PRINCIPAL COMPONENT ANALYSIS IN INBREDS OF MAIZE (*Zea mays* L.)

S. MALLIKARJUNA[1], V. ROJA[1], I. SUDHIR KUMAR[2], S.K. NAFEEZ UMAR[3] AND T. SRINIVAS[1]

[1]Department of Genetics and Plant Breeding,[3]Statistics and Computer Applications
Agricultural College, Bapatla - 522 101, Guntur District, Andhra Pradesh, INDIA
[2]Agricultural Research Station, Peddapuram - 533 437, East Godavari District, Andhra Pradesh, INDIA
e-mail: sanikommuashokreddy@gmail.com

ABSTRACT
The present study was carried out during *kharif,* 2018 using 49 maize inbred lines at Agricultural Research Station (ARS), Peddapuram, Andhra Pradesh. Principal Component Analysis (PCA) revealed first five principal components with eigen value greater than one. These principal components together accounted for 75.39 per cent of the total variability. The first Principal Component (PC1) contributed maximum towards variability (23.06%) followed by second (17.48%), third (14.66%) and fourth (11.52%) Principal Components, respectively. The characters, namely plant height, ear placement height, days to 50 per cent tasselling, cob length, grain yield plant-1, days to 50 per cent silking and cob girth contributed maximum for the variability. The results of scatter diagram revealed importance of hybridization between CML 1505, CML28708 and PDML 19-39 inbred lines with ADL 8070 and PDML15 inbreds for the production of heterotic and high yielding hybrids due to their diversity.

INTRODUCTION

Maize (*Zea mays* L.) is known as golden crop, because every part of the crop is useful to man, animals and the industries. It is one of the most important cereal crops of the world after wheat and rice. In India, the production of maize has increased linearly over the years and major breakthrough was experienced at the dawn of 21st century, with the development and release of several hybrids (Suryanarayana *et al.*, 2017). Globally, maize crop accounts for 9 per cent of the total food grain production crop has occupied a prominent place in Indian agriculture as it is widely grown in India under varied climatic situations throughout the year (Mounika *et al.*, 2018).

The seed yield, is a polygenic trait involving a number of genes contributing in it and their interaction with environment (Hassan *et al.*, 2013). So, it is important to measure the mutual relationship between various plant attributes and determine the component characters, on which selection procedure can be based for direct and indirect genetic improvement of crop yield. Seed yield is influenced by several yield contributing traits. These components are related among themselves and also with yield either positively or negatively. However, sometimes breeders obtain measures on a number of observed variables and wish to develop a smaller number of artificial variables (principal components) which will account for most of the variance in the observed variables.

The principal components may then be used as predictor or criterion variables in subsequent analyses. Actually, it is a variable reduction procedure and useful when breeders have obtained data on a number of variables (possibly a large number of variables), and believe that there is some redundancy in those variables. In this case, redundancy means some of the variables are correlated with one another, possibly because they are measuring the same construct (Jupp, 2006). Hence, the present investigation was undertaken in this context to study the importance of traits associated with seed yield of maize along with their inter-relationship and to cluster them using PCA analysis.

MATERIALS AND METHODS

The experimental material consisted of 49 inbreds of maize including one check. Details of the inbred lines studied are presented in Table 1. The material was sown in Simple Lattice Design (SLD) with two replications at Agricultural Research Station (ARS), Peddapuram, Andhra Pradesh during *kharif'* 2018. Each inbred was planted in two rows of 4 meter length in each replication with a spacing of 70 cm between rows and 20 cm within the row. All the recommended package of practices were followed in order to raise a healthy crop. Observations were recorded on five randomly selected plants for plant height (cm), ear placement height (cm), cob length (cm), cob girth (cm), kernel rows cob-1, number of kernels row-1, cob yield plant-1, kernel yield plant-1, days to maturity,

100 kernel weight and protein content (%). However, observations for the characters, namely, days to 50 per cent tasseling, days to 50 per cent silking and anthesis silking interval were recorded on plot basis. The data collected was subjected to standard statistical procedures given by Panse and Sukhatme (1967).

