ISSN (Online): 2091-2609
DOI Prefix: 10.3126/ijasbt

International Journal of Applied Sciences and Biotechnology (IJASBT)

Research Article

Clustering and Principal Component Analysis of Nerica Mutant Rice Lines Growing Under Rainfed Condition

Md. Nuruzzaman[1*], Md. Shohel Rana[2], Aleya Ferdausi[1], Md. Monjurul Huda[3], Lutful Hassan[1], Shamsun Nahar Begum[2]

[1]Department of Genetics and Plant Breeding, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh.
[2]Plant Breeding Division, Bangladesh Institute of Nuclear Agriculture (BINA), Mymensingh-2202, Bangladesh.
[3]Cotton Development Board, Khamabari, Dhaka, Bangladesh

Article Information

Received: 10 August 2019
Revised version received: 04 September 2019
Accepted: 07 September 2019
Published: 24 September 2019

Cite this article as:
M. Nuruzzaman et al. (2019) Int. J. Appl. Sci. Biotechnol. Vol 7(3): 327-334. DOI: 10.3126/ijasbt.v7i3.25703

***Corresponding author**
Md. Nuruzzaman,
Department of Genetics and Plant Breeding, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh.
Email: nuruzzamangpb@bau.edu.bd
Mobile No. (+88) 01733889293

Peer reviewed under authority of IJASBT
© 2019 International Journal of Applied Sciences and Biotechnology

OPEN ACCESS

This is an open access article & it is licensed under a Creative Commons Attribution Non-Commercial 4.0 International (https://creativecommons.org/licenses/by-nc/4.0/)

Abstract

A field experiment was conducted at subtropical region in Bangladesh to assess the contribution of morphological traits to variability in some NERICA mutant rice lines. The experiment was conducted following RCBD with three replications. Thirty-one NERICA rice genotypes (twenty-eight mutant lines along with three parents) of advanced generations were used. Data were collected on twelve morphological traits. The results of the principal component analysis showed that the first four components account for 80% of total variation giving a clear idea of the structure underlying the variables analyzed. This result was also supported by scree test. Cluster analysis using Ward's method classified the thirty-one genotypes into five distinct groups. The maximum inter-cluster distance was observed between clusters indicating the possibility of high heterosis if individuals from these clusters are cross-bred. The results of PCA were closely in line with those of the cluster analysis. These results can now be used by breeders to develop drought tolerant high yielding rice varieties and new breeding protocols for rice improvement.

Keywords: Cluster analysis; Principal component; NERICA; Mutant rice.

Introduction

NERICA is a new drought tolerant rice variety introduced by the ministry of agriculture for growing in drought prone areas of Bangladesh. The term NERICA stands for New Rice for Africa, an extended family of some 3000 siblings. So, NERICA is the product of interspecific hybridization between the cultivated rice species of Africa *(O. glaberrima)* and Asia *(Oryza sativa).* NERICA varieties have high yield potential and short growth cycle. Several of

them possess early vigor during the vegetative growth phase and this is a potentially useful trait for weed competitiveness. They also have higher protein content and amino acid balance than most of the imported rice varieties (WARDA, 2008).

Induced mutation has been used for improving both quantitative as well as qualitative characters (Fardous *et al.*, 2013). Mutants have made it possible to identify critical elements for developing high yield potential varieties exhibiting desirable traits such as semi-dwarfism, early maturity, greater number of panicles plant^{-1} and increased fertility. Continuous evaluation of germplasm should be done to broaden the genetic base of the species and identification of additional genes or alternative source that control a particular trait for use in crop improvement. Genetic variability in crop species should be exploited so as to develop new rice varieties with high stability to resist or tolerate adverse environments and biotic conditions (Gana, 2006). So the present study was performed to evaluate the relativeness of different traits with yield through PCA and clustering.

Materials and Methods

The experiment was conducted in Aus season, 2014 in Bangladesh lying under subtropical region. Thirty-one NERICA rice genotypes (twenty-eight mutant lines along with three parents) of advanced generations were used (Table 1). The seeds of mother NERICA plants were treated with physical (250, 300 and 350 gamma-rays, unit is roentgen, R) mutagens and were grown as M_1 generation. Populations were grown with advanced mutants. Finally, lines were selected from M_4 and M_5 generations for this experiment. Seeds of the rice genotypes were obtained from BINA. Field experiments were conducted following randomized complete block design (RCBD) with three replications. Data were collected on days to flowering (1st, 50%, 80%), days to maturity, plant height, total tillers and effective tillers hill^{-1}, filled and unfilled grains panicle^{-1}, 100-seed weight (g) and yield plant^{-1}(g). The principal component analysis was done by the method suggested by Holland (2008) using software R version 3.0.0. and cluster analysis was done using Ward's method.

