Principal Component Analysis for Yield and its Attributing Traits in Aromatic Landraces of Rice (*Oryza sativa* L.)

Maumita Burman[*], Sunil Kumar Nair and Arvind Kumar Sarawgi

Dept. of Genetics and Plant Breeding, Indira Gandhi Krishi Vishwavidyalaya (IGKV) Raipur, Chhattisgarh (492 012), India

Abstract

The present investigation was carried out in *Kharif* 2019 (July to November) to estimate the relative contribution of various traits for total genetic variability present in aromatic landraces by Principal Component Analysis. Here 90 aromatic rice landraces along with six check varieties were evaluated for 13 quantitative characters by Principal Component Analysis. Principal Component Analysis showed that, out of 13 quantitative characters studied, only five principal components (PCs) exhibited more than 1.00 eigen value and showed about 81.62% cumulative variability among the traits studied. Out of the five principal components exhibiting more than 1.00 eigen value PC1 had the highest variability (25.12%) followed by PC2 (21.8%). The first principal component PC1 was positively contributed mainly by two characters viz., Grain Length and 1000 grain weight. The second principal component PC2 was contributed mostly by three characters like grain yield plant[-1], panicle weight and spikelet fertility percentage. The third principal component PC3 is positively associated with panicle weight, grain yield plant[-1] and spikelet fertility percentage. The fourth principal component PC4 is positively associated with spikelet fertility percentage, Grain Length/ Breadth ratio and fertile grains panicle[-1]. The fifth principal component PC5 is positively associated with total grains per panicle[-1], grain width and 1000 grain weight. All the principal components were showing positive contribution for yield and its attributing traits. These variations can be exploited in crop improvement programme for developing high yielding varieties.

1. Introduction

Rice (*Oryza sativa* L.) is the most important cereal crop and primary energy source for two thirds of world's population (Khan et al., 2015). It is the staple crop for more than half of the world population, and it is also rich in genetic resources. Rice germplasm diversity is an important genetic information transmission system (Nambara and Nonogaki 2012; Nachimuthu et al., 2015). Adequate knowledge of genetic variation in different genotypes is a preliminary step in breeding programs for the selection and production of new varieties (Kumbhar et al., 2015; Ahmed et al., 2016). Wild ancestors and landraces with rich genetic diversity and wide adaptation to various environments provide valuable and useful genetic resources for crop improvement (Kovach and McCouch,

2008; Sang and Ge, 2013; Dwivedi et al., 2016). In spite of the richness of genetic resources, only a small proportion has been utilized in breeding programs, resulting in high genetic similarity in commercial rice cultivars (Das et al., 2013).

Aromatic rice is an important commodity worldwide and command premium prices in local and international market over non-scented varieties because of their superior grain quality and pleasant aroma (Nayak et al., 2002). In India, aromatic rice is grown in almost all the states, covering more than 30% of the total cultivated area (Chakravorty and Ghosh, 2013; Sarma et al., 2016). Majority of Indian aromatic rice genotypes are having small to medium grains. Since middle of the 20th century the conventional breeding methodologies contributed in the significant improvement of yield leading to the development of potential cultivars (Perez-de-Castro et al., 2012). Analysis of data related to yield and their attributing traits and their compression with the increasing population is an important aspect by which a good result may be made to fulfill the current demand.

Rice breeders generally observe a large number of traits, many of which may not be used in the discrimination of germplasm. In such cases principal component analysis (PCA) may be used to reveal patterns and eliminates redundancy in data sets (Maji and Shaibu, 2012). Principle component analysis (PCA) is used to assess variation as multivariate methods (Tiwari et al., 2020). It reduces the dimensions of a multivariate data to a few principal axes, generates an Eigen vector for each axis and produces component scores for the characters (Leonard and Peter, 2009). It analyses data consisting of several inter correlated quantitative dependent variables as observations (Nachimuthu et al., 2014; Mahendran et al., 2015). It extracts the information from a table and represents it as a set of new orthogonal variables called principal components (Table 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). 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. The higher the coefficients, regardless of the direction (positive or negative), the more effective they will be in discriminating between accessions. Considering the importance of PCA an investigation was carried out to study the principal component analysis for yield and its attributing traits in aromatic landraces of rice (*Oryza sativa* L.).

