The Agriculturists 8(2): 14-21(2010) ISSN-1729-5211
A Scientific Journal of Krishi Foundation

Diversity Analysis in Rice Using GENSTAT and SPSS Programs

M. M. Rahman[1], A. Ansari [2], and M. M. Rashid[1]

[1]*Bangladesh Rice Research Institute Regional Station, Sonagazi, Feni, Bangladesh*
[2]*Plant Breeding Division, Bangladesh Rice Research Institute (BRRI), Gazipur, Bangladesh*

**Corresponding author and Email: monir.brri@yahoo.com*

Received: 07 March 2010 Accepted: 17 March 2011

Abstract

A study was conducted to assess the morpho-physiological divergence among 21 T. Aman rice cultivars at BRRI regional station, Sonagazi, Feni, during July 2008 to December 2008. Data were collected on 13 morphological and 14 physiological traits. Cluster analysis were carried out separately by using two softwares viz. GENSTAT v 5.5 and SPSS v 12.0 where in, both the software divided the cultivars into five clusters in both cases. The resulted clusters seemed to be very similar to some swapping genotypes which indicate that the softwares showed little dissimilarity. When clusterings were carried out using 13 morphological data, multivariate analysis showed five clusters in both GENSTAT and SPSS software with 19.05% swapping genotypes. When multivariate analyses were done with GENSTAT and SPSS softwares based on 14 physiological data, five clusters were also found with 14.29 % swapping genotypes. Of the two programs, GENSTAT appeared to be more reliable than the SPSS program. Inclusion of BR3, BR5, BR11, BR23, BRRI dhan33, BRRI dhan38, BRRI dhan44 and BRRI dhan46 giving emphasis on BRRI dhan33, BRRI dhan38, BRRI dhan44 and BRRI dhan46 is recommended for effective development of a breeding strategy in diallel fashion.

Keywords: Clustering, multivariate analysis, rice, transplanted aman.

1. Introduction

Rice (*Oryza sativa* L.) is a self-pollinated cereal crop belonging to the family Gramineae having chromosome number 2n=24 (Hooker, 1979). It is the staple food for over one third of the world's population (Poehlman and Sleeper, 1995). Bangladesh is the forth largest producer and consumer of rice in the world with annual production of 27.318 million metric tons but needs 2.7% increase in rice production per year due to increasing population (Alam *et al.,* 2004). More than 90% of the world's rice is produced and consumed in Asia (Virmani, 1996). Rice is a all season crop in this country. Among three growing seasons, Aman occupies the highest area coverage (34% of gross cropped area) (Anonymous, 2009). So, we have to give more attention on the improvement of T. Aman rice varieties to increase rice production in order to satisfy our increasing need of food. In order to develop breeding strategy, assessment of variation in T. Aman rice and selection of parents is the prerequisite which could effectively be done using both GENSTAT and SPSS program.

2. Materials and Methods

Genetic diversity in respect of morphological and physiological traits of 20 BRRI developed T. Aman HYVs and one local cultivar named Rajasail was analyzed using GENSTAT 5.5 and SPSS 12.0 software programs. In GENSTAT 5.5 program, clustering was done using non-hierarchical classification using covariance matrix. The genetic divergence between two genotypes was calculated using the following formula proposed by Mahalanobis (1928).

$$pD^2 = W^{ij} (x^{-1}_i - x^{-2}_i) (x^{-1}_j - x^{-2}_j)$$

Where, pD^2 = genetic divergence between two genotypes.

W^{ij} = the inverse of estimated variance and co-variance matrix.

x_i, and x_j = the multiple measurements available on each individual.

Cluster analysis was also done through multivariate analysis under the software SPSS v 12.0 using Euclidean D^2 technique. Genetic divergence among the genotypes was assessed by using Euclidean D^2 technique. The genetic divergence between two genotypes was calculated as:

$$D_{ij}^2 = \sum_{i=j}^{k} (V_{ik} - V_{jk})^2$$

where, D_{ij}^2 = genetic divergence between i^{th} and j^{th} genotypes.