Results and Discussion

Cluster Analysis

Depending upon the range of diversity, 31 genotypes were grouped into five clusters (Table 2, Fig. 1). Significant differences among the genotypes for all the traits suggested the presence of variation among the genotypes for all the traits under study. The distribution pattern revealed maximum number of genotypes (12 genotypes) in cluster I whilst cluster V included minimum number of genotypes (2 genotype). Cluster II, III and IV included 6, 8 and 3 genotypes respectively. Limited number of the genotypes i.e. 2 and 3 were grouped in clusters V and IV respectively. The limited number of genotypes in the clusters was probably due to high correlation among most of the traits and duplication effect of the traits included in this study. The inter-cluster distances in all the cases were greater than the intra-cluster distances suggesting wider diversity among the genotypes of the distant groups (Table 3).

Table 1: List of the genotypes used in the experiment

S. N.	Symbol	Genotypes
1	G1	N_4/350/P-2(1)-32-11
2	G2	N_1/300/P-9-9-13
3	G3	N_{10}/300/P-2(1)-11-(1)
4	G4	N_1/350/P-2-2-4
5	G5	N_1/250/P-7-2-1
6	G6	N_{10}/300/P-2(1)-6-11
7	G7	N_4/250/P-2(5)-11-13
8	G8	N_{10}/300/P-3-7-1
9	G9	N_{10}/300/P-3-7-3
10	G10	N_{10}/300/P-5-7-5
11	G11	N_4/250/P-2(6)-26(1,3,4)
12	G12	N_4/300/P-3(4)-10-9
13	G13	N_1/250/P-7-3-12
14	G14	N_{10}/300/P-2-3-5
15	G15	N_1/300/P-9-5-12
16	G16	N_{10}/300/P-2(1)-8
17	G17	N_4/250/P-2(5)-11-10
18	G18	N_4/250/P-9-5-3
19	G19	N_1/250/P-7-3-11
20	G20	N_1/250/P-7-3-15
21	G21	N_1/250/P-7-3-12
22	G22	N_1/350/P-2-3
23	G23	N_1/300/P-9-5
24	G24	N_{10}/300/P-3-7-6
25	G25	N_1/300/P-8-3-3
26	G26	N_1/250/P-6-2-8
27	G27	N_1/250/P-7-3
28	G28	N_1/300/P-9-5-6
29	G29	N_1 Parent
30	G30	N_4 Parent
31	G31	N_{10} Parent

Table 2: Number, percent and name of genotypes in different clusters

Cluster number	Number of genotypes	Percent (%)	Name of genotypes
I	12	38.71	1, 2, 8, 11, 12, 14, 22, 23, 25, 26, 27 and 28
II	6	19.35	3, 4, 6, 9, 10 and 18
III	8	25.81	5, 7, 13, 15, 17, 20, 21 and 31
IV	3	9.68	16, 19 and 24
V	2	6.45	29 and 30

Table 3: Intra and inter cluster distance in 31 rice genotypes

Clusters	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V
I	126.59 (11.25)	273.05 (16.52)	185.12 (13.61)	175.89 (13.26)	468.21 (21.64)
II		130.65 (11.43)	213.14 (14.60)	172.68 (13.14)	398.48 (19.96)
III			122.88 (11.09)	87.94 (9.38)	264.37 (16.26)
IV				26.63 (5.16)	192.97 (13.89)
V					39.92 (6.32)

