Forage Res., **41** (2) : pp. 73-77 (2015) **http://forageresearch.in**

MULTIVARIATE ANALYSIS FOR SELECTION OF DIVERSE GENOTYPES IN PEARL MILLET GERMPLASM

RAMESH KUMAR, URMIL VERMA*[,1], VIRENDER MALIK AND DEV VART

Bajra Section, Department of Genetics & Plant Breeding
CCS Haryana Agricultural University,
Hisar-125 004 (Haryana), India
(e-mail : vermas21@hotmail.com)

(Received : 4 August 2015; Accepted : 25 September 2015)

SUMMARY

The focus of this work has been to study the genetic divergence among the pearl millet germplasm and grouping them into different clusters based on yield and yield contributing traits for the hybridization programme. The principal component analysis (PCA) revealed that the higher loading displaying variables on first four PCs were : plant height, ear diameter, number of effective tillers/plant and days to 50 per cent flowering. Genetic divergence analysis was performed on the basis of discriminant analysis using Mahalanobis' D^2-statistics. Based on the relative magnitude of D^2-values, 97 genotypes of pearl millet were grouped into six clusters and plant height, ear diameter and grain yield were found the best discriminatory characters for better selection of diverse genotypes.

Key words : Genetic diversity, principal component, cluster distance, D^2-value, cluster mean

In a crop improvement programme, the measurements are taken on several characters because of their inter-relationships, however, a breeder may be interested in selecting only few important characters in which the improvement is needed. A number of statistical procedures have been proposed from time to time for selection of important characters. Step-wise regression analysis and principal component analysis (PCA) can be used by researchers for the purpose. PCA has an edge over the other as it removes multicollinearity among the independent variables. Principal component analysis being a data reduction technique for investigating the interdependence attempts to simplify complex and diverse relationships existing among a set of observed variables, by revealing common dimensions or components that link seemingly unrelated variables.

Cluster analysis is also one of the methods of data reduction technique. It has similarity with discriminant analysis in respect of classification of observations. But discriminant analysis derives a rule for allocating an object to its known proper population based on some prior information of the group membership of the object, whereas the cluster analysis identifies homogeneous groups of clusters. It helps in grouping the materials in such a manner that similar types are grouped together while dissimilar ones belong to different groups. There are two main types of measures used to estimate this relation; distance and similarity measures.

The statistical technique of discriminant function can be used to discriminate between/among various groups of objects when the dependent variable is categorical and independent variables are metric. The objective of discriminant analysis is to classify the sample objects accurately on the basis of a linear combination of predictor variables. Three different methods, namely, (i) Maximum Likelihood Discriminant Rule, (ii) Fisher's (1936) Linear Discriminant Function and (iii) Bayes Discriminant Rule are in common use for identification of the populations.

Pearl millet is a staple diet for the vast majority of poor farmers and also forms an important fodder crop for livestock population in arid and semi-arid regions of the country. Increased emphasis on development of dual purpose (grain-cum-fodder) pearl millet, is, therefore, necessary for ensuring high grain yield as well as higher dry fodder yield under rainfed cultivation (Dangaria and Atara, 2004). The development of an effective breeding programme is dependent upon the

[1]Department of Mathematics, Statistics and Physics.

existence of genetic variability. Pearl millet is endowed with a rich reservoir of genetic variability for various yield components, adaptation and quality traits. Exploitation of the genetic variability in the available germplasm holds promise for producing high grain and fodder yielding hybrids. The more diverse the parents, the greater are the chances of obtaining new combinations of genes and therefore increasing the probability for crop improvement. Several measures have been used to assess genetic diversity among plant populations. Of these measures, multivariate analysis (Fisher, 1936; Jolliffe, 1972; Johnson and Wichern, 2006) provides the most reliable information. To cite a few references, Joshi *et al.* (1988), Vidyadhar *et al*. (2004) and Shanmuganathan *et al.* (2006), etc. have worked on genetic divergence in pearl millet. Among the multivariate procedures, Mahalanobis (1936) generalized distance (D^2) has been used extensively. Keeping in view the importance of the subject matter, an attempt has been made to perform the diversity analysis in pearl millet germplasm for grain yield and its associating characters.