2. Materials and Methods

2.1. Experimental site details

This experiment was carried out at Research cum Instructional Farm, Department of Genetics and Plant Breeding, College of Agriculture, Indira Gandhi Agricultural University, Raipur,

Table 1: List of 96 Rice Genotypes used in this study

Sl. No.	Name	Sl. No.	Name
1.	Aama Gohi	49.	Kali Muchh
2.	Aatma Shital	50.	Kali Muchha
3.	Atma Shital	51.	Lokti Machhi
4.	Akbar Badshah	52.	Sukra Phool
5.	Bag Muchh	53.	Tendu Phool
6.	Wasmati	54.	Basmati
7.	Banspatri	55.	Chhatri (C: 194)
8.	Barang	56.	Chhatri (C: 808)
9.	Badshah Bhog (B: 484)	57.	Bhatta Phool
10.	Badshah Bhog (B: 497)	58.	Til Kasturi
11.	Badshah Bhog (B: 2355)	59.	Kubri Mohar
12.	Basha Bhog	60.	Lalloo
13.	Bassa Bhog	61.	Angar Moti
14.	Bisnu Bhog (B:1094)	62.	Kadam Phool
15.	Muni Bhog	63.	Laichi Phool
16.	Chini Kapoor (C:30)	64.	Lallu
17.	Chini Kapoor	65.	Chini K Apoor (C: 459)
18.	Chinnur	66.	Bag Muchh
19.	Chinnour	67.	Tulsi Manjari Ii
20.	Chirai Nakhi	68.	Sukla Phool
21.	Dubraj (D: 268)	69.	Ganga Baru
22.	Dubraj (D: 421)	70.	Jai Gundi
23.	Dubraj (D: 433)	71.	Bayasa Bhog
24.	Dubraj (D: 874)	72.	Bisnu Bhog
25.	Dubraj (D: 1420)	73.	Raja Bhog
26.	Dubraj Dhan Deshi	74.	Vishnoo Bhog
27.	Majhali Dubraj	75.	Vishnu Bhog
28.	Ganga Balu	76.	Bikoni
29.	Bam Baijira	77.	Urai Buta
30.	Bhanta Phool (B: 1087)	78.	Chendara Chhal
31.	Jui Phool	79.	Chinnour
32.	Ramkali	80.	Duban Mua
33.	Dubraj (D: 925)	81.	Badshah Bhog (B: 562)
34.	Samund Chini	82.	Basa Bhog
35.	Samundar Fen	83.	Dubraj (D: 33)
36.	Sansari	84.	Dubraj (D:80)
37.	Sarsariya	85.	Dubraj (D: 90)
38.	Jira Shankar	86.	Dubraj (D: 272)

Table 1: Continue...

Sl. No.	Name	Sl. No.	Name
39.	Til Kasturi	87.	Dubraj (D: 934)
40.	Tulsi Amrit	88.	Dubraj (Deshi)
41.	Tulsi Bas	89.	Dudh Nag
42.	Tulsi Mala	90.	Bhanta Phool
43.	Tulsi Manjari	91.	Mahamaya (C)
44.	Badshah Bhog (B:1294)	92.	Tarun Bhog selection 1 (C)
45.	Badshah Bhog (B: 1510)	93.	C.G. Dev Bhog Selection 1 (C)
46.	Chhatri Bhog	94.	Badshah Bhog Selection 1 (C)
47.	Gobind Bhog	95.	Vishnu Bhog selection 1 (C)
48.	Raja Bhog (R: 399)	96.	Dubraj selection 1 (C)

Chhattisgarh during July, *kharif* 2019 Geographically, Chhattisgarh state lies between 17°14' to 24°06' North Latitudes and 80°14' to 84°24' East Longitude. Raipur, the capital of Chhattisgarh, is situated in East Central part of state at latitude of 21°16' N, longitude 81°36' E and at an altitude of 289.6 meters above mean sea level. The climate of the region is sub-humid with mean annual rainfall of about 1489 mm. The experimental material consists of 96 rice genotypes. Here, a core set of 90 aromatic landraces was prepared from 571 aromatic germplasm lines of I.G.K.V., Raipur, Chhattisgarh based on the aroma content of their leaves by KOH Sensory test method (Sood and Siddique, 1978). These 90 aromatic rice landraces along with six check varieties namely, Mahamaya, Tarun bhog Selection 1, C.G. Devbhog Selection1, Badshah bhog Selection 1, Vishnu bhog Selection1 and Dubraj Selection 1 were used in the present study.

