RESEARCH ARTICLE https://doi.org/10.58297/OWFQ8879

Estimation of Genetic Diversity by Principal Component Analysis of Yield Attributing Traits in Katarni Derived Lines

Divya Mahto[1], Singh PK [2], Rabiya Parveen[3], Sareeta Nahakpam[4] and Mankesh Kumar[5]*

[1, 3 &5]Department of Plant Breeding and Genetics, Bihar Agricultural University, Sabour, Bhagalpur, Bihar
[2]Director, Seed and Farms, Bihar Agricultural University, Sabour, Bihar
[4]Department of Plant Physiology and Biochemistry, Bihar Agricultural University, Sabour, Bihar
*Corresponding author's Email: drmankeshkumar@gmail.com

Received: 12th September 2022; Accepted: 30th October 2022

Abstract

Katarni is a non-basmati traditional aromatic rice cultivar being grown in the Bhagalpur, Banka, and Munger districts of Bihar. However, it is weak-strawed, tall, prone to lodging, and late maturing. Attempt was made to develop a semi-dwarf and early maturing lines of Katarni by crossing with three semi-dwarf high-yielding cultivars and was advanced to F_5 generation. In this study, 54 derived lines of Katarni were studied on the basis of 14 morphological traits. Five principal components (PCs) were observed which contributed 70% of cumulative variability and exhibited Eigenvalues> 1. The first PC (23.31%) and the second PC (35.59%) showed a cumulative variation of 63.90%. On the basis of genotype by trait biplot analysis, flag leaf length, plant height, and fragrance were found to be strongly positive. Genotypes biplot study revealed diverse genotypes like KIR-46 KIR-48, KRS-39, KRS-43, KRS-15, KRS-19, KRS-8, KRS-25, and KMTU-52 which can be further exploited for varietal development.

Keywords: Eigenvalue, Hybridization, Katarni, Principal component

Introduction

Rice (*Oryza sativa* L.) is a staple food for more than 2.7 billion people (Tannidi *et al.,* 2016) around the world, and about 32- 59% of the dietary energy and 25-44% of the dietary protein is obtained from rice in more than 39 countries (Prabhu *et al.,* 2017). According to Nagaraju *et al*., (2002), the aromatic Basmati lies in a separate group between *indica* and *japonica* in which the traditional Basmati and evolved Basmati varieties represent a major component of the Basmati gene pool of the Indian subcontinent. Katarni is a non-basmati traditional aromatic rice cultivar of the Bhagalpur district of Bihar. This rice is one of the famous fine-grained aromatic rices of India which is renowned for its unique aroma, special grain, and cooking qualities. Its flowering occurs between the end of October to the beginning of November and matures in the month of December. Plant height ranges from 140 to 160 cm (Smriti *et al.,* 2016). In view of its uniqueness, Katarni rice has been granted geographical indication in April 2018. However, the available Katarni is a poor yielder (25-30 t/ha), weak strawed, traditionally tall type, easily prone to lodging and late maturing (Kumar *et al.,* 2018). Principal component analysis (PCA) is generally used to estimate the relative contribution of various traits for total variability and a small number of factors that account for maximum variability can be identified easily. It also shows the pattern of similarity of the traits and relation among the traits. Further, PCA identifies the minimum number of components, which can explain the maximum variability out of the total variability (Anderson 1972), and also ranks genotypes on the basis of PC scores. Several researchers have characterized rice germplasm including the landraces, varieties, and advanced materials of diverse nature for morphological and physicochemical quality parameters (Bollinedi *et al.* 2020; Madhubabu *et al*. 2020), and reported a wide range of variability. Considering the importance of PCA, the present experiment was laid out to identify

diversified genotypes with short stature and early maturity with high-yielding ability in the segregating population generated by crossing Katarni with R. Sweta, IR-64, and MTU-7029. Among the derived lines, principal component analysis was carried out to identify diversified lines which can be utilised for future breeding programmes.