V_{ik} = transformed mean of the i^{th} genotype for the k^{th} character

V_{jk} = transformed mean of the j^{th} genotype for the k^{th} character

To divide the genotypes of a data set into some number of mutually exclusive groups, clustering was done using non-hierarchical classification. The algorithm is used to search for optimum values of chosen criterion. Starting from some initial classification of the genotypes into required number of groups, the algorithm repeatedly transferred genotypes from one group to another as long as such transfer improve the value of the criterion and the algorithm switches to and second stages which examines the effect of swapping two genotypes of different classes and so on. Then clustering as obtained using SPSS v12.0 software was compared with the clustering patterns of Genstat v 5.5 software.

3. Results and Discussion

3.1. Cluster analysis using morphological data

Cluster analysis was carried out with 13 morphological traits viz, plant height, panicle length, maximum number of tillers/m^2, number of effective tillers/m^2, tiller mortality, number of spikelets per panicle, number of effective spikelets per panicle, number of ineffective spikelets per panicle, spikelet fertility, 1000-grain weight, phenotypic acceptability (PACP), straw yield (t/ha), and grain yield (t/ha) using both GENSTAT v 5.5 and SPSS v 12.0 programs. The two-dimensional scattered diagram (Z_1- Z_2) using principal component score I as X-axis and principal component score II as Y-axis was constructed and the position of the genotypes in the scattered diagram was distributed into five groups by using GENSTAT and SPSS software (Fig. 1).

3.2. Cluster analysis using GENSTAT program

By the software GENSTAT v 5.5, clustering was done using covariance matrix where 21 T. Aman rice cultivars were grouped into five clusters (Table 1). The distribution pattern indicates that the maximum number of genotypes (7) were included in cluster V followed by cluster II (5), cluster I (4), cluster III (3) and cluster IV (2).

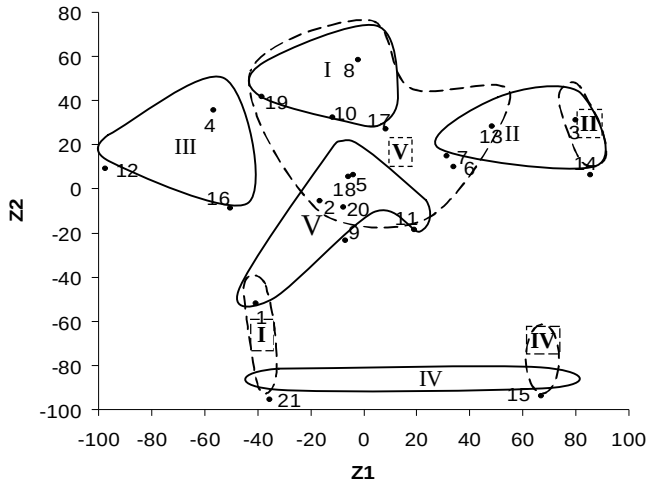

Fig. 1. Scattered diagram of 21 T. Aman rice cultivars based on their principal component scores super imposed in clustering using GENSTAT and SPSS program.

Table 1. Distribution of 21 T. Aman rice cultivars in five clusters through GENSTAT software on the basis of 13 morphological characters

Cluster	No. of variety	Variety
I	4	BR25, BRRI dhan31, BRRI dhan40 and BRRI dhan44
II	5	BR5, BR22, BR23, BRRI dhan34 and BRRI dhan37
III	3	BR10, BRRI dhan33 and BRRI dhan39
IV	2	BRRI dhan38 and Rajasail
V	7	BR3, BR4, BR11, BRRI dhan30, BRRI dhan32, BRRI dhan41 and BRRI dhan46

3.3. Cluster analysis using SPSS program

Cluster analysis was carried out with 13 morphological variables for grouping of 21 T. Aman rice cultivars using SPSS v 12.0 program. The pattern of distribution of genotypes into various clusters is shown in Table 2. The distribution pattern indicates that the maximum number of genotypes (13) were included in cluster V followed by cluster III (3), cluster II (2), cluster I (2) and cluster IV (1).

