Indian Journal of Agricultural Sciences **93** (2): 145–150, February 2023/Article
https://doi.org/10.56093/ijas.v93i2.132052

Genetic diversity and principal component analysis in cultivated rice *(Oryza sativa)* varieties of Assam

KANGKANA THAKUR[1*] and M K SARMA[2]

Biswanath College of Agriculture, Assam Agricultural University, Biswanath Chariali, Assam 784 176, India

Received: 5 January 2023; Accepetd: 28 January 2023

ABSTRACT

Rice (*Oryza sativa* L.) is the pre-eminent cereal crop of Assam in term of production and cultivable area. The measurement of genetic diversity is crucial to pick up the appropriate genotypes in a crop improvement programme. Identification of traits contributing towards genetic diversity in a given population can help in formulating effective selection criteria. This study was carried out in 64 indigenous rice genotypes collected from North Bank Plains and Upper Brahmaputra valley zones of Assam during rainy (*kharif)* seasons of 2021 and 2022. Based on the yield attributing characters, the cultivars were categorized into 8 clusters. Cluster evaluation showed that cluster II and V had farthest distance from cluster V (431.18), followed by cluster I and II (377.62). The study on the Principal component analysis showed the traits, days to 50% flowering (DF), days to maturity (DM), filled grains per panicle (FG), biological yield (BY) and harvest index (HI) had maximum contribution towards the total variability. The genotypes from the diverse clusters having desirable *per se* performance could be selected for inclusion as parents in hybridization programme. Identification of component traits through the principal component analysis will help in formulating efficient selection criteria for further genetic improvement of rice.

Keywords: Assam, Cultivated rice, Genetic diversity, PCA

Rice (*Oryza sativa* L.) is the most prominent edible cereal for half of the world's population which feeds around 90% people of south-east Asia. India has been recognized globally as a leading producer of rice, being second in position, with productivity of 2700 kg/ha (Anonymous 2021). India is a reservoir of large array of cultivated and wild rice genotypes grown in diverse conditions of altitude and climate. Out of the vast territorial area, the state of Assam is the home to a large spectrum of diverse land races which possesses many economically important traits including yield and adaptations (Parasar *et al.* 2017). Varietal development is a dynamic process and therefore, needs a constant supply of gene sources for an effective breeding programme. Moreover the traditional cultivars being grown for long undergo dynamic changes in genotypes over times leading to generation of new variability. For any crop improvement programme, the information on the nature and magnitude of genetic diversity is important. Also, it is important to understand the relative contribution and association of the component attributes on the grain yield in order to furnish a reliable guide for selection (Singh and Choudhary 1988). Cluster analysis and PCA are the reliable methods in this regard (Tiwari *et al.* 2020). Cluster analysis helps in grouping of genotypes on the basis of taking D^2 values as the measure of generalized distance between two closely associated genotypes (Rao 1952). Multivariate analysis with the D^2 technique estimates the nature of genetic diversity in a given population with respect to several traits (Sarma *et al.* 2021). Principal component analysis helps to pick up the component traits based on their relative importance contributing towards the total variability in a data set where PC1 describes the most variation, PC2 describes the second most variation and so on (Gupta and Khandelwal 2022). In the present investigation, therefore, it was attempted to assess the genetic diversity and to dissect out the traits contributing the grain yield.

MATERIALS AND METHODS

The present study was carried out at the experimental field of the Advanced Level Biotech Hub, Biswanath College of Agriculture, Assam Agricultural University, Biswanath Chariali, Assam during rainy (*kharif*) seasons of 2021 and 2022 using a set of 64 cultivated rice varieties of Assam. The experiment was laid down in a randomized block design with 2 replications. As per standard package of practice, the crop was raised under transplanted condition. Each plot consisted of 5 rows of 6 meters in length with a spacing