Materials and methods

The experimental material comprised 54 Katarni-derived families, four parental checks i.e. Katarni, R. Sweta, IR-64, MTU-7029, and two aromatic checks Sabour Surbhit and Rajendra Suwasini. The derived families of Katarni were in F_5 generation and were grown in alpha lattice design with two replications at Rice Section, Bihar Agricultural university, Sabour, Bhagalpur during *Kharif* 2018. For convenience, the genotypes of Katarni x Rajendra Sweta, Katarni x IR64, and Katarni x MTU7029 were denoted as KRS, KIR, and KMTU, respectively. The 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 fourteen quantitative and quality traits.

Principal Component analysis is a very important tool to minimize the large data set into a new set of uncorrelated variables (known as principal components) by a linear transformation of original variables. In the present study, genotypic means were used to determine genetic variability for the traits in PCA. The data analysis was conducted using SAS (Statistical Analysis System) version 9.2. For PCA, eigenvalues were calculated first which define the amount of total variation that was displayed on the PC axis. Then, loading values were standardized in such a way that the sum of squares of loadings within a PC was equal to one. The loading values depicted the contribution of each trait in the respective principal component.

Result and Discussion

The analysis of variance studied revealed the presence of significant variability for the traits which indicated diversity among the genotypes. Principal component analysis was performed to trace out major components and their contributing traits as well as genotypes in respective components. The PCA revealed up to seven principal components (PC1 to PC7). Among seven PCs, five components contributed 70% of cumulative variability and exhibited Eigenvalues > 1, i.e. PC1 (3.26), PC2 (2.28), PC3 (1.77), PC4 (1.25), PC5 (1.24), PC6 (0.94) and PC7 (0.73). The first PC (23.31%) and second PC (35.59%) showed for cumulative variation of 63.90%. Principal components, Eigenvalues, factor loading values, the percentage contribution of every variable to overall variance, and major contributing characteristics for each major component are described in **Tables 1** and **2**. The important characters and major contributors to variability in PC1 were kernel length, length and breadth ratio, thousand-grain weight, and panicle length. Whereas in PC2, important characters were flag leaf length, plant height and panicle length, and kernel breadth. Gelatinization temperature and number of tillers per plant in PC3 (**Table 3**); the number of tillers per plant and amylose content in PC4; fragrance, days to 50% flowering, and plant height in PC5 were major contributors to variability. Traits with high variability are essential during crop improvement program (Nachimuthu *et al.,* 2014). Therefore, the selection of kernel length, length and breadth ratio, plant height, number of tillers per plant, and panicle length can be used in the choice of diverse genotypes from the specific principal component. The outcomes of the current study were consistent with the findings of Sao *et al.,* (2019), Ojha *et al.,* (2017), and Gaur *et al.,* (2017). The factor loading value was found to be maximum for kernel length (0.88) in PC1, flag leaf length (0.75) in PC2, kernel breadth (0.71) in PC3, amylose content (0.79) in PC4, fragrance (0.65) in PC5, grain yield per plant in PC6 and PC7.

Maximum variability in PC1 was contributed by genotype KIR-46 (9.49%) followed by KIR-48 (7.21%) and KRS-39 (6.16%) (Table 3), whereas in PC2, maximum variability was contributed by KRS-43 (12.42%) followed by KRS-15 (8.58%) and KRS-19 (5.30%). In PC3, KRS-25 (11.99%) was followed by KRS-8 (11.97%) and KRS-4 (9.49%), in PCA4 KMTU-53 (11.94%) was followed by KRS-7 (11.38%) and KRS-39 (6.04%), were the major variability contributors. The highest contribution

Table 1. Eigen value, percentage of variance and eigenvector of Katarni derived lines