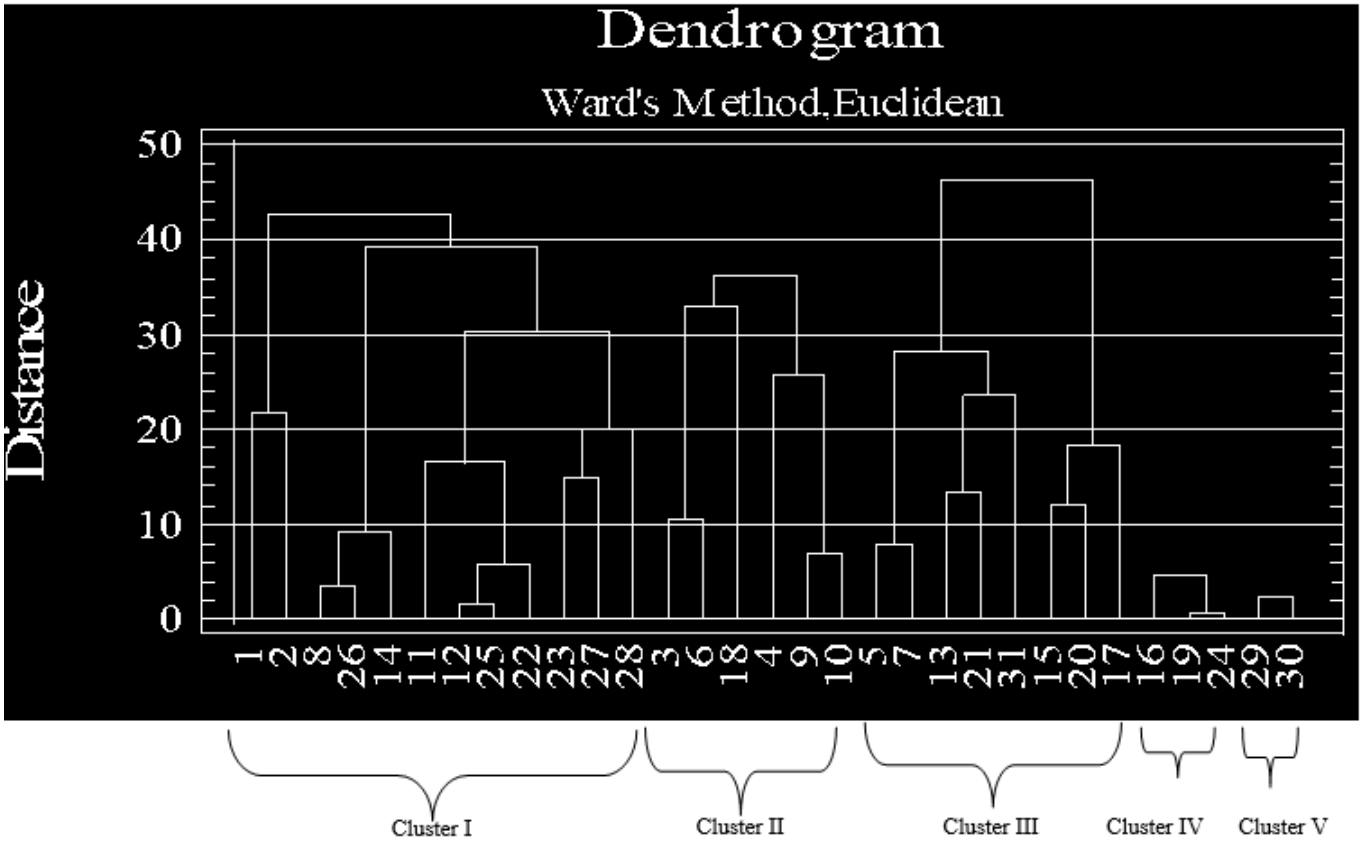

Fig. 1: Dendrogram based on summarized data on differentiation among 31 NERICA Mutant Lines according to Ward's method.

Mean performance of different clusters for different morphological traits studied (Table 4) reflected that all the early flowering (1st, 50% and 80% flowering) genotypes were grouped into cluster II whereas cluster V included all late flowering genotypes. Short duration (108.88 days) genotypes were grouped into cluster I whereas cluster V included long (126.0 days) duration genotypes indicating maximum contribution of this character towards the divergence between clusters I and V. Again all the high yielding genotypes with high number of total tillers hill^{-1}, high number of effective tillers hill^{-1}, maximum panicle length, high number of filled grain panicle^{-1} were grouped

