Non-hierarchical Euclidean Cluster Analysis in Aromatic Rice

Dharmendra Singh, VK Sharma and VK Shahi
Department of Genetics, Rajendra Agricultural University, Bihar, Pusa, Samastipur-848125 Bihar

The present investigation is an attempt to classify forty-six entries comprising advanced breeding lines, local and exotic collections and cultivars of aromatic rice into well characterized groups. Results revealed significant differences among the entries for seventeen agromorphological and quality characters under consideration in this study. With the exception of days to 50% flowering, days to maturity, panicle length and kernel elongation ratio, all characters had moderate to high estimates for genotypic and phenotypic coefficients of variability. By following the method of non-hierarchical Euclidean cluster analysis, the entries could be divided into five distinct, compact and well characterized groups. The distinctness of clusters was proven by distance matrix as inter-cluster distances were greater than all the intra-cluster distances. Some of the entries showed greater diversity despite having common parentage. Further, the distribution pattern of entries was not necessarily related to geographical diversity. Analysis based on cluster means suggested the importance of plant height, number of grains per panicle, kernel length before and after cooking and volume expansion ratio for selection of parents from distantly related clusters for hybridization. The entries with exceptionally good performance were identified in five clusters for their use as base material.

Key Words: Rice, Aromatic Lines, Cluster Analysis, Genetic Diversity, Grain Quality

Introduction

Presence of wide genetic variation provides a greater scope for genetic improvement and making progress in crop breeding. Although intensive breeding of improved cultivars in recent times has enhanced the availability of natural variability in rice, much emphasis has not been laid on systematic programme of developing semi-dwarf high yielding aromatic varieties with superior quality characteristics. With the consumers becoming more quality conscious, emphasis on quality breeding has assumed greater significance in recent years. Precise information on the nature and extent of genetic divergence and on the relative importance of characters in terms of their contribution to total divergence is crucial in any crop improvement programme as it helps in choosing the parents for hybridization. The method of non-hierarchical Euclidean cluster analysis is a powerful tool in classifying the entries into distinct groups and in discerning divergence among groups based on actual expression of multiple characters. Using this method in the present study, an attempt was made to classify the advanced breeding lines, local and exotic accessions and standard cultivars of aromatic rice into well characterized groups and to quantify the magnitude of genetic divergence for their further use in recombination breeding with expectation of getting potential transgressive segregants.

Materials and Methods

The experimental materials of the present study consisted of fourty-six entries comprising advanced breeding lines, local and exotic collections and cultivars of aromatic rice. The entries were evaluated in randomized complete block design with three replications at the research farm of Rajendra Agricultural University, Pusa, Bihar during *kharif*, 2002. Seedlings of each entry were transplanted to the field in three rows spaced at 20 cm and inter-plant distance within a row was maintained at 10 cm. While raising the crop, recommended agronomic practices were followed. The observations for seventeen agromorphological, physical grain quality and cooking quality characters with the exception of phenological characters were recorded on five plants per entry in each replication and their mean values were subjected to statistical analysis. Genotypic and phenotypic coefficients of variation were estimated as per standard statistical procedure (Johnson *et al.*, 1955). Using a computer program (Doshi and Gupta, 1991), the non-hierarchical Euclidean cluster anlaysis (Beale, 1969; Spark, 1973) was conducted to group the entries into different clusters and to estimate the inter-cluster and intra-cluster distances.

Results and Discussion

Analysis of variance revealed significant differences among the entries for each of the seventeen characters under consideration in the present study, suggesting adequate variability among the genotypes. A wide range of variability was noticed for different agronomic and quality traits. The estimates of genotypic coefficient of variation were found to be lower than

respective phenotypic coefficient of variation. However, a relatively higher magnitude of difference between genotypic and phenotypic coefficient of variation was noticed in respect of number of panicles per plant, uncooked kernel breadth, kernel elongation index and grain yield per plant. While days to 50% flowering, days to maturity, panicle length and kernel elongation ratio had lower estimates, the remaining characters recorded moderate to high genotypic and phenotypic coefficients of variability in the present study.