---

[1]Gauhati University, Guwahati, Assam; [2]Biswanath College of Agriculture, Assam Agricultural University, Biswanath Chariali, Assam. *Corresponding author email: kangkana.thakur@gmail.com

of 25 cm × 20 cm. Observations on 13 quantitative traits were recorded as per the standard evaluation system of International Rice Research Institute. The observations on all the traits except for days to 50% flowering and maturity were recorded based on 5 randomly taken plants per plot. Days to flowering and maturity were recorded on plot basis. The pooled data over two years were subjected to ANOVA following Panse and Sukhatme (1967). Further, the mean sum of squares, both year wise and pooled over years were subjected to estimation of genetic coefficient of variation as per Singh and Choudhury (1988). The mean data set pooled over the years was subjected to diversity analysis using Mahalanobis' D^2 Analysis (Mahalanobis 1936) and clustering was done by Tocher's method following Rao (1952) and single linkage rule (Sneath and Socal 1973). Further Principal component analysis was carried out following Singh and Choudhury (1988). Both the analysis were carried out using INDOSTAT software.

RESULTS AND DISCUSSION

The analysis of variance indicated significant variability with respect to all the 13 traits under study. The individual

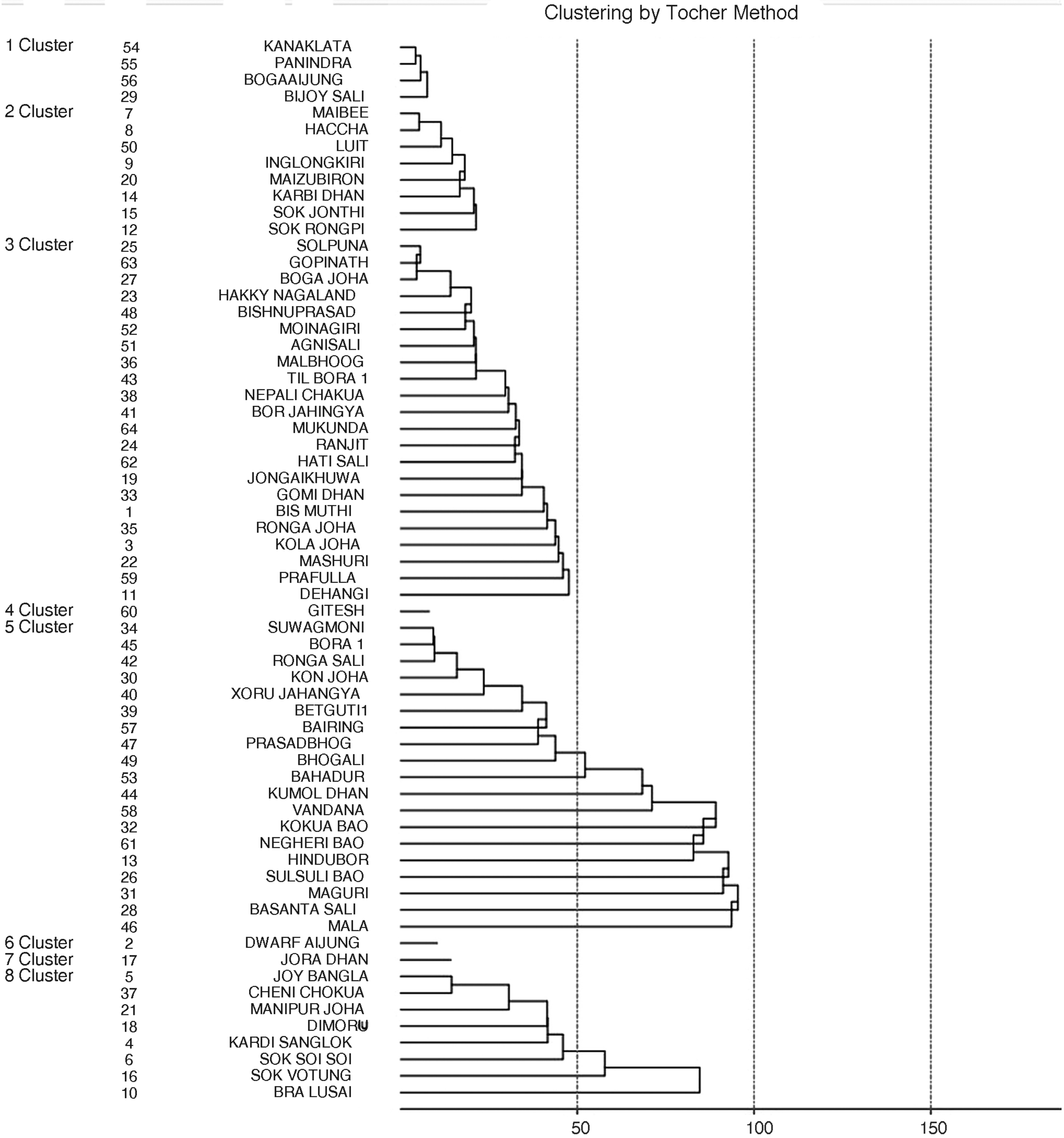

Fig 1 Dendogram showing relationship among 64 rice genotypes in 8 clusters.

