resari  | Hundar  | Swarna  | Kujjii  |
| 4      | Zinc rice-2 | Kurela | Kurela | Kurela | Kurela | Kurela | Kurela | Kurela |
| 5      | MTU 1010 | Resari    | Bora     | Bora   | Gathuwan | Laycha | Soth    | Laycha  |
| 6      | Mehдержан | Danwar   | Kal much | Maharaji | Jauphool | Sathi   | Sathi   | Korma   |
| 7      | Lal banko | Hundar    | Chepti gurmatiya | Zinc rice-2 | MTU 1010 | Jauphool | Zinc rice-2 | Raj banko |
| 8      | Kal much | Karhani   | Lal banko | Sathi   | Zinc rice-2 | Karhani | MTU 1010 | Parwatkala |
| 9      | Raj banko | Nagpuri gurmatiya | Khatbuti | Danwar | Kujjii | Lal banko | Davar | Bhejari |
| 10     | Aalcha  | Sathi     | Korma    | R-RKM-1 | Davar   | Tilkaturi | Karela | Swarna |

References
1. Anderson TW. An introduction to multivariate analysis. Wiley Eastern Pvt Ltd New Delhi, 1972.
2. Anonymous. Production and productivity of rice in India, USDA 2012-13.
3. Gour L, Maurya SB, Koutu GK, Singh SK, Shukla SS, Mishra DK, et al. 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):975-83.
4. Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of educational psychology 1933;24(6):417-441.
5. Morrison DE. Multivariate Statistical Methods (2nd ed. 4th Print, 1982). McGraw Hill Kogakusta Ltd, 1982.
6. Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S, et al. Evaluation of rice genetic diversity and variability in a population pannel by principal component analysis. Indian J Sci. Technol 2014;7(10):1555-1562.
7. Ojha GC, Sarawgi AK, Sharma B, 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. Pearson K. On lines and planes of closest fit to systems of points in space Philosophical Magazine 1901;2:559.
9. Richharia RH. An aspect of genetic diversity in rice. *Oryza* 1979;16:1-31.
10. Ringer M. What is principal component analysis? Nature Biotechnology 2008;26(3):303-304.
11. Sao R, Saxena RR, Sahu PK. Assessment of genetic variation and diversity in rice genotypes based on principal component analysis. Journal of Plant Development Sciences 2019;11(12):725-730.
12. Shlens J. A tutorial on principal component analysis Version 3.01 2009.
13. Thomson MJ, Septiningsih EM, Suwardjo F, Santosio TJ, Siliongta TS, McCouch SR, et al. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. Theor. Appl. Genet 2007;114:559-568.
14. Yadav S, Singh A, Singh MR, Goel N, Vinod KK, Mohapatra T, et al. Assessment of genetic diversity in Indian rice germplasm (*Oryza sativa* L.): use of random versus trait-linked microsatellite markers. J. of genet 2013;92(3):545-557.