Using correlation matrix based on mean values of entries, all the metric characters were transformed into a single index of similarity in the form of principal components, which yielded seventeen eigen roots and corresponding eigen vectors. Principal component analysis was used prior to cluster analysis to determine the relative importance of classification variables. Only first ten principal component scores were used further for clustering purpose as these accounted for 96.01 percent variation. Eigen values from the first ten principal component axes accounted for 28.63, 20.78, 13.74, 9.90, 5.63, 5.04, 4.02, 3.85, 2.86 and 2.47 per cent of the total variance present, respectively.

Non-hierarchical Euclidean cluster anlaysis was conducted and a series of cluster solutions with seven, six, five, four and three clusters were obtained. The sequential F-ratio tests for comparison of cluster solutions, however yielded the highest and significant F-value for making comparison between cluster solutions with five and four clusters. This indicated that going from five clusters to four clusters was inadequate to describe the data fully (Katyal *et al.*, 1985). Taking this into consideration, the cluster solution with five clusters was accepted for further study. Thus, fourty-six genotypes were classified into five separate and well characterized groups on the basis of their observed characteristics (Table 1), revealing the presence of considerable amount of genetic diversity in the material. Cluster A was dominated by traditional basmati varieties. The seven entries included in this cluster were characterized by superior grain and cooking quality characteristics. Cluster B was the largest group which consisted of fourteen entries, a majority of them being local collections possessing comparatively inferior grain quality attributes. Most of the entries included in this cluster were advanced breeding lines. Cluster D consisted of twelve entries including indigenous and exotic accessions and advanced breeding lines having better grain quality. Cluster E accommodated seven entries. This cluster was dominated by entries from indigenuous collections and characterized by superior grain and cooking quality attributes. It was interesting to find that some of the entries showed greater diversity as they were accommodated into distinct clusters, despite having common parentage. Further, the pattern of distribution of entries into different clusters was not observed to be necessarily related to geographical diversity as genotypes selected under diverse locations clustered together in consonance with the earlier reports (Singh *et al.*, 1999; Raju *et al.*, 2002; Shukla and Pandey, 2003; Datt and Mani, 2003). The results were in agreement with the earlier observations that the factors like history of selection, genetic drift and selection under diverse environments may cause greater genetic diversity than the geographical distance (Sardana *et al.*, 1997). Thus, it appeared that it is important to ascertain the level of parental divergence, since genetic diversity can not be taken as synonymous with geographical divergence when an initial choice of the parents has to be made in the hybridization programme for getting wide range of variability in segregating generations.

Statistical distances between the clusters represent the extent of diversity amongst the entries which are accommodated into different clusters. As it is evident from a perusal of distance matrix (Table 2), the inter cluser distance was maximum between B and E (5.30) reflecting greatest diversity between genotypes belonging to these two clusters. Similarly, the entries in clusters D and E (5.08) followed by those in clusters B and C (5.07) and clusters A and B (4.74) were much diverse to each other. The inter-cluster distance between A and

Table 1. Distribution of aromatic rice genotypes into five clusters

Cluster A : Type-3, Pakistani Basmati, Taroari Basmati, Katarni, PRR78, Seond Basmati, Basmati Sathi; **Cluster B:** Sugandha, Kamini, BR 9, BR 10, Kalanamak, BR 26, Sona Chur, Kamod-1, Chandan Chur, Shyma Jeera, Mallida, Laxman Bhog, Kariya Kamur, Gamkaua Dhan; **Cluster C:** RPS 1, Champaran Local-1, Pusa 1347-2, Pusa 1280-1, Pusa 1280-4, Pusa 1297-2; **Cluster D:** Kasturi, Champaran Local-2, Champaran Local-6, Basmati 385, RPS local, Kamod 2, BK 843-7, Azucena, Binam, Pusa 1347-3, Pusa 1347-1, Pusa 1280-3; **Cluster E:** Tilak Chandan, Haryana Basmati, RAU-SCT-13, RAU-SCT-14, RPS 2, Champaran Local-1, Champaran Local-4.