Data Description and Statistical Methodology

The experimental material comprising 97 entries of pearl millet germplasm was grown in Augmented Design during **kharif** 2014 at the research farm, Department of Genetics & Plant Breeding, CCSHAU, Hisar. Each genotype was sown in a plot consisting of two rows of size 3.6 m each with row to row spacing of 45 cm and plant to plant distance as 15 cm. The data were obtained for eight characters viz., days to 50 per cent flowering, plant height (cm), ear length (cm), ear diameter (mm), number of effective tillers/plant, number of nodes/plant, internode length (mm) and grain yield (g/plant). The analysis was carried out to identify the most informative and best discriminating characters in pearl millet. The genetic variability parameters including mean, standard deviation, coefficient of variation (CV%) of yield and yield contributing traits were obtained (Table 1). PC analysis was carried out to identify the most important characters. The procedure consists of finding the eigen roots and eigen vectors of the correlation matrix of explanatory variables. Interpretation of principal components is often facilitated by computing the components loadings. PC loadings measure the importance of each variable in accounting for the variability in the PC. One of the most commonly used criteria for solving the number of components problem is the eigen value-one, also known as the Kaiser's (1960) criterion (Table 2). With the Scree test, we plot the eigen values associated with each component and look for a "break" between the components with relatively large eigen values and those with small eigen values. The components that appear before the break are assumed

TABLE 1
Descriptive statistics of yield and its associated traits

Character	Mean	Std. deviation	C. V. (%)
Days to 50% flowering	51.48	4.49	8.72
Plant height (cm)	173.81	25.36	14.59
No. of effective tillers/plant	3.80	1.13	29.74
Ear length (cm)	20.19	5.46	27.04
Ear diameter (mm)	21.42	9.30	43.41
No. of nodes/plant	8.02	1.26	15.71
Internode length (mm)	17.05	2.71	15.89
Grain yield (g/plant)	16.64	6.84	41.10

TABLE 2
Total variance explained by different principal components

Component (s)	Eigen values	Per cent variance	Per cent cumulative variance
1	1.83	22.92	22.92
2	1.38	17.28	40.20
3	1.19	14.91	55.12
4	1.13	14.19	69.31
5	0.83	10.37	79.69
6	0.74	9.30	88.99
7	0.61	7.65	96.65
8	0.26	3.34	100.00

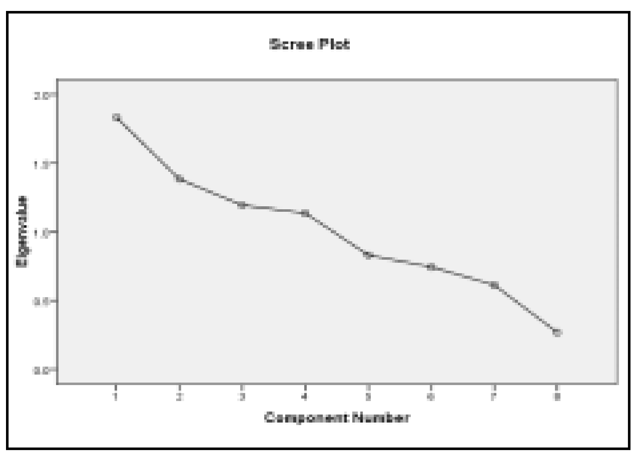

Fig. 1. Eigen values corresponding to different PC components.