into cluster I whereas cluster V included low yielding genotypes with less number of total tillers hill^{-1}, less number of effective tillers hill^{-1}, less panicle length, less number of filled grain panicle^{-1} indicating maximum contribution of these traits towards the divergence between cluster I and V. The cluster I was divergent from cluster III mainly due to total tillers hill$^{-1,}$ effective tiller hill^{-1} and number of filled grain panicle^{-1} indicating maximum contribution of these traits towards the divergence. The cluster II was divergent from cluster IV mainly due to days to 1st flowering, days to 50% flowering, days to 80% flowering, plant height, panicle length, plant height, total tiller hill^{-1}, effective tillers hill^{-1}, panicle length, filled grain panicle^{-1}, unfilled grain panicle^{-1}, and yield plant^{-1}indicating maximum contribution of these traits towards the divergence. The cluster III was divergent from cluster V mainly due to flowering (1st, 50% and 80%), days to maturity, plant height, filled grains panicle^{-1} and yield plant^{-1} indicating maximum contribution of these traits towards the divergence.

Table 4: Cluster mean for twelve yield and yield contributing characters of 31 rice genotypes

Characters	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V
Days to 1st flowering	78.30	69.63	78.81	80.35	89.45
Days to 50% flowering	83.83	74.36	84.18	85.82	94.44
Days to 80% flowering	89.37	79.66	90.23	90.29	100.75
Days to maturity	108.88	109.08	110.30	109.61	126.00
Plant height (cm)	99.68	91.36	99.06	107.19	76.22
Total tillers hill^{-1}	12.39	10.22	8.75	8.89	9.33
Effective tillers hill^{-1}	10.08	7.83	6.75	7.22	7.33
Panicle length (cm)	25.32	24.12	24.16	30.81	21.14
Filled grain panicle^{-1}	103.47	99.17	90.88	82.55	81.00
Unfilled grain panicle^{-1}	31.56	28.99	30.62	33.56	27.83
100 Seed weight (g)	2.56	2.03	2.76	2.36	2.23
Yield plant^{-1} (g)	25.79	14.82	16.02	13.74	6.40

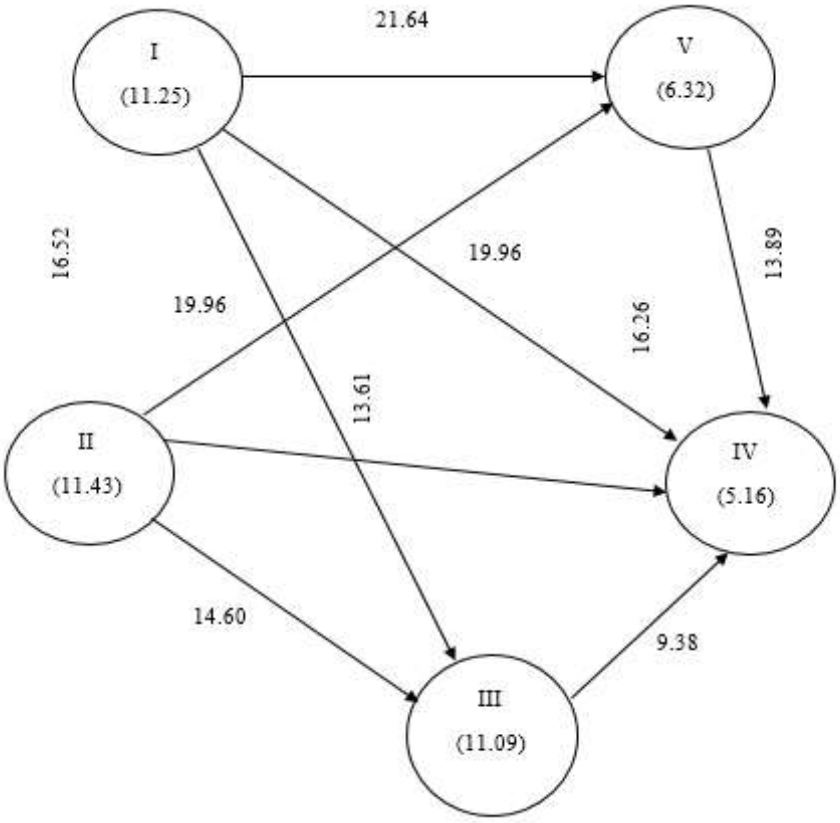

Fig. 2: Cluster diagram showing the average intra and inter cluster distances (D= $\sqrt{D^2}$ values)

Principal Component Analysis (PCA)

