Original Research Article

Cluster and Principal Component Analysis in Maize

L. Suryanarayana[1*], M. Reddi Sekhar[3], D. Ratna Babu[4], A.V. Ramana[2] and V. Srinivasa Rao[5]

[1]Department of Genetics and Plant Breeding, [2]Department of Agronomy, Agricultural College, Naira, A.P., India
[3]Department of Genetics and Plant Breeding, S.V. Agricultural College, Tirupati, A.P., India
[4]Plant Breeding, RARS, Lam farm, Guntur, A.P., India
[5]Department of Statistics and Mathematics, Agricultural College, Bapatla, A.P., India
Corresponding author

A B S T R A C T

In the present study all the 30 genotypes were grouped into six non overlapping clusters based on non- hierarchical Euclidean cluster analysis. Maximum numbers of genotypes (7) were grouped in cluster-I, IV and VI followed by cluster-II (5), cluster-V (3) and cluster III with one genotype. Inter-cluster distances was highest between cluster IV and VI (916.73) followed by III and V (846.80). Among the 21 characters studied, grain yield plant^{-1}, stover yield plant^{-1}, number of kernels per row and ear height contributed maximum towards the total divergence.

Introduction

In India, the production of maize over the years has increased linearly and major breakthrough was experienced at the dawn of 21st century, associated with the development and release of more number of hybrids during this period.

Recent emphasis on development of high yielding hybrids in preference to a continued dependence on composites has yielded rich dividends. This emphasis is likely to ensure still higher growth rates in productivity of this versatile crop in the years to come. There is still a considerable scope to improve productivity and adaptability by breeding

heterotic hybrids (Grzesiak, 2001). The success of any breeding method depends on the availability of genetic diversity in the base population. Hierarchical cluster analysis could serve as a basis for selection of parental types that could result to superior hybrids. Several authors suggested first principal component (PC) scores as input variables for the clustering process (Mujaju and Chakuya, 2008).

Hierarchical cluster analysis has been suggested for classifying entries of germplasm collections based on degree of similarity and dissimilarity (Van Hintum,

1995). Similarly, a combination of cluster analysis and principal component analysis has been used to classify maize (*Zea mays* L.) accessions (Crossa *et al.,* 1995).

Principal component analysis is a sort of multivariate analysis where canonical vectors or roots representing different axes of differentiation and amount of variation accounted for each of such axes, respectively, are derived (Rao, 1952). It is called principal component analysis as it reflects the importance of the largest contributor to the total variation at each axes of differentiation. The objective of this study was to analyze the genetic diversity among 30 Maize genotypes and to classify the genotypes in to different groups based on Euclidian distance.

Materials and Methods

Seeds of 30 maize inbred lines were obtained from Agricultural Research Institute, Hyderabad and were raised in Randomized Block Design (RBD) with three replications. Observations regarding 21 agronomic and physiological traits *viz.*, days to 50% flowering, days to 50% silking, days to maturity, plant height (cm), ear height (cm), ear length (cm), ear girth (cm), number of kernel rows per ear, number of kernels per row, 100 kernel weight (g), grain yield/plant (g), leaf area index at 30, 60 and at 90 DAS, LAD at 30-60 and at 60-90 DAS, SCMR, RGR at 30-60 and at 60-90 DAS, harvest index, stover yield/plant (g) were recorded in five randomly selected plants in each replication.

Data were subjected to analysis of Mahalanobis' D^2-statistics and intra-cluster and inter-cluster distance, cluster mean and contribution of each trait to the divergence were estimated as suggested by Singh and Chaudhary (1985).

Results and Discussion

Thirty genotypes were grouped into various clusters by using agglomerative hierarchical cluster analysis. Principal component scores of genotypes were used as input for clustering using Ward's minimum variance method and the tree like structure called dendrogram was constructed based on Euclidean^2 distance computed from PCA scores of genotypes (Fig. 1).