TABLE 3
Principal component matrix showing higher loading displaying characters

Character	Component (s)			
	1	2	3	4
Days to 50% flowering	.196	.189	.501	.581
Plant height (cm)	.913	.056	-.036	-.085
No. of effective tillers/plant	.261	-.257	.620	.128
Ear length (cm)	.330	.477	-.189	.575
Ear diameter (mm)	-.111	.778	.171	-.174
No. of nodes/plant	.723	-.129	.248	-.473
Internode length (mm)	.397	-.397	-.569	.353
Grain yield (g/plant)	.304	.520	-.327	-.254

Extraction method : Principal Component Analysis.

to be more meaningful (Fig. 1). The higher loading displaying variables on first four PCs were plant height (cm), ear diameter (mm), number of effective tillers/ plant and days to 50 per cent flowering (Table 3).

The descriptive statistics showed that a considerable amount of diversity existed in the material used. So, the data recorded on all the traits were subjected to cluster analysis and 97 genotypes were grouped into six clusters where each genotype within a cluster was closest to the cluster mean. The Ward's method was used to carry out the agglomerative hierarchical cluster analysis. The mean performance of different clusters calculated for different traits revealed wide range of differences among the clusters (Table 4).

The multiple discriminant analysis was perfomed with an idea to select the discriminator variables leading

TABLE 4
Mean values of clusters and contribution of different characters towards genetic divergence

Character	Cluster					
	1	2	3	4	5	6
Days to 50% flowering	54.75	51.86	50.38	50.88	56.00	51.52
Plant height (cm)	219.17	176.76	150.35	128.13	151.67	200.44
No. of effective tillers/plant	4.17	3.83	3.78	3.50	3.00	3.88
Ear length (cm)	30.92	19.10	19.46	19.13	20.33	21.07
Ear diameter (mm)	21.42	20.46	19.11	21.83	96.33	21.89
No of nodes/plant	9.17	8.34	7.28	6.42	7.00	8.67
Internode length (mm)	18.08	16.85	16.79	14.92	12.67	18.28
Grain yield (g/plant)	11.25	17.70	13.88	12.70	12.00	20.14

TABLE 5
Variables entered/removed[a,b,c,d]

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	Plant height (cm)	0.08	1	5	91.0	201.21	5	91.00	.000
2	Ear diameter (mm)	0.03	2	5	91.0	95.63	10	180.00	.000
3	Grain yield (g/plant)	0.02	3	5	91.0	52.76	15	246.09	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.
[a]Maximum number of steps is 16.
[b]Minimum partial F to enter is 3.84.
[c]Maximum partial F to remove is 2.71.
[d]F level, tolerance, or VIN insufficient for further computation.

TABLE 6
Percent variance explained by the discriminant functions

Function (s)	Eigen value	Per cent variance	Per cent cumulative variance	Canonical correlation
1	11.8	80.5	80.5	0.96
2	2.7	18.7	99.3	0.86
3	0.1	0.7	100.0	0.31

to the development of discriminant function which was then used for classifying the observations. Tests for differences in the means of different groups for each discriminator variable and also the tests for differences between the groups, considering all the variables simultaneously, were dealt using Wilk's test statistic (Tables 5 and 6). To determine the inter-cluster distances (Table 7), the data were analyzed on the basis of D^2-statistics to measure the genetic divergence among the genotypes and their inter- and intra-cluster distances are shown in Fig. 2. As there are genotypes superior for individual trait belonging to different clusters which indicate that none of the clusters contained genotypes with all the desirable characters. However, the genotypes superior for specific characters from different clusters may be selected for further utilization in breeding programme. Based on the relative contributions of different characters, plant height (cm), ear diameter (mm) and grain yield (g/plant) were found the best discriminatory characters. The distributing pattern of 97 genotypes into six clusters is shown below in Table 8.