Principal component analysis was carried according to procedure described by Banfield (1978). PCA can be performed on two types of data matrices *viz.*, variance-covariance matrix and correlation matrix. With characters of different scale, a correlation matrix standardizing the original data set is preferred. If the characters are of same scale, a variance-covariance matrix can be used. In the present study, PCA was performed on the correlation matrix of traits, thereby removing the effects of scale (Jackson, 1991). A data matrix of all 49 genotypes over 14 variables of each genotype was prepared and used for analysis by WINDOSTAT computer programme.

Eigen values and eigen vectors

The eigen values and eigen vectors were computed from data matrix. Eigen values define the amount of total variation that is displayed on principal components. The proportion of variation accounted for each principal component (PC) is expressed as the eigen value divided by the sum of the eigenvalues.

$$\text{Percent variance explained for PCI} = \frac{\text{Eigen value (PCI)}}{\text{Sum of eigen values}}$$

The eigen vector (loading) defines the correlation of each variable with the principal components. The principal components were identified by following procedure. The jth principal component (Yj) of the observations X is the linear combination given as follows:

$$Y_j = A_{1j}X_1 + \ldots\ldots + A_{pj}X_p$$

Where,

Aij are found such that Yj is uncorrelated Y1, Y2...Yj-1 the jth largest variance.

The Aij are the elements of the normalized eigen vector associated with largest jth eigen value. The variance of the jth principal component of the λ j and the total system variance(S)

$$= \lambda 1 + \lambda 2 + \ldots + \lambda p.$$

The importance of the jth principal component is given by,

$$\frac{\lambda j}{\text{Trace}}$$

This is informative about the proportion of total variation that can be accounted for the ith principal component. The correlation between the ith original variable Xi

and the jth principal component Yj is given by

$$\rho(Xi, Yj) = \frac{Aij\sqrt{\lambda j}}{\sqrt{Si}}$$

Where, Si is the standard deviation of Xi.

Thus, a principal component is linear function of the test

variables given as follows,

Principal component = ax1 + bx2 + ... + hx8

Where, a, b,.... are coefficients and x1, x2 ..... *etc.*, are the variables in such a way that the principal component has a unit variance as reported by Ehrenberg (1985).

PCA scores for each genotype under concerned PCs were computed and utilized to derive 2D or 3D (dimensional) scatter plot of individuals.

RESULTS AND DISCUSSION

The results on analysis of variance (ANOVA) for yield and yield component traits (Table 2) revealed significant differences among the genotypes for all the characters studied, indicating the existence of sufficient variation among the inbreds and therefore effective selection of inbred lines for hybridization program aimed at the development of high yielding and heterotic hybrids. The results on Principal Component Analysis (PCA) are presented in Table 3-4 and Fig. 1. The eigen values, proportion of total variance represented by Principal Components of importance and the component loadings of different characters for the Principal Components are presented in Table 3.

A perusal of these results revealed first five Principal Components with eigen value more than one as contributing to 75.39 per cent towards the total variability, and hence, were considered for interpreting the results. The first Principal Component (PC1) contributed maximum towards total variability (23.06%). The characters namely plant height (0.469), ear placement height (0.442), days to 50 per cent tasseling (0.360) and cob length (0.347) had explained maximum variance in this component. The second Principal Component (PC2) contributed 17.48 per cent of total variance. The characters, namely, number of kernels row-1 (0.343), cob length (0.146) and days to 50 per cent silking (0.138) recorded maximum loading for the component. The third Principal Component (PC3) was characterized by 14.66 per cent contribution towards the total variability. The characters,
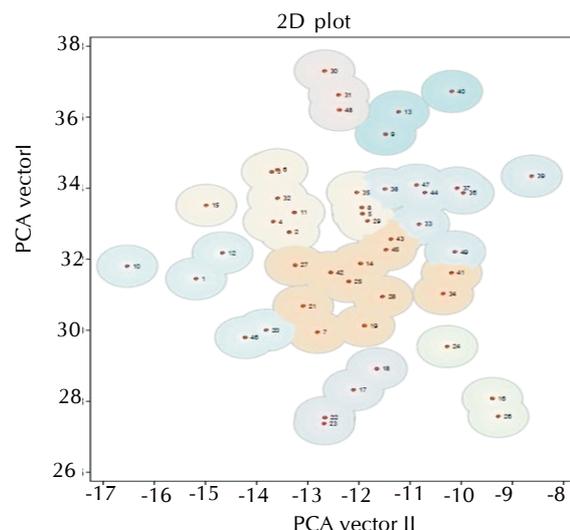