2.2. Planting pattern, plot size and statistical analysis

Nursery sowing was done in well prepared raised seed bed in first week of July 2019 and the crop was ready to harvest by mid of November 2019. Here twenty eight days old seedlings were transplanted in well puddle field in Augmented Block Design as suggested by Federer, 1956. 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. The distance between each block was maintained at 50 cm. The randomization of check varieties was done within each block. Each genotype was transplanted without replication. Five random but robust plants from inner rows were tagged from each plot for data collection. A total of thirteen quantitative traits viz., days to 50 % flowering, plant height (cm), effective tillers plant[-1], panicle weight, thousand grain weight (g), grain length (mm), grain breadth (mm), grain length breadth ratio, fertile grains panicle[-1], sterile grains panicle[-1], total grains panicle[-1], spikelet fertility percentage

and grain yield plant[-1] (g) were measured at particular stages of rice plant following the minimal descriptor of rice. The observations recorded were statistically analyzed using PAST v3.14 software.

3. Results and Discussion

The results of Principle Component Analysis explained the genetic variation among the genotypes for all agro-morphological and grain quality characters under study. Principal components with eigen values more than 1 and variation percent more than 4% were considered as main PC (Brejda et al., 2000). The outcome of the PCA described the genetic diversity among rice genotypes for the studied traits. 'Eigen values' measure the importance and contribution of each component to total variance, whereas each coefficient of eigen vectors indicates the degree of contribution of every original variable with which each principal component is associated. There are no standard tests to prove significance of Eigen values and the coefficients (Jolliffe, 2002) (Table 2 and Figure 1). Proper values measure the importance and contribution of each component to total variance, while each value indicates the degree of contribution of each

Table 2: Eigen value, proportion and cumulative variation of analyzed components

Principal components	Eigen value	The proportion of variation %	Cumulative variation %
PC1	3.27	25.12	25.12
PC2	2.83	21.80	46.92
PC3	2.06	15.86	62.78
PC4	1.29	9.96	72.74
PC5	1.15	8.88	81.62
PC6	0.89	6.85	88.47
PC7	0.79	6.04	94.51
PC8	0.44	3.38	97.89
PC9	0.20	1.53	99.42
PC10	0.04	0.31	99.73
PC11	0.03	0.26	99.99
PC12	0.00	0.03	100.02
PC13	0.00	0.00	100.02

original variable associated with each main component. The characters coming together in different principal components explaining the variability show the tendency to remain together and must be taken into consideration during the exploitation of these characters in the breeding program (Chakravorty et al., 2013). Principle Component Analysis revealed that out of thirteen characteristics studied, only five principal components (PCs) exhibited more than 1.00 eigen value and showed about 81.62% cumulative variability among the traits studied while the other components were rejected
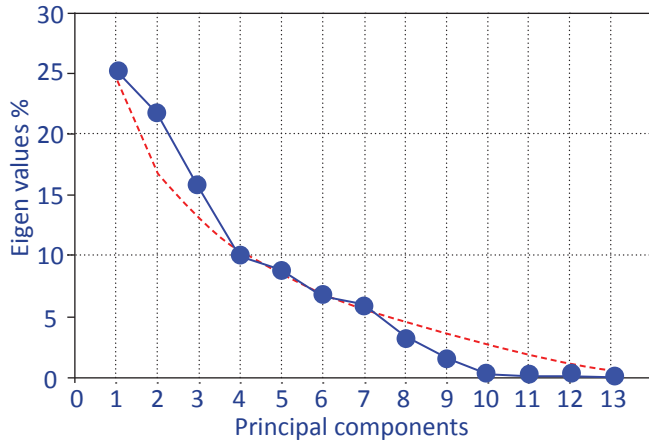
Figure 1: Scree plot diagram of Eigen values constructed on thirteen traits recorded in 96 rice genotypes

the same positive effect.