The principal component analysis was done by the method suggested by Holland (2008). The result of the PCA explained the genetic diversity of the rice genotypes (Table 5-7). 'Proper values' measure the importance and contribution of each component to total variance, whereas each coefficient of proper vectors indicates the degree of contribution of every original variable with which each principal component is associated. The higher the coefficients, regardless of the direction (positive or negative), the more effective they will be in discriminating between accessions. There are no standard tests to prove significance of proper values and the coefficients. In this study, the criterion corroborated by Guei *et al.*, (2005) was chosen, which suggested that the first four principal components are often the most important in reflecting the variation patterns among genotypes, and the traits associated with these are more useful in differentiating genotypes. According to this criterion, the first four components account for 80% of total variation (Table 5), giving a clear idea of the structure underlying the variables analyzed. This result was also supported by scree test (Figure 3) indicating that the consideration of first four components which account for 80% of total variation. The first component differentiated those genotypes that flowered and mature earlier in the season and identified mainly phonological variables presenting negative contributions. Guei *et al.* (2005) and Sanni *et al.* (2010) observed similar result for days to 50% flowering and days to maturity of rice. The second principal component differentiated the high yielding genotypes (Table 6). The third principal component differentiated the genotypes with good architecture. The fourth principal component accounted for more than 9% of total variance and associated with unfilled grain panicle^{-1}. These findings agree with Caldo *et al.* (1996), who reported that maturity, heading, plant height, culm length, leaf length, and tillering ability were the major factors contributing to the variation of parental lines of modern Philippine rice cultivars.