Figure 1: Two-dimensional graph demonstrating relative position of 49 maize (*Zea mays* L.) genotypes based on PCA scores

Table 1: List of maize genotypes employed in the present investigation

S. No	Source	No. of Genotypes	Genotypes
1	Winter Nursery Centre, Hyderabad	17	VML 15036, VML 16008, OML 17-47, ZL161032, ZL1381388081, VML 15028, ADL1608, ADL1608, ADL8070, ADL1620, AL8178, KDML-115, ADL1619, AL8106, VM 45, VM 51, TA5084,
2	CIMMYT, Regional Center, Hyderabad	10	CAL1511, CAL1711830, CML08-292, CAL1784, CAL17834, CAL1612, CAL1810, CML1505, CML28708, CML28208
3	Agricultural Research Station, Peddapuram	21	PDML9644, PDML124, PDML7470, PDML506, PDML15, PDML8082, PDML19-34, PDML19-35, PDML19-36, PDML19-37, PDML19-38, PDML19-39, PDML19-40, PDML19-41, PDML19-42, PDML19-43, PDML19-44, PDML19-45, PDML19-46, PDML19-47, PDML19-48
4	PJTSAU, Hyderabad	1	BML6

Table 2: Analysis of variance for kernel yield and yield attributing characters in maize

Source of variation	Df	DT	DS	ASI	DM	PH	EPH	CL	CG	KR	KPR	CYP	KYP	100 GW	PC
		Mean Sum of Squares													
Replications	1	1.02	1.23	0.01	40.5	36.73	102.04	1.78	0.97	0.16	4.94	5527.51	4039.72	0.26	1.4
Treatments (Unadj.)	48	28.97**	30.16**	0.50*	59.44**	837.15**	314.89**	4.29*	1.06**	1.79**	21.02**	1488.52*	1100.80*	27.76**	2.19**
Blocks within replications (Adj.)	12	4.57	4.09	0.28	20.92	60.07	30.23	2.61	0.39	0.59	23.82	1028.96	760.27	9.59	0.9
Error	36	1.67	1.56	0.28	4.47	81.87	30.12	2.06	0.43	0.69	7.94	813.28	540.46	8.59	0.88

** Significant at 1% level; *Significant at 5% level;

Df: Degrees of freedom; DT: Days to 50 per cent tasseling; DS: Days to 50 per cent silking; CL: Cob length; KYP: Kernel yield plant-1; ASI: Anthesis silking interval; CG: Cob girth; 100 GW: 100 grain weight; DM: Days to maturity; KR: Kernel rows cob-1; PC: Protein content; PH: Plant height; KPR: Kernels row-1

Table 3: Eigen values, proportion of the total variance represented by first six Principal Components, cumulative per cent variance and component loading of different characters in maize

	PC I	PC II	PC III	PC IV	PC V	PC VI
Eigen Value (Root)	3.228	2.448	2.052	1.612	1.213	0.852
% Var. Exp.	23.061	17.488	14.659	11.518	8.665	6.089
Cum. Var. Exp.	23.061	40.548	55.207	66.725	75.39	81.479
Days to 50 per cent tasselling	0.36	0.086	0.333	0.327	0.148	0.175
Days to 50 per cent silking	0.239	0.138	0.442	0.397	0.148	0.106
Anthesis silking interval	-0.162	-0.009	0.367	-0.229	-0.247	0.183
Days to maturity	0.119	-0.372	-0.243	-0.056	-0.275	0.599
Plant height (cm)	0.469	-0.087	-0.1	0.005	-0.04	-0.224
Ear placement height (cm)	0.442	0.007	-0.141	-0.228	0.091	0.013
Cob length (cm)	0.347	0.146	-0.135	-0.34	0.245	-0.128
Cob girth (cm)	-0.283	-0.386	0.048	0.123	-0.012	-0.503
Kernel rows cob-1	0.031	-0.038	-0.432	0.483	0.004	-0.165
Number of kernels row-1	0.02	0.343	0.303	0.005	-0.453	-0.259
Cob yield plant-1 (g)	0.115	-0.491	0.141	0.294	0.147	-0.06
Grain yield plant-1 (g)	-0.345	0.045	0.066	0.057	0.518	0.326
100 grain weight (g)	0.152	-0.465	0.264	-0.077	-0.295	0.038
Protein content (%)	0.022	-0.281	0.278	-0.411	0.41	-0.2