The third principal component PC3 is positively associated with panicle weight (0.42), grain yield plant^{-1} (0.39) and spikelet fertility percentage (0.33). The fourth principal component PC 4 is positively associated with spikelet fertility percentage (0.44), Grain Length/Breadth ratio (0.42) and fertile grains panicle^{-1} (0.37). The fifth principal component PC5 is positively associated with total grains panicle^{-1} (0.44), grain width (0.42) and 1000 grain weight (0.37). (Table 3 and Figure 2). Through PCA we could identify the characters responsible for genotypic variation within the group. Four principal components with Eigen value greater than >1 and

because they have eigen values less than one. So, these five PCs were given due importance for further explanation. Out of the five principal components (PCs) exhibiting more than 1.00 eigen value PC1 had the highest variability (25.12%) followed by PC2 (21.8%).

The first principal component PC1 was mainly contributed by two characters viz., Grain Length (0.48) and 1000 grain weight (0.42). The second principal component PC2 was contributed mostly by three characters like grain yield plant^{-1} (0.44), panicle weight (0.43) and spikelet fertility percentage (0.41). Grain yield plant^{-1} and yield attributing traits i.e. panicle weight, spikelet fertility percentage positively affect the PC2 on the contrary days to 50% flowering is negatively correlated with PC2. This indicates the increase in grain yield plant^{-1} and yield attributing traits i,e panicle weight, spikelet fertility percentage as a result of early flowering. To be sure, improving a given yield trait will direct the improvement of other yield traits collected in the same PC as long as they have
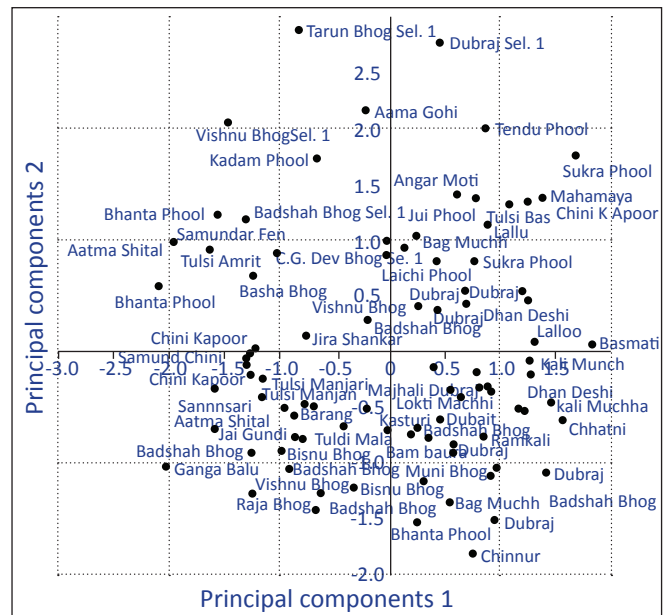

Figure 2: Biplot distribution of 96 rice genotypes and studied traits depending on principal component axes PC1 and PC2