3.4. Comparison of clustering patterns

Multivariate analysis showed five clusters in both GENSTAT and SPSS software with some dissimilarity. The cluster distribution pattern through the software GENSTAT v 5.5 indicates that cluster I, II, III, IV and cluster V was composed of 4, 5, 3, 2 and 7 genotypes, respectively where as in SPSS v 12.0 software cluster I, II, III, IV and cluster V was composed of 2, 2, 3, 1 and 13 genotypes, respectively. All genotypes of cluster I formed by GENSTAT software were found to be included in cluster V of SPSS software. Out of 5 genotypes in cluster II by GENSTAT, 2 genotypes formed cluster II in SPSS and remaining 3 genotypes found to be

merged in cluster V. Cluster III was identical in both GENSTAT and SPSS and were composed of 3 genotypes (BR10, BRRI dhan33 and BRRI dhan39). Cluster IV of GENSTAT was composed of 2 genotypes of which one formed an individual cluster (cluster IV) in SPSS and the other positioned to the cluster I. All the six genotypes of cluster V in GENSTAT were included in the cluster V except BR3 which was under cluster I in SPSS. Among 21 cultivars, 4 (BR3, BR5, BRRI dhan37 and Rajasail) changed their clustering group when the GENSTAT clustering was compared with the SPSS clustering. So, 19.05 % cultivars swapped in their clustering while clustering is done by SPSS software. It is to be noted that some varieties may change their cluster number in SPSS but if they are grouped in the same cluster, then it was considered to be similar clustering or grouping. The dissimilarities of clustering found in two different softwares might be due to the difference in the degree of sensitivity and calculation of genetic distance method of the software during clustering.

3.5. Comparison in parent selection considering 13 morphological traits

Selection of parents based on genetic divergence has been successfully utilized in different crop species including rice (Gaur *et al*., 1978 and Murty and Anand 1966). Selection of one parent from each cluster and testing them by a series of diallel analysis may prove to be highly fruitful (Chowdhury *et al*., 1975). If we select one genotype from each cluster bearing better traitsis selected, then in case of clustering by GENSTAT v 5.5 software, BRRI dhan44, BR5, BRRI dhan33, BRRI dhan38 and BRRI dhan 46 may be selected from cluster I, cluster II, cluster III, cluster IV and cluster V, respectively. On the other hand, if clustering of SPSS v 12.0 software is considered, then BR 3, BR5, BRRI dhan33, BRRI dhan38 and BRRI dhan46 may be selected from cluster I, cluster II, cluster III, cluster IV and cluster V, respectively. So, only one genotype differs in comparison between these two softwares, in case of selection of five parents from five clusters only. Here, for future breeding program, inclusion of six genotypes considering both the software's clustering and selection followed by diallel cross might be more effective.

Table 2. Distribution of 21 T. Aman rice cultivars in five clusters through SPSS software on the basis of 13 morphological characters

Cluster	No. of variety	Variety
I	2	BR3 and Rajasail
II	2	BR5 and BRRI dhan37
III	3	BR10, BRRI dhan33 and BRRI dhan39
IV	1	BRRI dhan38
V	13	BR4, BR11, BR22, BR23, BR25, BRRI dhan30, BRRI dhan31, BRRI dhan32, BRRI dhan34, BRRI dhan40, BRRI dhan41, BRRI dhan44 and BRRI dhan46

3.6. Cluster analysis using physiological data

Results of cluster analysis with 14 physiological traits viz, seedling vigor (mg/cm), days to flowering (50%), panicle exsertion rate (%), flag leaf area (cm^2), days to maturity, LAI at panicle initiation and at flowering, CGR at panicle initiation and at flowering, RGR at panicle initiation and at flowering, NAR at panicle initiation and at flowering and on harvest index (HI) for grouping of 21 T. Aman rice cultivars using (GENSTAT v 5.5 and SPSS v 12.0) have been presented in Table 3 and Table 4, respectively. The two-dimensional scattered diagram (Z_1- Z_2) using principal component score I as X-axis and principal component score II as Y-axis was constructed and the position of the genotypes in the scattered diagram was distributed into five clusters by using GENSTAT and SPSS software (Fig. 2).