PCA / Components	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	3.26	2.28	1.77	1.25	1.24	0.94	0.73
Variability (%)	23.31	16.28	12.67	8.95	8.85	6.72	5.24
Cumulative %	23.31	39.59	52.26	61.21	70.06	76.78	82.02
Component matrix	**Factor Loading Value**						
PH	-0.03	0.67	0.05	0.20	0.31	0.34	-0.15
DOF	-0.59	-0.06	0.07	0.19	0.46	0.02	-0.31
FLL	-0.03	0.75	0.00	0.08	-0.36	-0.11	0.25
PL	0.46	0.65	0.29	0.12	0.01	-0.14	0.00
NOT	-0.11	-0.52	0.35	0.57	-0.09	-0.37	-0.05
GPP	-0.76	0.29	-0.18	-0.10	-0.05	0.14	0.15
GW_100	0.71	0.05	0.10	-0.22	0.27	0.39	-0.08
GY	-0.05	-0.52	0.30	0.15	-0.01	0.49	0.55
ASV	0.17	0.15	0.64	-0.24	-0.52	-0.02	-0.10
AMY	0.32	0.25	0.11	0.79	0.02	0.15	0.05
FRAG	0.00	0.22	0.27	-0.19	0.65	-0.45	0.43
KL	0.88	-0.17	-0.28	0.02	0.06	-0.09	0.02
KB	0.41	-0.19	0.71	-0.17	0.15	0.03	-0.13
L/B	0.73	-0.10	-0.59	0.08	-0.01	-0.11	0.06

Table 2. Contribution of each trait in different principal components

PCA / Traits	PC1	PC2	PC3	PC4	PC5	PC6	PC7
PH	0.02	19.98	0.16	3.11	7.87	12.44	2.87
DOF	10.67	0.17	0.31	2.91	17.42	0.03	13.23
FLL	0.02	24.59	0.00	0.53	10.50	1.39	8.67
PL	6.46	18.61	4.79	1.23	0.00	2.17	0.00
NOT	0.37	11.68	6.80	25.52	0.69	14.23	0.38
GPP	17.59	3.62	1.79	0.76	0.23	2.06	2.95
GW_100	15.42	0.12	0.62	4.01	5.74	15.93	0.90
GY	0.08	11.73	5.06	1.69	0.01	25.56	41.26
ASV	0.84	1.03	23.03	4.52	21.49	0.05	1.49
AMY	3.11	2.85	0.66	49.83	0.03	2.28	0.39
FRAG	0.00	2.20	4.25	2.96	33.88	21.69	24.84
KL	23.99	1.32	4.35	0.03	0.34	0.77	0.05
KB	5.15	1.67	28.46	2.40	1.78	0.08	2.42
L/B	16.29	0.44	19.71	0.51	0.02	1.31	0.54