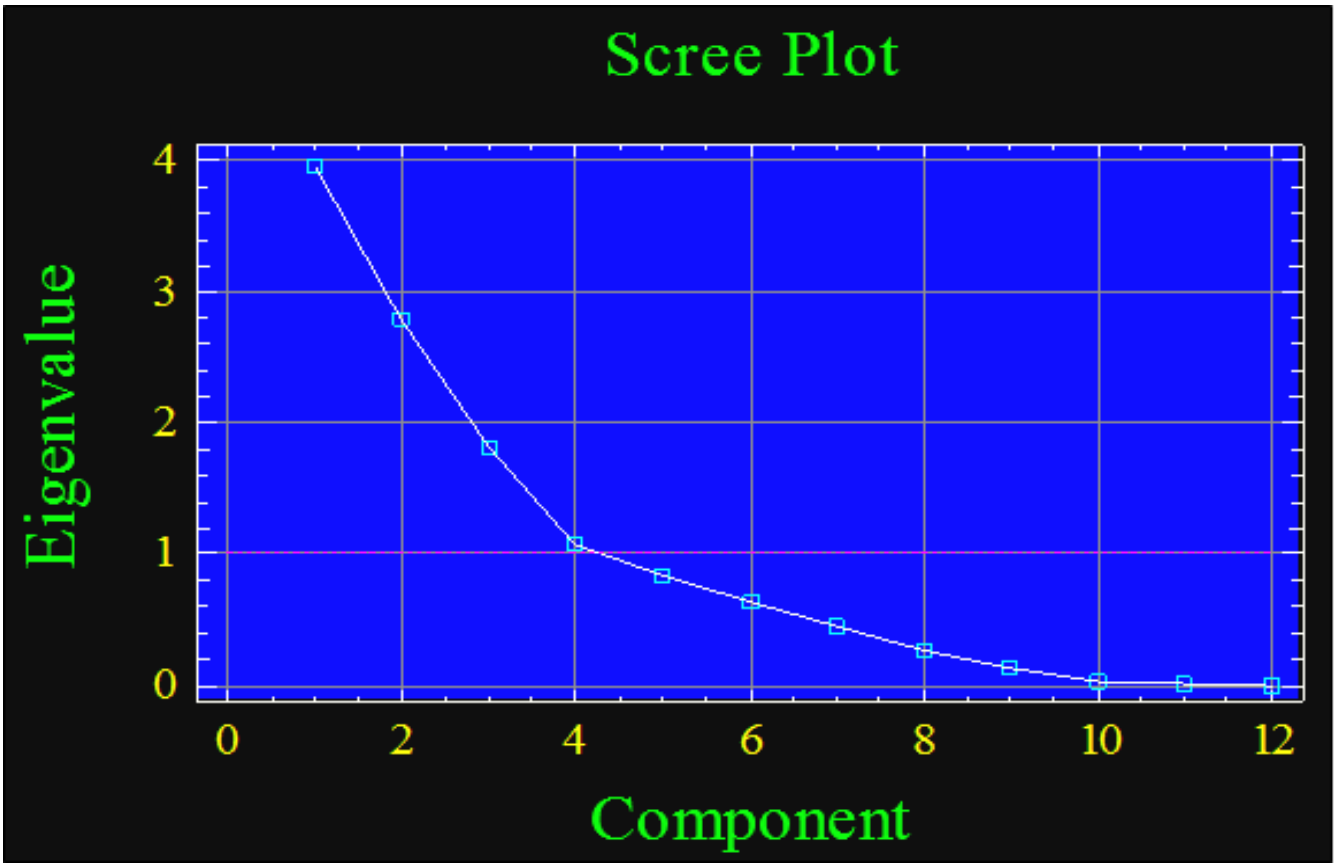

Fig. 3: Scree plot of principal component analysis of 31 genotypes between their Eigen values and the number of principal component.

Table 5: Eigen value, % variance and cumulative (%) total variance of the principal components

Characters	**Eigen value**	**% Variance**	**Cumulative (%) total variance**
Days to 1st flowering	3.95	32.91	32.91
Days to 50% flowering	2.78	23.19	56.10
Days to 80% flowering	1.81	15.11	71.21
Days to maturity	1.08	9.02	80.24
Plant height (cm)	0.837	6.97	87.21
Total tillers hill^{-1}	0.635	5.29	92.50
Effective tillers hill^{-1}	0.449	3.74	96.24
Panicle length (cm)	0.268	2.23	98.47
Filled grain panicle^{-1}	0.132	1.10	99.57
Unfilled grain panicle^{-1}	0.039	0.325	99.90
100 seed weight (g)	0.009	0.076	99.97
Yield plant^{-1} (g)	0.003	0.028	100.00

Table 7: Principal components (PCs) for morphological traits of 31 genotypes under field condition

Characters	**PC1**	**PC2**	**PC3**	**PC4**
Days to 1st flowering	-0.391	0.360	0.011	-0.014
Days to 50% flowering	-0.014	0.377	0.003	-0.005
Days to 80% flowering	-0.006	0.375	0.033	-0.010
Days to maturity	-0.010	-0.131	0.217	0.074
Plant height (cm)	0.074	0.255	-0.428	-0.244
Total tillers hill^{-1}	-0.244	0.286	0.441	-0.015
Effective tillers hill^{-1}	-0.015	0.348	0.406	0.080
Panicle length (cm)	0.080	0.050	-0.498	0.240
Filled grain panicle^{-1}	0.241	0.109	-0.081	0.060
Unfilled grain panicle^{-1}	0.060	-0.003	-0.020	0.926
100 seed weight (g)	0.926	0.332	-0.390	0.014
Yield plant^{-1} (g)	0.015	0.422	0.037	0.091