TABLE 7
Distances between final cluster centers

Clusters	2	3	4	5	6
1	44.61	70.02	91.99	101.57	23.19
2		26.78	48.99	80.35	23.97
3			22.52	77.57	50.63
4				78.34	72.83
5					89.67
6					

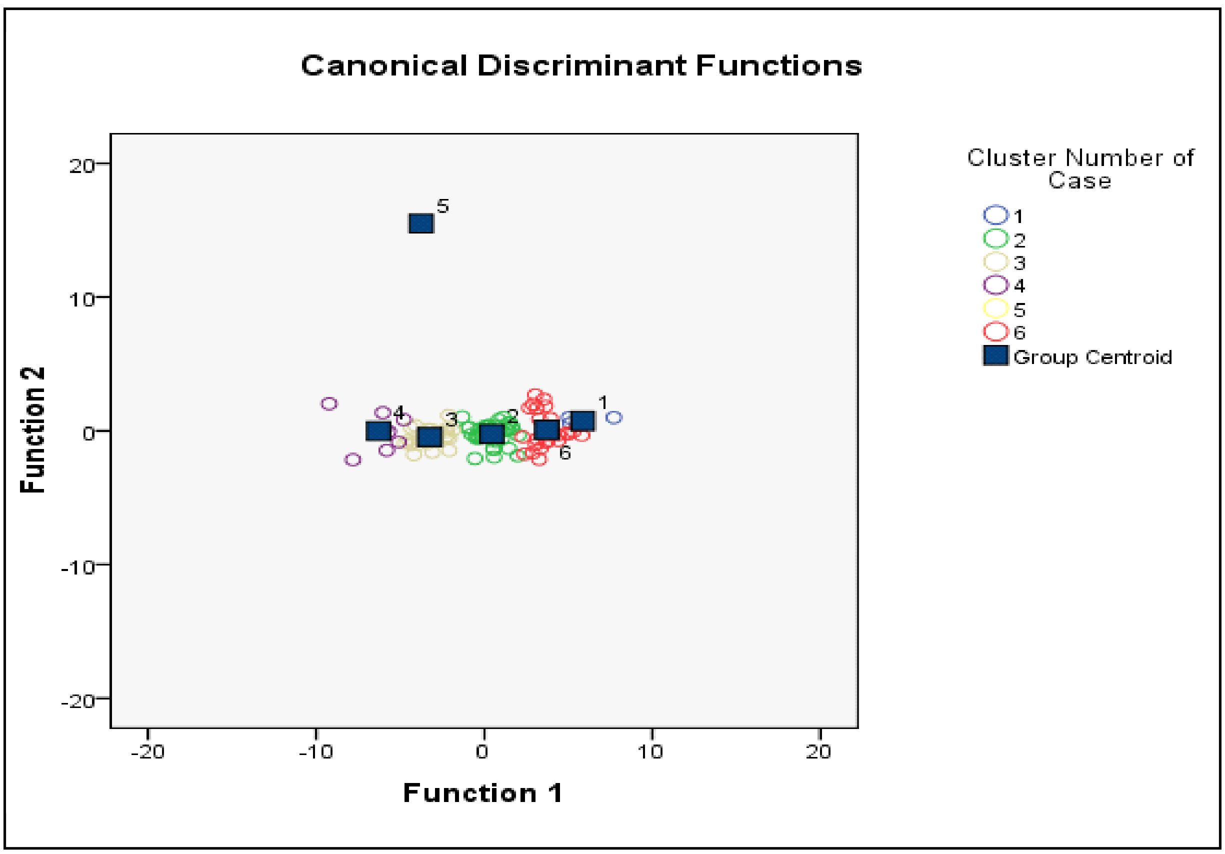

Fig. 2. For inter- and intra-cluster distances.