Table 3. Contribution of each genotype in different principal components

Sl. No.	Genotypes	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7
1.	KIR-44	0.74	0.31	0.06	0.09	2.80	4.89	0.80
2.	KIR-45	4.07	2.36	0.75	0.04	3.37	0.30	2.83
3.	KIR-46	9.49	1.25	0.96	2.80	0.12	6.59	4.87
4.	KIR-47	3.02	0.25	2.75	0.00	0.34	0.08	11.60
5.	KIR-48	7.21	0.00	0.48	0.95	0.00	0.02	2.98
6.	KIR-49	5.38	0.10	0.56	1.24	6.31	5.79	2.62
7.	KMTU-50	1.13	0.70	0.11	0.57	0.14	0.68	0.53
8.	KMTU-51	0.99	0.94	0.59	0.20	0.22	0.26	0.42
9.	KMTU-52	2.70	0.07	0.20	1.27	1.98	0.25	10.56
10.	KMTU-53	1.19	1.06	2.00	11.94	0.17	0.11	0.11
11.	KMTU-54	0.41	1.56	0.49	5.62	12.94	0.69	3.35
12.	KRS-1	0.02	0.44	1.34	0.76	0.18	0.90	0.05
13.	KRS-10	2.36	1.61	0.80	4.84	2.53	0.02	0.83
14.	KRS-11	0.77	0.23	3.15	0.13	1.26	7.23	0.00
15.	KRS-12	0.22	0.00	3.48	0.16	0.79	0.18	1.18
16.	KRS-13	0.66	1.05	2.59	4.26	0.05	0.39	1.38
17.	KRS-14	0.89	1.43	2.00	0.00	0.79	0.00	0.30
18.	KRS-15	0.30	8.58	0.03	2.36	0.29	0.12	0.03
19.	KRS-16	0.06	4.20	0.08	0.33	0.80	0.51	3.52
20.	KRS-17	0.59	3.66	0.03	0.02	0.53	0.03	0.03
21.	KRS-18	0.03	2.14	1.12	0.90	3.31	3.85	0.03
22.	KRS-19	0.24	5.30	0.32	0.24	0.61	0.01	1.10
23.	KRS-2	0.00	0.80	0.20	2.26	1.05	1.89	0.92
24.	KRS-20	0.11	1.20	3.06	0.31	0.04	0.54	0.28
25.	KRS-21	0.67	1.08	0.08	0.37	1.29	1.85	1.52
26.	KRS-22	0.04	4.40	0.09	0.58	0.15	0.21	0.03
27.	KRS-23	0.21	0.15	0.20	0.02	0.31	0.99	0.24
28.	KRS-24	0.06	0.24	0.52	1.83	0.94	9.09	5.84
29.	KRS-25	1.34	1.78	11.99	0.00	2.38	0.25	8.79
30.	KRS-26	0.14	0.51	0.19	0.11	0.12	0.50	0.97
31.	KRS-27	0.54	0.30	2.65	0.08	1.58	2.25	0.45
32.	KRS-28	0.01	0.12	1.86	2.10	1.42	4.13	0.09
33.	KRS-29	2.73	0.12	0.54	0.73	0.52	0.04	1.85
34.	KRS-3	2.45	0.35	0.95	0.12	3.39	5.54	0.43
35.	KRS-30	5.46	0.52	0.00	0.03	0.22	1.17	1.83
36.	KRS-31	0.97	0.74	8.82	3.89	0.16	6.78	1.18
37.	KRS-32	4.26	0.52	2.27	3.83	0.00	0.76	0.70

Sl. No.	Genotypes	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7
38.	KRS-33	1.03	0.00	4.23	2.31	0.00	0.03	0.22
39.	KRS-34	1.80	1.12	1.46	0.12	0.24	0.61	0.71
40.	KRS-35	2.01	0.21	0.05	0.25	0.00	0.13	0.67
41.	KRS-36	0.16	0.06	0.02	0.04	1.21	0.03	0.23
42.	KRS-37	0.90	0.00	4.23	0.83	0.22	2.25	0.66
43.	KRS-38	0.68	0.45	1.30	3.65	3.28	0.33	0.09
44.	KRS-39	6.16	0.19	0.01	6.04	1.34	1.32	0.05
45.	KRS-4	0.08	0.42	9.49	3.39	0.42	0.03	3.93
46.	KRS-40	3.16	1.70	0.00	0.00	2.90	0.01	1.69
47.	KRS-41	0.02	1.77	0.50	0.05	0.05	2.70	0.71
48.	KRS-42	0.91	0.27	0.01	0.73	6.90	1.12	1.86
49.	KRS-43	0.46	12.42	0.04	1.52	9.47	7.78	2.48
50.	KRS-5	0.06	3.49	1.54	0.18	0.40	0.63	0.01
51.	KRS-6	0.01	4.26	0.02	0.11	0.82	4.45	0.14
52.	KRS-7	0.90	0.45	0.00	11.38	1.24	0.07	0.93
53.	KRS-8	0.06	0.27	11.97	2.20	0.83	2.48	0.06
54.	KRS-9	4.06	0.40	1.12	1.46	0.83	0.07	2.80
55.	MTU7029	0.44	1.32	0.50	0.00	0.09	2.94	2.38
56.	IR-64	4.42	2.18	0.49	0.96	1.22	1.23	0.12
57.	Katarni	2.01	14.75	1.59	8.52	9.81	1.00	1.35
58.	R. Sweta	0.22	1.35	1.98	0.63	1.47	0.80	1.17
59.	R. Suwasini	4.02	2.66	1.49	0.35	4.12	0.12	2.63
60.	S. Surbhit	4.97	0.19	0.64	0.29	0.00	0.98	0.81