Table 1 Genetic variation and cluster mean values in 64 rice cultivars for 13 quantitative traits

Genetic parameter	Year	DF	DM	PH (cm)	PL (cm)	ET per plant	FG per panicle	GW (g)	GL (mm)	GB (mm)	Grain L:B	BY (g)	HI (%)	GY (g) per plant
Mean	2021	119.3	151.45	151	26.3	9.9	127.8	2.19	8.35	2.96	2.82	100.34	20.02	25.13
	2022	116.64	147.8	139.76	25.56	10.12	122	2.15	8.11	2.9	2.97	97.26	21.97	22.15
	Mean over years	117.97	149.67	145.38	25.93	10.01	124.9	2.17	8.23	2.93	2.88	98.80	28.47	23.64
Range	2021	69-152	105-191	89.72-204.9	17.3-34.26	5.1-15.4	45.8-258.34	1.6-3.17	7.2-9.89	2.12-4.46	2.14-4.93	19.12-249.7	7.67-56.98	7.23-53.62
	2022	72-154	103-188	91.76-202.3	18.3-33.1	4.8-14.2	44.9-257.44	1.3-3.19	7.1-9.67	1.96-4.38	2.09-4.51	19.049-250.73	7.234-54.67	7.88-51.5
GCV	2021	16.2	12.15	20.03	11.46	17.2	33.6	16.65	6.12	15.2	15.02	46.34	41.3	40.21
	2022	14.76	11.47	16.25	9.82	13.02	31.78	14.33	5.9	13.4	14.82	45.16	37.5	38.35
	Mean over the years	15.48	11.81	18.14	10.64	15.11	32.69	15.49	6.01	14.3	14.92	45.75	39.40	39.28
CV (%)	2021	1.76	1.57	8.68	4.34	12.0	6.03	0.13	0.127	0.12	0.14	13.5	7.32	5.35
	2022	2.1	1.82	9.54	5.35	11.76	7.54	0.15	0.16	0.14	0.13	14.3	9.3	7.2
Cluster mean values for 13 quantitative traits	Cluster I	131.17	160.8	120.51	25.32	9.35	114.31	2.1	8.24	2.86	2.90	142.20	14.15	21.29
	Cluster II	86.83	120.1	147.82	26.20	9.53	92.44	2.05	8.27	2.81	2.98	55.33	32.33	20.46
	Cluster III	118.67	150.7	151.50	25.87	10.28	133.46	2.2	8.3	2.94	2.88	93.93	28.49	25.34
	Cluster IV	122.67	154.7	116.57	24.08	11.33	124.23	2.17	7.85	2.57	3.07	46.06	48.51	35.56
	Cluster V	130.23	161.6	144.78	25.51	10.22	121.63	2.14	8.1	2.86	2.88	136.26	22.64	23.64
	Cluster VI	130.33	161.7	108.47	27.19	9.76	159.89	1.91	8.28	2.78	2.98	54.75	45.58	24.39
	Cluster VII	96.33	132	137.25	29.03	6.67	84.72	2.15	8.67	2.00	4.34	54.87	22.14	11.51
	Cluster VIII	110.33	142.7	149.25	26.83	9.87	147.68	2.35	8.53	3.39	2.56	62.55	41.70	28.44

NB, Bold figures indicate the extreme values (highest and lowest) of cluster mean for different traits. DF, days to 50% flowering; DM, days to maturity; PH, plant height; PL, panicle length; ET, effective tillers per plant; FG, filled grains per panicle; GW, grain weight; GY, grain yield per plant; GL, grain length; GB, grain breadth; BY, biological yield; and HI, harvest.

year wise as well as average of years for Mean, Range, genotypic coefficient of variation (GCV) and CV (error) are presented in Table 1. The highest value for GCV was observed for biological yield (45.75), followed by harvest index (39.40) and grain yield per plant (39.28). The presence of significant variations indicated the worth of the materials under study for further diversity and PCA analysis. The values of CV being within the limit of 20% indicated that the experimental error was at minimum. The results of Genetic diversity and Principal component analysis are presented in Table 1, 2 and 3.