Table 2. Estimates of average intra and inter-cluster distances based on non-hierarchical cluster analysis

Cluster	A	B	C	D	E
A	2.72	4.74	3.67	4.49	3.32
B		3.11	5.07	4.06	5.30
C			3.08	3.54	4.30
D				2.73	5.08
E					2.64

Diagonal values represent intra-cluster distances

E (3.317) was smallest. However, it was greater than all the intra-cluster distances indicating that the five clusters obtained by non-hierarchical cluster analysis were distinct and more or less compact. While cluster E with the lowest value for intra cluster distance was most compact, the clusters A and D seemed to be relatively more compact than clusters B and C which had comparatively greater intra cluster diatance. Thus, the genotypes in clusters B and C were relatively more heterogenous, that is, diverse among themselves. The genotypes belonging to the clusters separated by high estimated distance could be utilized in hybridization programme for obtaining wide variation among the segregarits (Singh *et al.*, 1996; Gupta *et al.*, 1999).

The data on cluster wise mean values for different characters also indicated substantial variation among different clusters as well as superiority of different clusters for different characters (Table 3). None of the clusters contained genotypes with all the desirable characters which could be directly selected and utilized. The entries of cluster A appeared to be superior than others in respect of kernel length, kernel L/B ratio, cooked kernel length, cooked kernel L/B ratio and volume expansion ratio. In addition, cluster A also had the second highest man value for 100-grain weight, number of grains per panicle, kernel elongation ratio and kernel elongation index. The entries forming cluster B had more plant height and number of grains per panicle, least 100-grain weight, higher grain yield per plant, delayed flowering behaviour, late maturity and lower values for grain and cooking quality characteristics. The entries of cluster C were characterized by lower mean values for plant height, grain yield per plant, number of panicles per plant and number of grains per panicle, superior grain quality but relatively inferior cooking quality characteristics. The higher mean values for grain yield per plant, 100-grain weight and number of panicles per plant, moderately high plant height and number of grains per panicle, superior grain quality attributes, whereas lower mean values for cooking quality characteristics appeared to be the features of entries which were accommodated in cluster D. Similarly, the lower mean values for days to flowering, days to maturity, number of panicles per plant, number of grains per panicle and grain yield per plant, whereas higher mean values for grain and cooking quality components were the characteristics of the entries which were included in cluster E. Based on cluster means for different characters, plant height, number of grains per panicle, kernel length before and after cooking and vlume expansion ratio appeared to be the main factors responsible for differentiation among the lines evalutated in the present study. The importance of these characters as main contributors to genetic diversity in rice has also been emphasized earlier (Pandey *et al.*, 1999; Raju *et al.*, 2002; Roy *et al.*, 2002; Shukla and Pandey, 2003) supporting the observations that the greatest contributors to genetic diversity in grain crops are flowering time, plant height, primary branches or tiller number per plant (Murty and Arunachalam, 1966).

Considering individual genotypic performance as the basis for selection, three genotypes which registered exceptionally good performance were identified in each

Table 3. Cluster wise mean values for seventeen characters in aromatic lines of rice

Character	A	B	C	D	E
Days to 50% flowering	106.38±2.67	112.31±2.57	108.50±3.27	107.39±3.31	93.29±6.24
Days to maturity	131.67±6.25	138.98±3.67	133.11±5.47	132.67±4.92	122.14±4.21
Panicles per plant	6.03±0.90	6.09±0.96	4.86±1.04	5.71±1.13	4.85±0.98
Panicle length	22.03±1.54	23.84±2.02	20.93±1.72	21.86±1.46	21.53±0.95
Grains per panicle	123.84±47.98	147.13±49.64	92.48±50.65	103.50±31.80	82.70±10.85
100-grain weight	2.13±0.41	1.60±0.37	2.06±0.27	2.30±0.42	2.01±0.38
Grain yield per plant	10.45±2.22	12.84±1.64	8.22±3.83	13.24±1.40	9.88±2.84
Plant height	124.91±27.66	138.39±9.18	96.51±23.76	117.44±24.12	114.19±12.72
Kernel length	7.09±0.47	4.96±0.74	7.04±0.64	6.80±0.60	6.13±0.90
Kernel breadth	1.79±0.16	1.91±0.31	1.78±0.13	2.07±0.26	1.86±0.16
Kernel L/B ratio	4.00±0.40	2.64±0.59	3.82±0.75	3.35±0.50	3.19±0.62
Cooked kernel length	13.46±0.93	8.92±1.59	10.96±1.42	10.60±1.17	11.71±1.39
Cooked kernel breadth	2.29±0.34	2.52±0.39	2.21±0.34	3.01±0.26	2.28±0.31
Cooked kernel L/B ratio	5.92±0.81	3.62±0.83	5.08±1.00	3.58±0.53	5.22±0.83
Kernel elongation ratio	1.90±0.08	1.79±0.15	1.56±0.17	1.57±0.14	1.92±0.08
Kernel elongation index	1.49±0.21	1.38±0.22	1.33±0.12	1.09±0.09	1.67±0.24
Volume expansion ratio	3.82±0.18	3.57±0.30	2.88±0.34	2.99±0.40	3.81±0.19