Table 8: Mean principal components (PC) scores of first four PCs of 31 Genotypes

Genotypes	**PC1**	**PC2**	**PC3**	**PC4**
N_4/350/P-2(1)-32-11	3.998	2.253	1.718	1.460
N_1/300/P-9-9-13	2.000	1.056	1.957	-0.269
N_{10}/300/P-2(1)-11-(1)	3.223	-1.066	1.579	-1.140
N_1/350/P-2-2-4	-0.043	-3.038	1.923	0.099
N_1/250/P-7-2-1	-0.782	0.443	0.543	0.788
N_{10}/300/P-2(1)-6-11	2.496	-2.444	1.642	-1.066
N_4/250/P-2(5)-11-13	-0.648	0.780	0.045	0.274
N_{10}/300/P-3-7-1	1.355	0.221	0.166	2.064
N_{10}/300/P-3-7-3	0.028	-2.813	-0.603	0.103
N_{10}/300/P-5-7-5	-0.358	-0.886	-0.158	-0.618
N_4/250/P-2(6)-26(1,3,4)	0.549	0.881	0.623	-0.971
N_4/300/P-3(4)-10-9	0.099	1.588	-0.417	-0.885
N_1/250/P-7-3-12	-0.945	0.337	-1.327	0.170
N_{10}/300/P-2-3-5	0.260	-0.558	-0.676	0.708
N_1/300/P-9-5-12	-0.074	-0.287	-2.217	-0.701
N_{10}/300/P-2(1)-8	-1.113	-0.671	-2.076	0.983
N_4/250/P-2(5)-11-10	0.657	1.148	-1.742	-1.878
N_4/250/P-9-5-3	2.822	-3.751	-0.639	1.104
N_1/250/P-7-3-11	-0.725	0.016	-1.861	0.064
N_1/250/P-7-13-15	0.003	-1.099	-1.418	-2.337
N_1/250/P-7-13-12	-2.899	-0.171	-1.290	0.214
N_1/350/P-2-3	0.543	0.263	0.550	-1.353
N_1/300/P-9-5	0.733	3.716	0.549	-0.667
N_{10}/300/P-3-7-6	-0.198	0.504	-1.346	0.399
N_1/300/P-8-3-3	0.659	0.754	-0.110	-0.026
N_1/250/P-6-2-8	1.204	0.372	-1.014	1.308
N_1/250/P-7-3	0.828	1.683	-0.505	0.594
N_1/300/P-9-5-6	-1.598	3.095	0.753	0.748
N_1 Parent	-5.012	-0.061	1.795	0.310
N_4 Parent	-4.751	-0.536	2.675	-1.007
N_{10} Parent	-2.311	-1.728	0.881	1.532
SEM	0.357	0.299	0.241	4
P value	0.01	0.01	0.01	1.460

Notes: SEM = Standard error of mean, P = Probability of statistical significance in generalized linear model

Conclusion

The rice genotypes studied showed considerable variability for most of the characteristics that could be exploited for crop improvement. Cluster analysis based on D^2 values and principal component analysis suggested that the genotypes could be grouped into various clusters. Such groupings are useful to breeders in identifying possible genotypes that may be used as parent in future breeding program. Above all, the information generated will save time required by

plant breeders to screen large populations for potential breeding stock.

Author's Contribution

Md. Nuruzzaman, Lutful Hassan and Shamsun Nahar Begum designed the research plan; Md. Nuruzzaman and Md. Shohel Rana performed experimental works & collected the required data. Md. Nuruzzaman and Aleya Ferdausi analysed the data; Md. Nuruzzaman, Md. Shohel Rana and Aleya Ferdausi prepared the manuscript. Critically revised and finalized the manuscript. Final form of manuscript was approved by all authors.

Conflict of Interest

The authors declare that there is no conflict of interest with present publication.

Acknowledgements

The authors are thankful to the Biotechnology Division, Bangladesh Institute of Nuclear Agriculture (BINA), Mymensingh. Financial support from Ministry of Science and Technology, Bangladesh is dully acknowledged.

References

Caldo RA, Sebastian LS, Hernandez JE (1996). Morphology-based genetic diversity analysis of ancestral lines of rice in Philippine rice cultivars. *Philipines journal of crop science,* **21(3):** 86–92.

Fardous AE, Mohammed, Mahrouk E, Magd El-Din, Rida F, Mary N (2013). Effects of Gamma Radiation on Germination, Growth Characteristics and Morphological Variations of *Moluccellalaevis*L. *Am-Euras. Journal of Agricultural and Environmental science*, **13(5)**: 696-704.

Gana AS, Shaba SZ and Tsado EK (2006). Principle component analysis of morphological traits in thirty-nine accessions of rice (*Oryzasativa* L.) grown in a rainfed lowland ecology of Nigeria. *Journal of plant breeding and crop science*, **5(1)**: 120-126.

Guei RG, Sanni KA, Abamu FJ and Fawole I(2005).Genetic diversity of rice (*Oryza sativa* L.). *Journal of AgronomyAfricaine*, **5(1)**: 17–28.

Holland SM (2008) Principal component analysis. Department of Geology, University of Georgia, Athens, 30602-2501.

Sanni KA, Fawole I, Ogunbayo D, Tia EA, Somado K, Futakuchi M, Sié FE, Nwilene and Guei RG (2010). Multivariate analysis of diversity of landrace rice germplasm. Second Africa Rice Congress, Bamako, Mali: Innovation and Partnerships to Realize Africa's Rice Potential, 12-15.

WARDA (West Africa Rice Development Association).Rice Interspecific Hybridization Project: Research Highlights 2008, WARDA, Bouaké, Côted'Ivoire, 36.