Table 3: Principle components for studied traits in 96 rice genotypes

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Days to 50% flowering	-0.21	-0.18	0.29	0.20	-0.27	-0.07	-0.62	0.58	0.03	-0.01	0.03	0.02	0.00
Plant height	-0.21	-0.06	-0.01	0.01	-0.01	0.97	0.04	0.06	0.02	0.02	0.01	0.01	0.00
Effective tillers plant^{-1}	-0.08	0.28	0.18	-0.18	-0.42	-0.06	0.64	0.50	0.04	0.00	0.04	0.01	0.00
Panicle weight	0.02	0.43	0.42	-0.13	-0.11	0.05	-0.19	-0.24	0.05	0.57	-0.43	-0.02	0.00
1000 seed weight	0.42	0.17	0.09	0.12	0.37	0.10	0.01	0.35	-0.71	0.08	-0.03	0.00	0.00
Grain length	0.48	-0.02	0.21	0.24	0.12	0.09	0.06	0.10	0.46	-0.05	0.06	-0.64	0.00
Grain width	0.15	0.31	-0.21	-0.42	0.42	0.03	-0.20	0.35	0.46	-0.05	0.01	0.33	0.00
Grain L/B ratio	0.38	-0.17	0.29	0.42	-0.08	0.05	0.16	-0.09	0.21	-0.02	-0.03	0.69	0.00
Fertile grains panicle^{-1}	-0.39	0.24	0.09	0.37	0.33	-0.07	0.12	0.04	0.07	-0.21	-0.25	-0.01	0.64
Sterile grains panicle^{-1}	-0.10	-0.35	0.32	-0.31	0.28	-0.05	0.11	-0.03	0.00	0.36	0.51	0.03	0.33
Total grains panicle^{-1}	-0.40	0.05	0.28	0.19	0.44	-0.09	0.17	0.02	0.07	-0.02	0.01	0.00	-0.70
Spikelet fertility %	-0.07	0.41	0.33	0.44	-0.08	-0.01	-0.06	-0.02	0.06	0.42	0.58	0.03	0.00
Grain yield	0.05	0.44	0.39	-0.14	-0.12	0.08	-0.20	-0.29	-0.12	-0.57	0.40	0.02	0.00

explained 72.48% of the total variance were also recorded by Ilieva et al. (2019). It indicates that the identified traits within the axes exhibited great influence on the phenotype of germplasm lines. PCA has been used by various researchers like Gana et al. (2013), Yugandhar et al. (2018) and Ilieva et al. (2019) for characterization of different rice germplasm lines. PCA helps us to indentify the characters which have great impact in phenotype of different landraces of rice, and this is very much important in the selection procedure of breeding programme.

4. Conclusion

The first principal component PC1 was positively contributed mainly by two characters viz., Grain length and 1000 grain weight. The second principal component PC2 was contributed mostly by three characters like grain yield plant^{-1}, panicle weight and spikelet fertility percentage. On the basis of principal component analysis as depicted in Figure 2, the landraces Sukla Phool, Tendu Phool, Agar Moti, Jui Phool and Tulsi Bas are showing maximum contribution for yield attributing traits.

5. References

Ahmed, M.S., Bashar, M.K., Shamsuddin, A.K.M., 2016. Diversity level, Spearman's ranking and core collections from 98 rice germplasm through quantitative, qualitative and molecular characterizations. Rice Genomics and Genetics 7(2), 1–10.

Brejda, J.J., Moorman, T.B., Karlen, D.L., Dao, T.H., 2000. Identification of regional soil quality factors and indicators in central and southern high- plains. Soil Science Society of American Journal 64, 2115–2124.

Chakravorty, A., Ghosh, P.D., 2013.Characterization of landraces of rice from Eastern India. Indian Journal of Plant Genetic Resources 26(1), 62–67.

Chakravorty, A., Ghosh, P.D., Sahu, P.K., 2013. Multivariate analysis of phenotypic diversity of landraces of rice of West Bengal. American Journal of Experimental Agriculture 3(1), 110–123.

Das B., Sengupta S., Parida, S.K., Roy, B., Ghosh, M., Prasad, M., Ghose, T.K., 2013. Genetic diversity and population structure of rice landraces from Eastern and North Eastern states of India. BMC Genetics 14, 71.

Dwivedi, S.L., Ceccarelli, S., Blair, M.W., Upadhyaya, H.D., Are, A.K., Ortiz, R., 2016. Landrace germplasm for improving yield and abiotic stress adaptation. Trends in Plant Science 21(1), 31–42.

Federer, W.T., 1956. Augmented (or Hoonuiaku) designs. Hawaiian Planters' Record 55, 191–208.

Gana, A.S., Shaba, S.Z., Tsado, E.K., 2013. Principal component analysis of morphological traits in thirty-nine accessions of rice (*Oryza sativa*) grown in a rainfed lowland ecology of Nigeria. Journal of Plant Breeding Crop Science 5, 120–126.