Cluster analysis: The wider the genetic distance between the parents, wider is the variability generated in the segregating generation (Sarma *et al.* 2015). Therefore, in order to generate variability, genetic diversity of the parents is of foremost importance. In this investigation, the 64 genotypes were grouped into 8 clusters with cluster III having the highest entries of 22, followed by cluster V with 19 entries, cluster II and VIII with 8 entries each, cluster I with 4 entries (Fig 1). Gitesh, Dwarf Aijong and Jora Dhan remained distinct and were placed alone in cluster IV, VI, VII, respectively (Table 2). Cluster analysis to study intra and inter-cluster distances between groups is an effective tool used for classifying genotypes in plant breeding (Latif *et al.* 2011). The intra and inter-cluster distance amongst the 8 clusters are also presented in Table 2 and Fig 2. The highest intra-cluster D^2-value was recorded in cluster V (213.3), followed by cluster VIII (174.96), and cluster III (103.28). Intra-cluster distance was observed zero for the mono-genotypic clusters, IV, VI and VII. The highest intra-cluster distance in cluster V indicated the presence of relatively wide variation amongst the entries within it. Such type of clustering was used by many workers with respect to morphological as well as molecular variation in various crops including rice (Mohapatra *et al.* 1993, Sarma *et al.* 2015, Parasar *et al.* 2017, Sarma *et al.* 2019).

Highest inter cluster distance was observed between cluster II and cluster V (431.18), followed by, between cluster I and II (377.62) and cluster V and VIII (372.08). The closest clusters observed were cluster IV and VI (27.99). Three cultivars, viz. Gitesh, Dwarf Aijong and Jora Dhan identified separately in clusters IV, VI, VII, respectively, is of special significance. These cultivars could be crossed with other entries belonging to other diverse clusters based on specific per se performances. The total diversity deciphered into inter se distance between various clusters would have implication in rice breeding. As the range of variability generated in a segregating generation depends on the genetic distance of the parents, the observed diversity pattern would be of values to the rice breeders as also indicated by other researchers earlier (Khan *et al.* 2008, Rabbani *et al.* 2008).

The cluster mean values for the 13 traits under study are presented in Table 1. A wide range of cluster mean values in 8 clusters indicated the large magnitude of diversity amongst the genotypes. Mean value for DF was highest in cluster I (131.17) and lowest in cluster II (86.83). DM was highest in clusters VI (161.67) and V (161.57) and lowest in cluster II (120.08). Mean value for PH was highest in cluster III (151.5) and lowest in cluster VI (108). PL had the highest mean value for cluster VII (29.03) and lowest

Table 2 Intra and inter-cluster distances for 64 rice genotype along with the names of rice varieties

Cluster	Name of rice varieties	Clusters							
		I	II	III	IV	V	VI	VII	VIII
Cluster I	Kanaklata, Boga Aijung, Panindra, Bijoy Sali	23.1	377.65	144.23	299.83	120.11	272.43	310.99	336.78
Cluster II	Maibee, Haccha, Luit, Inglonkiri, Maizubiron, Karbi dhan, Sok Jongthi, Sok rongpi	377.65	57.73	209.48	221.09	431.18	277.81	79.79	197.52
Cluster III	Solpuna, Gopinath, Boga Joha, Hakky Nagaland, Bishnuprasad, Moinagiri, Agnisali, Malbhoog, Til bora 1, Nepali Chakua, Bor jahingya, Mukunda, Ranjit, Hatisali, Joingaikhua, Gomidhan, Bismuthi, Ronga joha, kola joha, Mashuri, Prafulla, Dehangi	144.23	209.48	103.28	145.09	204.68	140.35	191.64	174.24
Cluster IV	Gitesh	299.83	221.09	145.09	0	317.71	27.99	221.93	152.29
Cluster V	Suagmoni, Bora 1, Ronga Sali, Konjoha, Xoru jahingya, Betguti, Bairing, Prasadbhog, Bhogali,Bahadur, Kumol dhan, Vandana, Kokua Bao, Negheri Bao, Hindubor, Sulsuli bao, Maguri, Basanta Sali, Mala,	120.11	431.18	204.68	317.71	213.36	301.09	376.86	372.08
Cluster VI	Dwarf Aijung	272.43	277.81	140.35	27.99	301.09	0	267.87	151
Cluster VII	Joradhan	310.99	79.79	191.64	221.93	376.86	267.87	0	223.8
Cluster VIII	Joy bangla, Cheni chokua, Manipuri Joha, Dimoru, Kardi sanglok, Sok soi soi, Sok Votung, Bra lusai	336.78	197.52	174.24	152.29	372.08	151	223.8	174.96