All the 30 genotypes were grouped into six clusters. The distribution of genotypes into six clusters is presented in table 1. Among all the clusters, cluster I, IV and VI were the largest containing seven genotypes each followed by cluster II with five genotypes, cluster V with three genotypes and cluster III is solitary with one genotype. The average intra and inter-cluster Euclidean^2 distance were estimated based on Ward's minimum variance and are presented in table 2.

The mutual relationship between these clusters is represented diagrammatically by taking average intra and inter- cluster Euclidean^2 distances. Cluster II recorded the maximum intra cluster Euclidean^2 distance (214.31) followed by cluster V (177.94), cluster IV (122.34), cluster VI (90.14) and cluster I (89.18).

The PCA scores for individual genotypes were used for clustering the genotypes as suggested by Anderberg (1993). Principal components (Eigen value greater than one), Eigen values (Latent Root), per cent variability, cumulative per cent variability and component loading of different characters are presented in table 3.

In the present study, the six principal components with Eigen values greater than one contributed 85.31 per cent towards the total variability. It was therefore inferred that

the essential features of data set had been represented in the first six principal components. The first principal component contributed maximum towards variability (39.65).

The significant factors loaded in PC_1 towards maximum genetic divergence were LAD at 30-60 DAS, 100-kernel weight, grain yield per plant, LAI at 30 DAS, LAI at 60 DAS, SCMR, number of kernels per row and ear length. The second principal component (PC_2) described 16.26 per cent of total variance and

the characters *viz.,* ear height, stover yield per plant, days to 50% silking, days to 50% tasseling, RGR at 30-60 DAS and plant height contributed significant factor loadings.

The third principal component (PC_3) explained 9.93 per cent of total variance and the characters *viz.,* LAD at 60-90 DAS, LAI at 90 DAS, number of kernels per row and days to maturity were the contributors for the maximum variance in this principal component.