PH: Plant height, DOF: Days to 50% flowering, FLL: Flag leaf length, PL: Panicle length, NOT: Number of tillers/hill, GPP: Number of grains/panicle, GW-100: 1000-grain weight, ASV: Alkali spreading value, AMY: Amylose content, FRAG: Fragrance, KL: Kernel length, KB: Kernel breadth, LB: L/B ratio and GY: Grain yield/plant

for variability in PC5 was contributed by genotypes KMTU-54 (12.94%) followed by KRS-43 (9.47%) and KIR-49 (6.31%). Among the checks, Katarni (14.75%) was the major variability contributor in PCA2. The derived information of PCA on F_5 lines of Katarni would be very useful to select potential and diverse breeding lines for future rice improvement programmes.

Scree plot explained the percentage of variation by plotting a graph between eigenvalues and cumulative variability (%) on the Y axis and the mean value of 14 characters under study on the X axis (**Figure 1**). The scree plot explained the percentage of variation by each PC and its eigenvalues. As depicted in the graph majority of variations were contributed by the first three PCs. The distribution of the scores for the 14 different characters in the scree plot indicated the presence of large diversity.

Comparison of genotypes on the basis of measured multiple variables are possible by Genotype by Trait (GT) biplot which identifies those genotypes that are particularly superior in certain traits. The GT biplot can be effectively used as an independent selection criterion of genotypes on the basis of yield (Yan and Rajcan, 2002). The distance to the biplot origin, known as the vector length of a trait is indicative