NB: The diagonal values are the intra-cluster values. All other values are inter- cluster values.

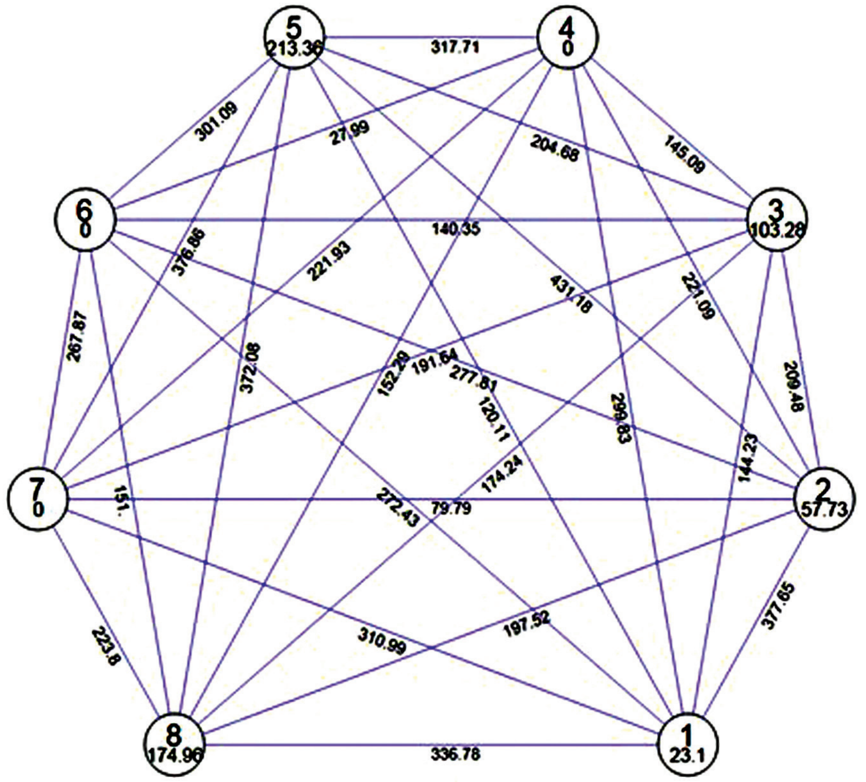

Fig 2 Intra and inter-cluster distance amongst 8 clusters formed out of 64 rice cultivars.

for cluster IV (24.08). The range of mean for ET ranged from 11.3 in cluster IV to 6.67 in cluster VII. FG was highest for cluster VI (159.89) and lowest for cluster VII (84.72). GW observed highest in cluster VIII (2.35) and lowest in cluster VI (1.19). Highest GL was in cluster VII (8.67) and lowest in cluster IV (7.85). GB was highest in cluster VIII (3.39) and lowest in cluster VII (2.00). L: B ratio was found highest in cluster VII (4.34) and lowest in cluster VIII (2.56). Cluster I had highest mean for BY (142.2) with lowest mean for HI (14.15). Lowest mean for BY was in cluster IV (46.06) with highest means for HI (48.51) and GY (35.56). It was also observed that the distribution of highest and lowest mean values of traits in different clusters was also quite diverse. This could be a reliable guide in picking appropriate genotypes to complement each other for a hybridization programme. Cluster analysis to study the diversity between groups is an effective tool used for classifying genotypes for plant improvement programmes. Ahmedikah *et al.* (2008), Yadav *et al.* (2011), Anandan *et al.* (2011), Chakma *et al.* (2012), and Shrestha *et al.* (2021) also made similar investigations in order to cluster various groups of rice genotypes and elucidated the values of them in the planning hybridization programme for rice improvement.