Fig.1 Dendrogram showing relationship of 30 maize inbredlines (wards minimum)
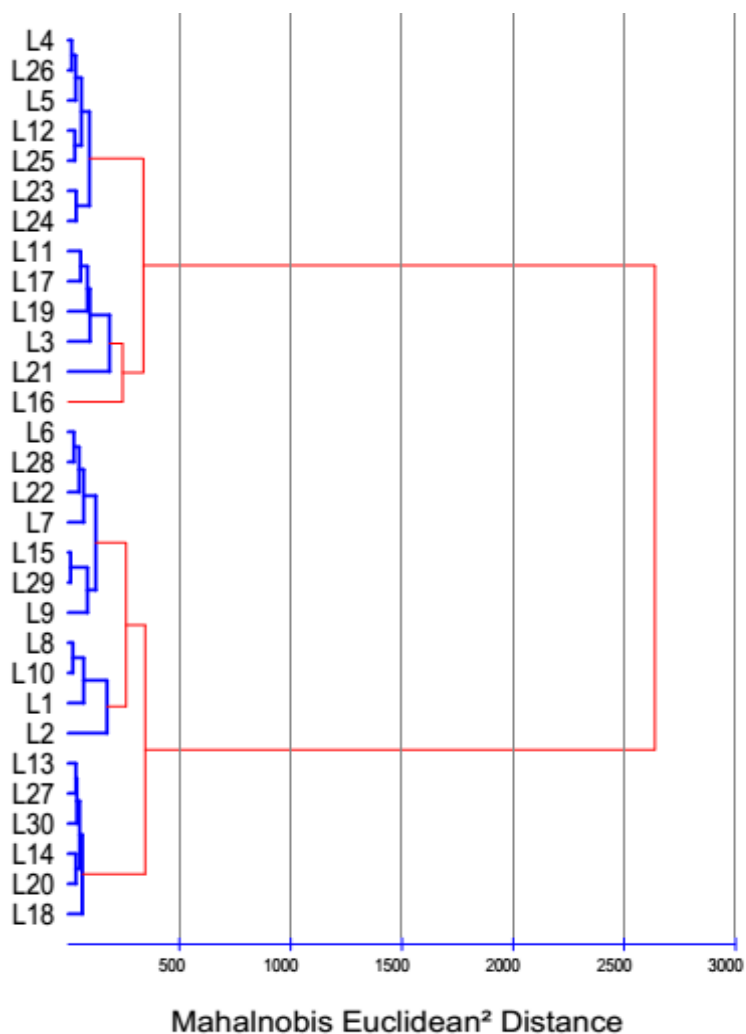
Fig.2 Two dimensional graph showing relative position of
30 maize inbred lines based on PCA scores
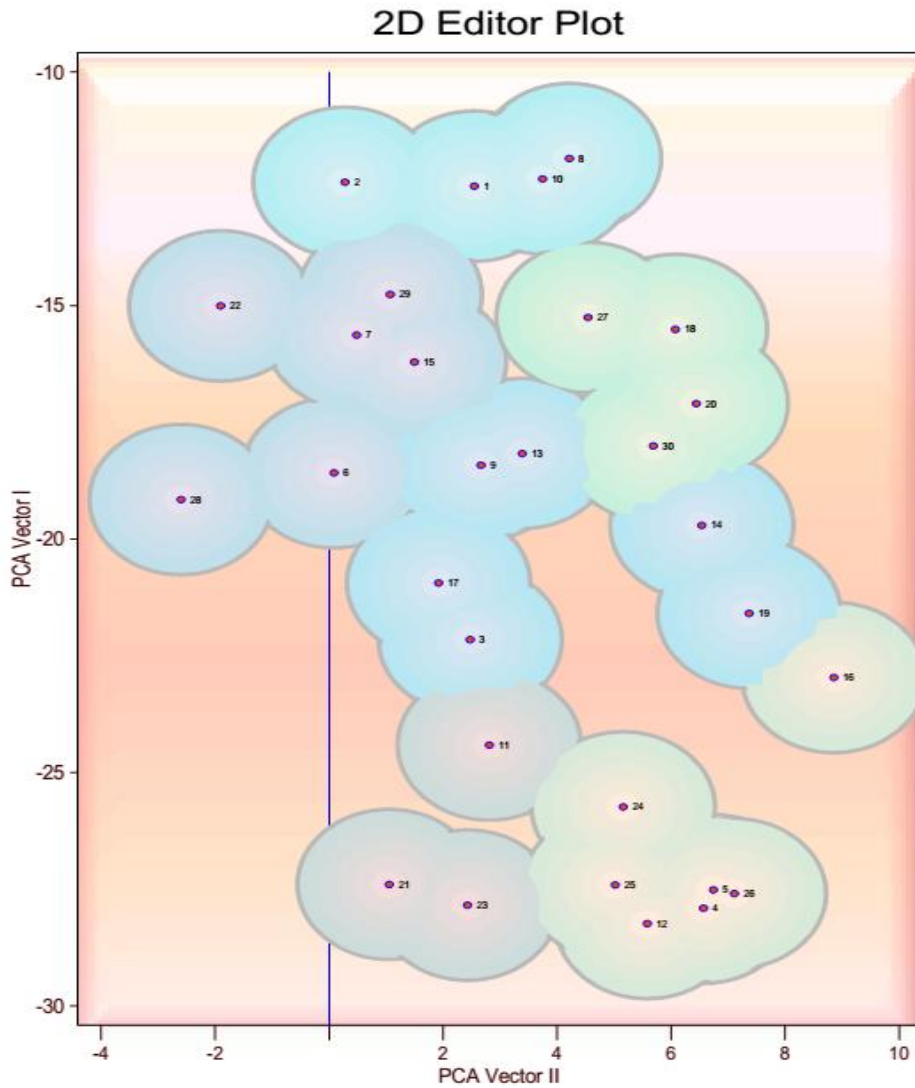
Table.1 Cluster composition of 30 genotypes of maize, complete linkage dendrogram