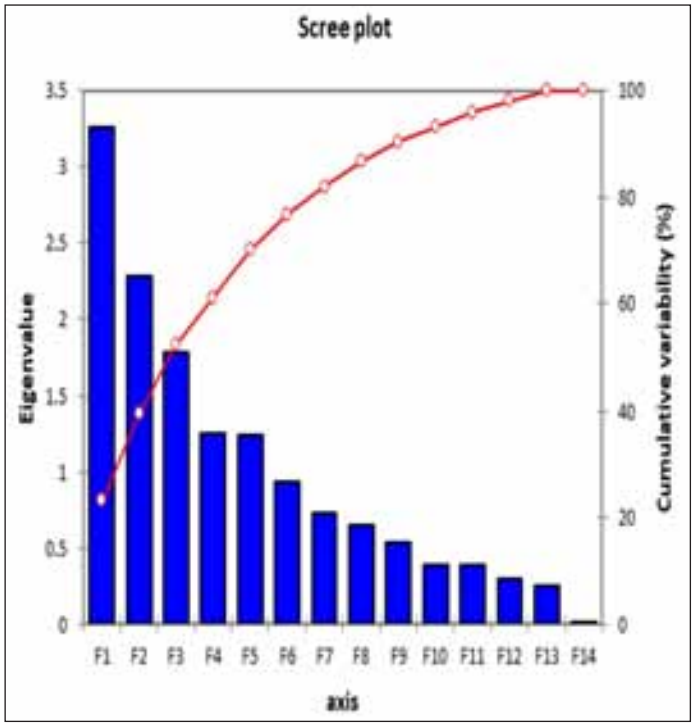

Figure 1: Scree plot of different components with Eigen values

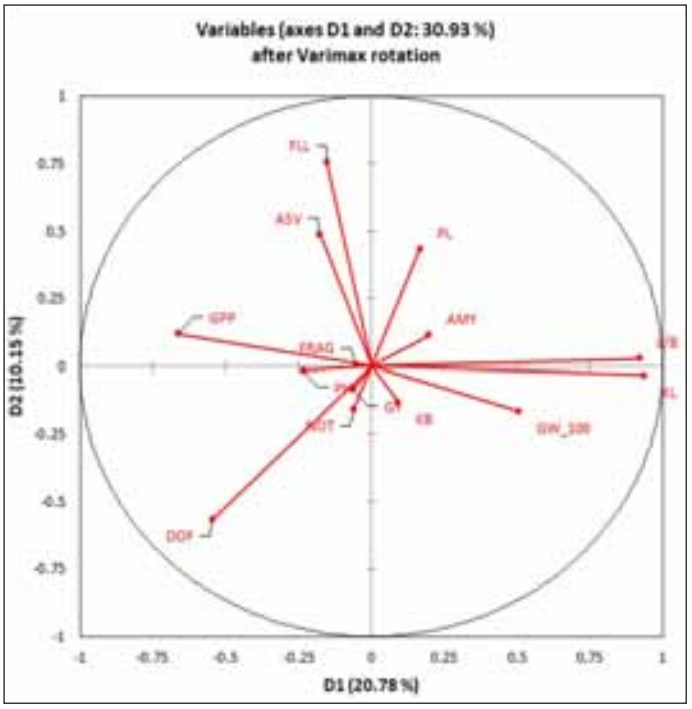

Figure 2: Biplot of 14 different morphological characters

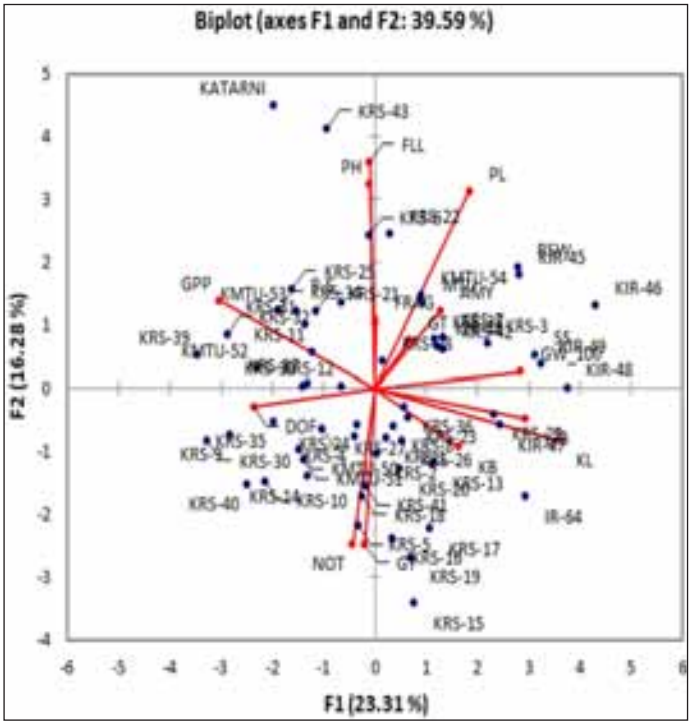

Figure 3: Biplot of 60 genotypes including checks

of the trait representation in the biplot. A relatively short vector indicates that the variation of the trait across genotypes is either small or not well presented in the biplot due to its weak or lack of correlation with other traits (Yan and Fregeau-Reid, 2018). PC1 and PC2 variables in biplot analysis showed both positive and negative associations among the traits. Flag leaf length, plant height, and fragrance were strongly positively correlated as the axes recorded an angle less than 90^0 (**Figure 2**). Similarly, panicle length with gelatinization temperature and amylose content; length/ breadth ratio with kernel length and kernel breadth, and number of tillers per plant with grain yield per plant were positively correlated as these traits showed axes angle less than 90°. A few traits like the number of grains per panicle with kernel breadth and thousand seed weight with days to 50% flowering were negatively associated as these traits are placed at approximately 180° angle on PC1 and PC2 axes. Similarly, the genotypes biplot study (**Figure 3**) revealed that entries KIR-46 KIR-48, KRS-39, KRS-43, KRS-15, KRS-19, KRS-8, KRS-25, KMTU-52, and Katarni are distantly placed from the origin of axes indicating their diversity with respect to other genotypes under study. Therefore, the selection of kernel length, length and breadth ratio, plant height, number of tillers per plant, and panicle length can be used in the selection of diverse genotypes and a

hybridization breeding program can be initiated by using the diverse genotypes obtained in the present study.