Table 3. Distribution of 21 T. Aman rice cultivars in five clusters based on their physiological characters using GENSTAT software

Cluster	No. of variety	Variety
I	6	BR22, BR23, BRRI dhan30, BRRI dhan 40, BRRI dhan 41 and BRRI dhan 44
II	6	BR4, BR10, BR25, BRRI dhan31, BRRI dhan32 and BRRI dhan 46
III	3	BRRI dhan33, BRRI dhan39 and Rajasail
IV	5	BR3, BR5, BRRI dhan34, BRRI dhan37 and BRRI dhan38
V	1	BR11

Table 4. Distribution of 21 T. Aman rice cultivars in five clusters based on their physiological characters using SPSS software

Cluster	No. of variety	Variety
I	7	BR4, BR10, BR22, BRRI dhan30, BRRI dhan40, BRRI dhan41 and BRRI dhan44
II	9	BR3, BR5, BR25, BRRI dhan31, BRRI dhan32, BRRI dhan34, BRRI dhan37, BRRI dhan38 and BRRI dhan46
III	3	BRRI dhan33, BRRI dhan39 and Rajasail
IV	1	BR23
V	1	BR11

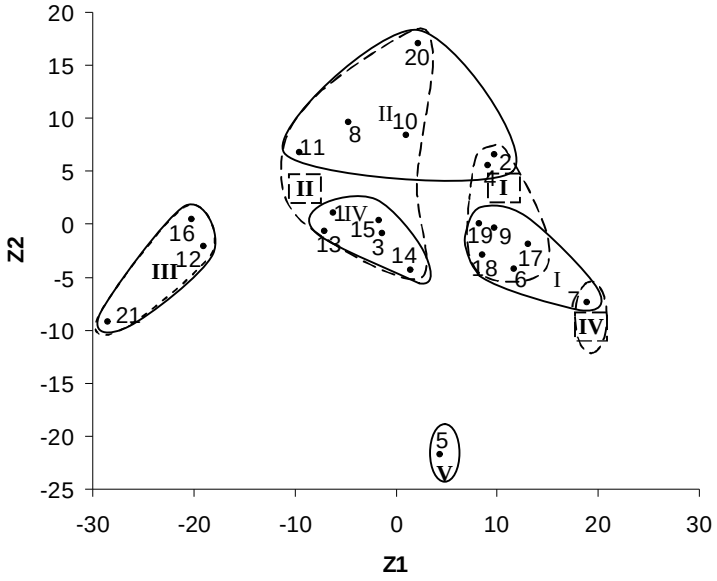

Fig. 2. Scattered diagram of 21 T. Aman rice cultivars based on their principal component scores super imposed in clustering using GENSTAT and SPSS program.

Non-hierarchical clustering using co-variance matrix grouped 21 T. Aman rice cultivars into five clusters based on their 14 physiological traits. The pattern of distribution of genotypes into various clusters is given in Table 3. The distribution pattern indicate that the maximum number of genotypes (6) were included in cluster I followed by cluster IV (5), cluster II (4), cluster III (3) and cluster V (3). Cluster analysis was carried out with 14 physiological traits for grouping of 21 T. Aman rice cultivars using SPSS v 12.0 program. shows that the maximum number of genotypes (9) were included in cluster II followed by cluster I (7), cluster III (3), cluster IV (1) and cluster V (1).

3.7. Comparison of clustering pattern based on two different multivariate analyses

Multivariate analysis showed five clusters in both GENSTAT and SPSS program with some dissimilarity. The cluster distribution pattern through the software GENSTAT v 5.5 indicates that cluster I, II, III, IV and cluster V was composed of 4, 5, 3, 2 and 7 genotypes, respectively where as in SPSS v 12.0 software cluster I, II, III, IV and cluster V was composed of 6, 4, 3, 5 and 3 genotypes, respectively. Among 6 genotypes of cluster I as indicated by GENSTAT software, 5 genotypes were included in the cluster I of SPSS software. The remaining one (BR23) formed a separate cluster in SPSS. All the four cultivars under the cluster II by GENSTAT constituted cluster II in SPSS clustering. Cluster III was identical in case of both GENSTAT and SPSS and was composed of 3 genotypes (BRRI dhan33, BRRI dhan39 and Rajasail). All genotypes (5) of cluster IV indicated by GENSTAT constituted cluster II in SPSS software. Among 3 cultivars under cluster V in GENSTAT, 2 cultivars were merged in cluster I of SPSS and the remaining one (BR11)

formed alone a separate cluster (V). Among 21 cultivars, 3 cultivars (BR23, BRRI dhan40 and BRRI dhan44) changed their clustering group when compared GENSTAT clustering with SPSS clustering. So, 14.29 % cultivars swapped in their clustering while clustering was done by SPSS software. It is to be noted that some varieties may change their cluster number in SPSS but if they was grouped in the same cluster, then the genotypes were considered to be of similar clustering or grouping. The dissimilarities of clustering might be due to the degree of sensitivity of the software used.