Cluster	No. of genotypes	Cluster Members
1	7	MRC- 152, MRC- 170, MRC- 180, MRC- 219, MRC- 134, MRC- 163, MRC- 191.
2	5	MRC- 190, MRC- 126, MRC- 157, MRC- 216, MRC- 185.
3	1	MRC- 139
4	7	MRC- 186, MRC- 184, MRC- 160, MRC- 151, MRC- 197, MRC- 163, MRC- 203.
5	3	MRC- 153, MRC- 130, MRC- 180.
6	7	MRC- 194, MRC- 179, MRC- 167, MRC- 147, MRC- 206, MRC- 132, MRC-139

Table.2 Inter and Intra (diagonal) cluster average Euclidean^2 and Euclidean values (parenthesis) of 30 genotypes of maize - complete linkage dendrogram

	1 Cluster	2 Cluster	3 Cluster	4 Cluster	5 Cluster	6 Cluster
1 Cluster	**89.81 (9.47)**	234.28 (15.30)	362.61 (19.04)	597.98 (24.08)	916.73 (30.27)	440.79 (20.99)
2 Cluster		**214.31 (14.63)**	379.88 (19.48)	373.68 (19.33)	616.88 (24.83)	325.80 (18.04)
3 Cluster			**0.00 (0.00)**	720.12 26.83)	846.80 (29.09)	486.07 (22.04)
4 Cluster				**122.34 (11.06)**	221.38 (14.87)	192.77 (13.88)
5 Cluster					**177.94 (13.33)**	235.53 (15.34)
6 Cluster						**90.14 (9.49)**

Table.3 Eigen values, proportion of the total variance represented by first six principle components, cumulative percent variance and component loading of Different characters in maize inbred lines

	PC$_1$	PC$_2$	PC$_3$	PC$_4$	PC$_5$	PC$_6$
Eigene Value (Root)	**8.33**	**3.41**	**2.09**	**1.78**	**1.35**	**0.96**
% Var. Exp.	**39.65**	**16.26**	**9.93**	**8.49**	**6.42**	**4.56**
Cum. Var. Exp.	**39.65**	**55.91**	**65.84**	**74.33**	**80.75**	**85.31**
Days to 50% tasseling	0.20	0.33	0.03	0.31	0.00	0.20
Days to 50% silking	0.13	0.34	-0.03	0.26	0.02	0.45
Days to maturity	0.09	-0.08	-0.23	0.44	0.40	-0.32
Plant height (cm)	-0.17	-0.30	-0.01	0.09	-0.10	0.55
Ear height (cm)	-0.19	-0.38	0.01	0.17	-0.10	0.18
Ear length (cm)	-0.25	0.24	-0.15	-0.13	-0.16	-0.23
Ear girth (cm)	-0.20	0.17	-0.12	0.30	0.27	-0.22
No of kernel rows per ear	-0.11	-0.13	0.07	-0.49	0.28	-0.12
No of kernels per row	-0.25	0.17	-0.24	-0.10	-0.28	-0.12
100 kernel weight (g)	-0.31	0.04	-0.12	0.04	0.18	0.10
Grain yield/ plant (g)	-0.31	0.17	-0.18	0.03	0.07	0.05
Leaf area index at 30 DAS	-0.29	-0.18	-0.08	0.09	-0.13	-0.10
Leaf area index at 60 DAS	-0.27	0.26	0.20	0.00	0.06	-0.03
Leaf area index at 90 DAS	-0.13	0.00	0.57	0.15	0.11	-0.09
LAD at 30-60 DAS	-0.33	0.04	-0.18	-0.02	0.06	0.14
LAD at 60-90 DAS	0.16	-0.09	-0.59	-0.06	-0.11	0.06
SCMR	-0.26	-0.19	0.14	0.07	0.07	0.07
RGR at 30-60 DAS	-0.04	-0.30	0.06	0.22	-0.35	-0.18
RGR at 60-90 DAS	0.19	-0.02	-0.04	-0.38	0.34	0.18
Harvest Index (%)	-0.25	-0.04	-0.07	-0.03	0.39	0.23
Stover yield/ plant (g)	0.16	-0.36	-0.14	0.15	0.29	-0.10

The analysis thus identified the maximum contributing variables *i.e.,* days to 50 % silking, days to 50% tasseling, number of kernels per row, leaf area index at 30 DAS, ear length, leaf area duration at 60-90 DAS and leaf area index at 90 DAS for total divergence.