Principal component analysis: Principal component analysis (PCA) was used to find out the relative contribution of the traits towards the total variability and to provide a guide for selection of traits. The main focus of PCA was to extract the total variation due to the studied variables into a limited number of factors. The component traits, chiefly responsible for the extracted PCs, were then delineated. The estimated Eigen vector values, percentage of variance and cumulative percentage are presented in Table 3. The first three PCs accounted for the major portion of variability i.e. 80.75%. Thus, the traits loaded in the three PCs appeared to be of more importance. The first PC contributed maximum towards variability i.e. 43.54%. Characters, viz. BY (0.701), DF (0.446), DM (0.424), ET (0.084), and FG (0.117), were positively loaded while GL (-0.16), HI (-0.244), and GB (-.0428) was negatively loaded. The second PC accounted for 25.12% of variability. The major traits having positive loading in this PC were DF (0.50), DM (0.49), and HI (0.28). However, considerable negative loading was observed by BY (-0.60). The third PC accounted for 12.09% of the total variations. DF, DM, PH and ET were positively loaded while the rest of 13 traits were negatively loaded. Negative loading of FG (-0.78), GW (-0.51) and HI (0.15)

Table 3 Principal components extracted with Eigen values, percentage of variance explained and factor loading of different traits (Based on data over two years).

Eigen value (Root)		PCA I 3005.43200	PCA II 1734.36400	PCA III 834.88100	PCA IV 312.66340
Per cent Var. Exp.		43.53562	25.12338	12.09379	4.52913
Cum. Var. Exp.		43.53562	68.65900	80.75279	85.28192
Factor loading for different traits on the PCs	DF	0.44604	0.50143	0.05098	0.08800
	DM	0.42493	0.48590	0.03103	0.10259
	PH	-0.00421	0.01543	0.00829	-0.11098
	PL	-0.05640	-0.02711	-0.00946	-0.06106
	ET	0.08466	0.10003	0.02209	0.02862
	FG	0.11748	0.12675	-0.78234	0.02608
	GW	0.00765	0.00119	-0.19513	-0.23122
	GL	-0.16368	-0.14761	-0.22395	0.28703
	GB	-0.04280	-0.04210	-0.51434	0.21752
	L:B	-0.11491	-0.12327	-0.04280	0.09352
	BY	0.70321	-0.60716	-0.06415	-0.31618
	HI	-0.24461	0.27906	-0.15153	-0.81623
	GY	0.00965	0.01211	-0.04530	-0.07789

DF, days to 50% flowering; DM, days to maturity; PH, plant height; PL, panicle length; ET, effective tillers per plant; FG, filled grains per panicle; GW, grain weight; GY, grain yield per plant; GL, grain length; GB, grain breadth; BY, biological yield; and HI, harvest.

was prominent in this axis. PCA clearly showed that DF, DM, FG, GW, BY and HI were the most important traits showing a strong effect on total variation. The distinguished characters coming together in different principal components and contributing towards variability have the tendency to remain interrelated (Sinha and Mishra 2013). Thus, this elucidation of the relative merit of the traits could be a reliable guide to formulate an effective selection strategy for further rice improvement programme.

The present study elucidated the worth of a set of 64 rice genotypes of Assam in terms of genetic diversity and the dissection of component traits contributing towards variability. Variability is the prerequisite in breeding programme. On the other hand, selection to be effective, the traits must contribute towards the variability. Therefore for a trait based selection, elucidation of the relative merits of component traits would be meaningful. The present study grouped the 64 rice cultivars of Assam into 8 diverse clusters and delineated few important traits contributing towards total variability. This could be the basis for planning an effective hybridization programme and selection of genotypes for further rice improvement programme.

REFERENCES

Anonymous. 2021. Annual Report, ICAR-National Rice Research Institute, Cuttack, India.

Anandan A, Eswaran R and Prakash M. 2011. Diversity in rice genotypes under salt affected soil based on multivariate analysis. *Pertanika Journal of Tropical and Agricultural Science* **34**(1): 33–40.