Acknowledgements

We sincerely acknowledge the grants received from the SERB, DST, New Delhi (Sanction no. SB/YS/LS-81/2013) and BAU, Sabour (State plan code: SNP/CI/2017-2). Research Article bears BAU communication no. 1218/220722.

References

Anderson TW. 1972. Principal Components. P. 459-483. In An introduction to multivariate analysis: Wiley Eastern Private Limited, New Delhi, Third Edition.

Bollinedi H, Vinod K, Bisht K, Chauhan A, Krishnan SG, Bhowmick PK, Nagarajan M, Rao DS, Ellur RK and Singh AK. 2020. Characterising the diversity of grain nutritional and physico-chemical quality in Indian rice landraces by multivariate genetic analyses. *Indian Journal of Genetics*, 80: 26-38.

Gour L, Maurya SB, Koutu GK, Singh SK, Shukla SS, Mishra DK. (2017. Characterization of rice (*Oryza sativa* L.) genotypes using principal component analysis including scree plot & rotated component matrix. *International Journal of Chemical Studies*, 5:975-83.

Kumar M, Satyendra, Singh SP, Kumar A, Smriti, Shashikant, Singh PK and Sohane RK. 2018. Characterisation of Katarni rice for its geographical indication in Bihar, India. *Journal of AgriSearch,* 4: 223-229.

Madhubabu P, Surendra R, Suman K, Chiranjeevi M, Fiyaz RA, Rao DS, Chaitanya U, Lella VSR, Babu VR and Neeraja CN. 2020. Assessment of genetic variability for micronutrient content and agro-morphological traits in rice (*Oryza sativa* L.). *Indian Journal of Genetics*, 80: 130-139.

Nachimuthu VV, Robin S, Sudhakar D, Raveendran M, Rajeswari S, Manonmani S. 2014. Evaluation of rice genetic diversity and variability in a population by principal component analysis. *Indian Journal of Science and Technology*, 7: 1555-1562.

Nagaraju J, Kathirvel M, Kumar RR, Siddiq EA and Hasnain SE. 2002. Genetic analysis of traditional and evolved Basmati and Non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proceedings of the National Academy of Sciences,* U. S. A. 99: 5836-5841.

Ojha GC, Sarawgi AK, Sharma B, Parikh M. 2017. Principal component analysis of morpho-physiological traits in rice germplasm accessions (*Oryza sativa* L.) under rainfed condition. *International Journal of Chemical Studies*, 5:1875-1878.

Prabhu S, Ganesan NM, Jeyaprakash P, Selvakumar R and Prabhakaran NK. 2017. Assessment of genetic variability studies among the backcross populations in Rice (*Oryza sativa* L.). *International Journal of Pure & Applied Bioscience,* 5:368-372.

Sao R, Saxena RR, Sahu PK. 2019. Assessment of genetic variation and diversity in rice genotypes based on principal component analysis. *Journal of Plant Development Sciences*, 11:725-730.

Smriti, Deepti, Chattopadhyay T, Satyendra, Singh SP and Kumar M. 2016. Characterization and purity analysis of a few aromatic indigenous rice genotypes of Bihar using an aroma-specific marker. *International Journal of Environmental Science and Technology,* 5: 1167-1176.

Tannidi S, Govintharaj P, Ameenal M, Manonmani S and Robin S. 2016. Improving blast resistance in parental line of rice hybrid through marker assisted selection. *International Journal of Agricultural Science Research,* 6: 339-346.

Yan W and Rajcan I. 2002. Biplot analysis of test sites and trait relations of Soybean in Ontario. *Crop Science,* 42:11-20.

Yan W, Frégeau-Reid J. 2018. Genotype by Yield*Trait (GYT) Biplot: A Novel Approach for Genotype Selection based on Multiple Traits. *Scientific Reports*, 8:8242. doi: 10.1038/s41598-018-26688-8.