Results of cluster analysis based on PCA scores were compared with the results of the principal component analysis on a visual aid in discerning clusters in 2D (Fig. 2) and 3D scattered diagrams. The genotypes falling in same cluster were present closer to each other in scattered diagrams. In the 2D and 3D scattered diagrams, the genotypes, MRC 139, MRC 216, MRC 127, MRC 180, MRC 163 were present distantly from the other genotypes and the inter cluster distance among these genotypes is also high indicating their usefulness in breeding programmes. Alika *et al.,* (1993), Okporie (2008*)*, Mehrnaz *et al.,* (2014), Muhammad *et al.,* (2015) and Hughes *et al.,* (2015) studied the utilization of principal component analysis combined with clustering of Ward's method in genetic divergence studies in maize.

References

Anderberg, M.R. 1993. *Cluster Analysis for Application*. Academic Press, New York.

Alika, J.E., Aken, M.E.O and Fatokun, C.A. 1993. Variation among maize (*Zea mays* L.) accessions of Bendal state, Nigeria and multivariate analysis of agronomic data. *Euphytica*. 66: 65-71.

Crossa, J., Delacy, I.H., Taba, S. 1995.*The use of multivariate methods in developing a core collection*. p. 77-92. In: Hodgkin, T., Brown, A.H.D., Van Hintum, Th. J. L., Morales, E.A.V. (Eds.) Core collections of plant genetic resources. John Wiley & Sons, Chichester, UK.

Grzesiak, S. 2001. Genotypic variation between maize (*Zea mays* L.) single-cross hybrids in response to drought stress. *Acta Physiologiae Plantarium*, 23(4): 443-456.

Hughes, A. N., Louise Akanvou., Arsène Irié, Z. B and Charles Konan, K. 2015. Phenotypic diversity of farmer's traditional maize (*Zea mays* L) Varieties in Cote d'Ivoire. *Maydica*. 60(2): 112-118.

Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Proceedings of Natinal Institute of Sciences. India*. 12: 49-55.

Mehrnaz, T., Elham, B., Ali, R, A and Ali, R, A. K. 2014. Genetic Diversity of 13 Maize (*Zea mays* L.) Hybrids based on multivariate Analysis Methods. *International Journal of Farming and Allied Sciences*. 2014-3-5:467-470.

Muhammad, R.W., Qayyum, A., Hamza, A., Ahmad, M.Q., Naseer. N.S., Liaqat, S., Ahmad, B., Malik, W and Noor, E. 2015. Analysis of genetic traits for drought tolerance in maize. *Genetics and Molecular Research*. 28: 14(4): 13545-65.

Mujaju, C., Chakuya, E. 2008. Morphological variation of sorghum landrace accessions on-farm in Semi-arid areas of Zimbabwe. *International Journal of Botany*. 4:376-382.

Okporie, E.O. 2008. Characterization of maize (*Zea mays* L.) germplasm with principal component analysis. *Agro Science*. 7 (1): 66-71.

Rao, C. R. 1952. Advanced statistical methods in biometrical research. *John Wiley and Sons Inc., New York*.236-272.

Van Hintum, T.J.L. 1995. *Hierarchical approaches to the analysis of genetic diversity in crop plants*. In: Hodgkin, T., Brown, A.H.D., Van Hintum T.J.L., Morales, E.A.V. (Eds.) Core collection of plant genetic resources. John Wiley and Sons, pp. 23-34.

Singh, R.K and Chaudhary, B.D. 1985. Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi. pp. 102-157.