3.8. Comparison in parent selection considering 14 physiological traits

Since selection of one parent from each cluster and testing them by a series of diallel cross may prove to be highly fruitful (Chowdhury *et al*., 1975), selection of one genotype from each cluster bearing better traits may be made. In case of clustering by GENSTAT v 5.5 software, BRRI dhan44, BRRI dhan46, BRRI dhan33, BRRI dhan38 and BR11 from cluster I, cluster II, cluster III, cluster IV and cluster V, respectively may be selected. In case of SPSS v 12.0 software, BRRI dhan44, BRRI dhan46, BRRI dhan33, BR23 and BR11 from cluster I, cluster II, cluster III, cluster IV and cluster V, respectively may be selected. Only one genotype differs in case of selection of five parents from five clusters. So, inclusion of all the six genotypes considering both the programs which might be more effective in breeding program.

3.9. Parent selection based on both morphological and physiological traits

Considering both morphological and physiological traits, both GENSTAT and SPSS softwares suggest the selection of BRRI dhan33, BRRI dhan38, BRRI dhan44 and BRRI dhan46 as parent for future breeding program in all the cases. In addition, GENSTAT suggests BR5 and SPSS software suggests BR3 for considering morphological traits, while GENSTAT suggests BR11 and SPSS suggests BR 23 as parent for future breeding program. Inclusion of all the suggested genotypes giving emphasis on BRRI dhan33, BRRI dhan38, BRRI dhan44 and BRRI dhan46 might be more effective to develop a breeding strategy in diallel fashion.

3.10. Comparison of GENSTAT and SPSS programs' efficacy

In GENSTAT program distance between two cultivars were estimated using the formula proposed by Mahalanobis (1928), where as in SPSS program it was estimated using the Euclidean D^2 statistic. So the efficacy of clustering patterns depends on how effectively they estimated distance between the genotypes. In SPSS software, intergenotypic distances are calculated giving equal importance on all the data collected under different variables where as in GENSTAT software these distances are calculated emphasizing on how much a character contribute towards the total diversity. So, in GENSTAT software some characters which have ignorable contribution towards the divergence may be discarded and the remaining characters are prioritized according to their degree of contribution towards the divergence. Moreover, SPSS program is a crude method where the level of significance is relatively higher and is specially suitable for Social Science. On the contrary, GENSTAT program is more sophisticated than SPSS program and its level of significance is relatively lower which provides more accurate result for the field of genetics. So, GENSTAT software is recommended for diversity analysis in the field of genetics.

4. Conclusions

Both GENSTAT and SPSS softwares are used in the field of genetics, specially GENSTAT is extensively used. But if both the softwares are used and if their outcomes are more or less similar, the findings might provide more accurate

findings in this field. However, parent selection based on both the softwares would provide higher probability of finding expected progeny from the segregating generations.

References

Alam, M. M., Buresh, R. J., Ladha, J. K. and Khan, A. H. 2004. Site Specific Nutrient Management for Rice. BRRI, Gazipur, Bangladesh. 1 p.

Anonymous. 2009. Statistical Pocket Book of Bangladesh (Bangladesh Burea of Statistics), Ministry of Planning, Government of the People Republic of Bangladesh. 175 p.

Chowdhury, B. D., Bhatt, P. N. and Singh, V. P. 1975. Genetic diversity in cluster bean. *Indian Journal of Agriculture Science*, 45**:** 530-535.

Gaur, P. C., Gupta, P. K. and Hishore, H. 1978. Studies on genetic dicergence in potato. *Euphytica.* 27:361-368

Hooker, J. D. 1979. The Flora of British India. Vol. 2L. Reeve Co. Kent, England. 25 p.

Mahalanobis, P.C. 1928. On the generalized distance in statistics. *National Institute of Science India* Bull., 2**:** 49–55.

Murty, B. R. and V. Anand. 1966. Combining ability and genetic diversity in some varieties of *Linum usitatissimum*. *Indian Journal of Genetics*, 26**:** 21-36.

Poehlman, J. M. and Sleeper, D. A. 1995. Breeding Field crops. Panima Publishing Corporation. New Delhi, India. 278 p.

Virmani, S. S. 1996. Hybrid rice. *Advanced Agronomy*, 57**:** 379.