TABLE 8
Distributing pattern of 97 genotypes of pearl millet into six clusters

Cluster 1 (4)	Cluster 2 (35)	Cluster 3 (24)	Cluster 4 (8)	Cluster 5 (1)	Cluster 6 (25)
97 Raj 62	H 94/17	H 77/833-2	2194	99 ABL-52	HTP 92/5
[(RVTP 93/102-19 x	(HTBC-208-1-B-2-1-1x B-	[(ICMB 92333 x	1234		HTP-0814-1
	line Bulk]-28	EEBC C1-1)-5-B-B]	1219/2		96 AC-27-01K-1
1307(Tall)]	SPF2 98-2	HTP 3/14	99 HS-13		S 97/97
HTP 94/54	(H 90/4-5 x 77/29-2)	PAT 202	1246		AC-04/22
PMRLN 98-37	HTP 91/43	HBL-0566 205-1-2	77/273		2231
	(B.Senegal-2-5 x 700651)-	H 90/4-5	HTP 93/37		99 ABL-122
	2-1-4	IP-495			1449/98K 1
	[K-560-2 x (J834-7 x	HTP 91/13			PMRLN 98-37
	700544-7-2-1)]	VCF5 279			S 97/120
	-1-1-K-06-1	ICMV 95501			[(K-560-2 x (J 934
	HBL-0505	[ICMB 91777 x (91777B			-7 x 700544-7-2-1)]
	HBL-0521	x HHVBC)]-11-B-3			-4-1-3-3-2-2 (A1R)
	IP 17862 (ICTP 8203)	IP-326			HTP-03/9-K-06-1
	TCH 19-1	ICR 159-K-06-1			HBL-0510
	HBL-0551	PAT 216-1			ICMS 7704-S1-52
	IP 18535	HTP 03/13			-3-1-2-1-2-1-6-B
	H 72-2-2/98K-1	HTP 07-35			HC 20 S1-16
	AC-04/6	(H 77/833-2 x IP 15336)			HTP 07-19
	S 97/28	A5R-08—106			HTP 07-37
	IPS-98-2-03K-1	HBL-0526			HTP 07-63
	AC-04/17	LPB-10-112			HPT-10-105
	[ICMB 91777 x (91777b	PT-1-10-1019			HPT-10-110
	x HHVBC)]				PT-1-10-1060
	-6-B-4				TP-10-141
	JBV 3 S1-44-3B-4-B				TCF3-10-1-9
	HBL-34				HFIT-3-11-102
	S3-24P1				TPT-A2-1-11-131
	TCH 10-13				
	Raj-3				
	HTP 07-20				
	HTP 07-67				
	HTP 92/45				
	TCH 19-1-3				
	HBL-621 (13577)				
	(R-234 x R-238)-1				
	TCF3-10-21-5				

REFERENCES

Dangaria, C. J., and S. D. Atara. 2004 : GHB-558–A newly developed dual-purpose hybrid of pearl millet. In : *Millet Research and Development – Future Policy Options in India*, Sharma, Y. K and Khairwal, I. S. (eds.). 11-12 March 2004. Agricultural Research Station, Mandor, Jodhpur All India Coordinated Pearl Millet Improvement Project. p. 6.

Fisher, R. A. 1936 : The use of multiple measurements in taxonomic problem. *Ann. Eugen, London*, **7** : 179-188.

Johnson, R. A., and D. W. Wichern. 2006 : *Applied Multivariate Statistical Analysis*. Pearson Education.

Jolliffe, I. T. 1972 : Discarding variables in a principal component analysis. I. Artificial data. *Appl. Statistics*, **21** : 160-173.

Joshi, R. P., G. S. Chauhan, and H. S. Yadav. 1988 : Genetic divergence in pearl millet. *Exptl. Genet.*, **4** : 16-28.

Kaiser, H. F. 1960 : The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20** : 141-51.

Mahalanobis, P. C. 1936 : On the generalised distance in statistics. *J. Genet.*, **41** : 159-193.

Shanmuganathan, M., A. Gopalan, and K. Mohanraj, 2006 : Genetic variability and multivariate analysis in pearl millet germplasm for dual purpose. *The J. Agric. Sci.*, **2** : 73-80.

Vidyadhar, B., P. Chand, S. L. I. Devi, M. V. S. Reddy, and D. Ramachandraiah. 2004 : Variability and correlation analysis for the yield attributes in pearl millet germplasm. In : *3rd National Seminar on Millet Research and Development–Future Policy Options in India*, Y. K. Sharma, and I. S. Khairwal (eds.). Jodhpur, India. pp. 12.