Table 4. Superior genotypes selected from different clusters

Cluster	Genotypes with desirable characters
A	Type-3, Pakistani Basmati, Taroari Basmati
B	Kamod-1, Gamkaua Dhan, Kariya Kamur
C	Pusa 1347-2, Pusa 1280-4, Pusa 1297-2
D	Binam, Azucena, BK 843-7
E	RAU-SCT-13, RAU-SCT-14, Champaran Local-4

of the five distinct and well characterized clusters (Table 4). These entries can serve as parents to generate desirable segregarits in breeding programmes. The hybridization among genetically diverse parental genotypes may be helpful in bringing the new gene pool in population and producing a wide spectrum of variation in the segregating progeny.

References

Beale, EML (1969) Euclidean cluster analysis. A paper contributed to 37th session of the International Statistical Institute, UK.

Datt S and SC Mani (2003) Genetic divergence in elite genotypes of basmati rice (*Oryza sativa* L.) *Indian J. Genet.* **63:** 73-74.

Doshi SP and KC Gupta (1991) Statistical package for agricultural research data analysis. Indian Agricultural Statistics Research Institute, New Delhi.

Gupta A, RK Sharma, VP Mani and VS Chauhan (1999) Pattern of genetic diversity and variability in rice germplasm of UP Hills. *Crop Improv.* **26**: 81-87.

Johnson HW, HF Robinson and RE Comstock (1955) Estimates of genetic and environmental variability in Soybean. *Agron. J.* **47:** 314-318.

Katyal JC, SP Doshi and PK Malhotra (1985) Use of cluster analysis for classification of benchmark soil samples from India in different micronutrient availability groups. *J. Agric. Sci.* Camb. **104:** 421-424.

Murty BR and V Arunachalam (1966) The nature of divergence in relation to breeding system in some crop plants. *Indian J. Genet.* **26A:** 188-198.

Pandey DK, HS Gupta and PS Pandey (1999) Assessment of divergence in rice grown in iron toxic soil of Meghalaya: *Indian J. Agric. Sci.* **69:** 69-70.

Raju Ch S, MVB Rao, GLK Reddy, JSP Rao and KS Reddy (2002) Genetic divergence in rice for physiological and quality attributes. *Madras Agric. J.* **89:** 474-478.

Roy B, AK Basu and AB Mandal (2002) Genetic diversity in rice (*Oryza sativa* L.) genotypes under humid tropics of Andaman based on grain yield and seed characters. *Indian J. Agric. Sci* **72:** 84-87.

Sardana S, DN Borthakur and TN Lakhanpal (1997) Genetic divergence in rice germplasm of Tripura, *Oryza* **34:** 201-208.

Singh AK, SB singh and SM Singh (1996) Genetic divergence in scented and fine genotypes of rice (*Oryza sativa* L.). *Ann. Agric. Res.* **17:** 163-166.

Singh UK, SB Mishra and R Thakur (1999) Genetic divergence in boro rice. *Oryza.* **36:** 76-77.

Shukla SK and MP Pandey (2003) Genetic divergence of elite indica thermosensitive genic male sterile lines in rice (*Oryza sativa* L.). *Indian J. Genet.* **63:** 71-72.

Spark DN (1973) Euclidean cluster analysis. Algorithm. *Applied Statistics.* **22:** 126-130.