Ilieva, V., Ruzdik, N.M., Mihajlov, L., Ilievski, M., 2019. Assessment of agro-morphological variability in rice using multivariate analysis. Journal of Agriculture and Plant Sciences 17(1), 79–85.

Jolliffe, I.T., 2002. Principal component analysis (2nd Edn.). Springer-Verlag, New York, USA.

Khan, M.H., Dar, Z.A., Dar, S.A., 2015. Breeding strategies for improving rice yield -A review. Agricultural Sciences 6, 467–478.

Kovach, M.J., McCouch, S.R., 2008. Leveraging natural diversity: back through the bottleneck. Current Opinion in Plant Biology 11(2), 193–200.

Kumbhar, S.D., Kulwal, P.L., Patil, J.V., Sarawate, C.D., Gaikwad, A.P., Jadhav, A.S., 2015. Genetic diversity and population structure in landraces and improved rice varieties from India. Rice Science 22(3), 99–107.

Leonard, K., Peter, R.J., 2009. Finding Groups in Data: An Introduction to Cluster Analysis, 344.

Mahendran, R., Veerabadhiran, P., Robin, S., Raveendran, M., 2015. Principal component analysis of rice germplasm accessions under high temperature stress. International Journal of Agricultural Science and Research 5(3), 355–360.

Maji, A.T., Shaibu, A.A., 2012. Application of principal component analysis for rice germplasm characterization and evaluation. Journal of Plant Breeding and Crop Science 4(6), 87–93.

Nachimuthu, V.V., Muthurajan, R., Duraialaguraja, S., Sivakami, R., Pandian, B.A., Ponniah, G., Sabariappan, R., 2015. Analysis of population structure and genetic diversity in rice germplasm using SSR markers: an initiative towards association mapping of agronomic traits in *Oryza sativa*. Rice 8(1), 30.

Nachimuthu, V.V., Robin, S., Sudhakar, D., Raveendran, M., Rajeswari, S., Manonmani, S., 2014. Evaluation of rice genetic diversity and variability in a population pannel by principal component analysis. Indian Journal of Science and Technology 7(10), 1555–1562.

Nambara, E., Nonogaki, H., 2012. Seed biology in the 21st century: Perspectives and new directions. Plant and Cell Physiology 53(1), 1–4.

Nayak, A.R., Chaudhury, D., Reddy, J.N., 2002. Genetic variability, heritability and genetic advance in scented rice. Indian Journal of Agricultural Science 46(12), 45-47.

Perez-de-Castro, A.M., Vilanova, S., Canizares, J., Pascual, L., Blanca, J.M., Diez, M.J., Prohens, J., Pico, B., 2012. Application of genomic tools in plant breeding. Current Genomics 13(3), 179–195.

Ringer, M., 2008. What is principal component analysis? Nature Biotechnology 26(3), 303–304.

Sang, T., Ge, S., 2013. Understanding rice domestication and implications for cultivar improvement. Current Opinion in Plant Biology 16(2), 139–146.

Sarma, B., Basumatary, N.R., Nahar, S., Tanti, B., 2016. Effect

of drought stress on morpho-physiological traits in some traditional rice cultivars of Kokrajhar district, Assam, India. Annals of Plant Sciences 5(8), 1402–1408.

Sood, B.G., Siddiq, E.A., 1978. A rapid technique for scent determination in rice. Indian Journal of Genetics and Plant Breeding 38, 268–271.

Tiwari, S., Yadav, M.C., Dikshit, N., Yadav, M.C., Pani, D.R., Latha, M., 2020. Morphological characterization and genetic identity of crop wild relatives of rice (*Oryza sativa* L.) collected from different ecological niches of India. Genetic Resources and Crop Evolution 67, 2037–2055.

Yugandhar, R.P., Suneetha, K., Kiran, U.B., Sridhar, M., 2018. Principal component analysis for agro-morphological and quality characters in germplasm of rice (*Oryza sativa* L.). International Journal of Advanced Biological Research 8(2), 268–273.