Ahmadikhah A, Nasrollanejad S and Alishah O. 2008. Quantitative studies for investigating variation and its effect on heterosis of rice. *International Journal of Plant Production* **2**: 297–08.

Chakma S P, Huq H, Mahmud F and Husna A. 2012. Genetic diversity analysis in rice (*Oryza sativa* L.). *Bangladesh Journal of Plant Breeding and Genetics* **25**(1): 31–39.

Gupta D and Khandelwal V. 2022. Principal component analysis for yield and its attributing characters of pearl millet (*Pennisetum glaucum* [L.] R.Br.). *Annals of Plant and Soil Research* **24**(3): 408–14.

Khan A A, Nasrollanejad S and Alishah O. 2008. Quantitative studies for investigating variations and its effect on heterosis of rice. *International Journal of Plant Production* **2**(4): 297–07.

Latif M A, Rahman M M, Kabir M S, Ali M A, Islam M T and Rafii M Y. 2011. Genetic diversity analyzed by quantitative traits among rice (*Oryza sativa* L.) genotypes resistant to blast disease. *African Journal of Micobial Research* **5**(25): 4483–91.

Mahalanobis P C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Science of India* **2**: 49–55.

Mohapatra K C, Mishra H P, Mishra P K and Acharya B. 1993. Genetic diversity in mutants of upland rice. *Oryza* **30**: 100–05.

Panse V G and Sukhatme P V. 1967. *Statistical Methods for Agricultural Workers*. ICAR, New Delhi, India.

Parasar N, Bairagi P, Sarma M K, Bora R and Baruah S. 2017. Genetic variability and diversity in rice (*Oryza sativa* L.) under rainfed upland situation of Assam. *Contemporary Research in India* **7**(3): 214–18.

Rabbani M A, Pervaiz Z H and Masood M S. 2008. Genetic diversity analysis of traditional and improved cultivars of Pakistani rice (*Oryza sativa* L.) using RAPD markers. *Electronic Journal of Biotechnology* **11**(3): 1–10.

Rao C R. 1952. *Advanced Statistical Methods in Biometrical Research*. Wiley and Sons, New York.

Sarma M K, Goswami R K, Baruah M, Sarma D and Neog P. 2015. Genetic variability and diversity in indigenous rice germplasm of Assam under aerobic condition. *Progressive Agriculture* **15**(1): 66–70.

Sarma M K, Sharma A A, Ramchiary N, Baruah S and Bora P. 2019. Assessment of genetic diversity and inheritance studies for pre-harvest sprouting resistance in soybean (*Glycine max* L. Merr.). (*In*) *Proceedings of International conference on Molecular Biology and Genetic Engineering (Allied Academies, London, UK),* Melbourne, Australia, November 7–8.

Sarma M K, Ahmed N A S, Saharia D D, Nath M K, Talukdar J, Devi S N, Deka S D, Zaman S and Bhuyan J. 2021. Pattern of genetic diversity in indigenous *Ahu* rice germplasm of Assam. *Electronic Journal of Plant Breeding* **13**(1): 21–27.

Singh R K and Chaudhary B D. 1988. *Biometrical Methods in Quantitative Genetic Analysis*. Kalyani Publishers, New Delhi.

Shrestha J, Subedi S, Kushwaha U K S and Maharjan B. 2021. Evaluation of growth and yield traits in rice genotypes using multivariate analysis. *Heliyon* **7**(1). DOI: 10.1016/j.heliyon.2021.e07940

Sneath P H and Sokal R R. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, 1st Edn. W. H. Freeman, San Francisco.

Sinha K A, and Mishra P K. 2013. Agro-morphological characterization and morphology based genetic diversity analysis of landraces of rice variety (*Oryza sativa* L.) of Bankura district of west Bengal. *Inernational Journal of Current Reasearch* **5**(10): 2764–69.

Tiwari S, Yadav M C, Dikshit N, Yadav M C, Pani D R and 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**: 2037–55.

Yadav V K, Singh Y, Soni S K and Yadav A K. 2011. Genetic divergence analysis in salt tolerance rice (*Oryza sativa* L.) genotypes. *Plant Archives* **11**(2): 593–95.