Table 4: PCA scores of 49 maize inbred lines studied

Table 4 (cont.)

S. No.	GENOTYPE	PCA I X Vector	PCA II Y Vector	PCA III Z Vector
1	VML 15036	29.134	-13.749	9.582
2	VML 16008	30.403	-11.919	9.171
3	OML 17-47	32.175	-12.338	7.917
4	CAL1511	30.782	-12.302	8.192
5	ZL161032	31.123	-10.667	6.577
6	ZL1381388081	32.175	-12.166	9.097
7	CAL1711830	27.444	-11.247	10.65
8	VML 15028	31.314	-10.65	8.721
9	ADL1608	33.265	-10.044	9.749
10	ADL8070	29.357	-15.119	7.641
11	ADL1620	31.076	-11.912	8.342
12	AL8178	29.938	-13.351	7.396
13	CML08-292	33.98	-9.951	6.794
14	PDML9644	29.563	-10.511	9.588
15	PDML124	30.998	-13.528	8.255
16	KDML-115	26.091	-8.363	5.261
17	CAL1784	26.035	-10.832	7.598
18	CAL17834	26.567	-10.281	8.087
19	ADL1619	28.12	-10.736	6.414
20	AL8127	27.908	-12.629	6.982
21	ADL8106	28.409	-11.741	7.996
22	VM 45	25.671	-11.457	9.109
23	VM 51	25.249	-11.489	7.169
24	PDML7470	27.385	-8.94	8.357
25	PDML506	29.045	-10.808	8.334
26	PDML15	25.531	-8.198	4.945
27	CAL1612	29.542	-11.821	8.629
28	TA5084	28.477	-10.198	8.685
29	CAL1810	30.914	-10.519	8.258
30	CML1505	34.999	-11.491	7.428
31	CML28708	34.325	-11.116	6.952
32	CML28208	31.595	-12.335	7.662
33	PDML8082	30.757	-9.532	6.473
34	PDML19-34	28.565	-8.692	11.358
35	PDML19-35	31.398	-10.402	10.76
36	PDML19-36	31.461	-8.429	9.601
37	PDML19-37	31.605	-8.761	6.332
38	PDML19-38	31.769	-10.28	6.226
39	PDML19-39	31.985	-7.203	7.593
40	PDML19-40	34.379	-8.828	8.281
41	PDML19-41	29.262	-8.815	7.038
42	PDML19-42	29.47	-11.168	8.185
43	PDML19-43	30.374	-10.152	6.558
44	PDML19-44	31.265	-8.997	11.534
45	PDML19-45	29.787	-10.15	9.108
46	PDML19-46	27.236	-12.74	8.329
47	PDML19-47	31.633	-9.4	8.438
48	PDML19-48	33.756	-10.671	12.851
49	BML6	29.721	-8.496	9.695

namely, days to 50 per cent silking (0.442), anthesis silking interval (0.367) and days to 50 per cent tasseling (0.333) explained maximum variation in this component. The fourth principal component (PC4) contributed 11.52 per cent towards the total variability and the characters kernel rows cob-1 (0.483), days to 50 per cent silking (0.397) and days to 50 per cent tasseling (0.327) accounted for maximum variance in this component. The fifth Principal Component (PC5) contributed to 8.66 per cent of the total variance and the characters, namely, grain yield plant-1 (0.518) and protein content (0.410) contributed maximum towards divergence. The PCA analysis thus identified that the maximum

contributing traits towards the existing variability as plant height, ear placement height, days to 50 per cent tasseling, cob length, kernel yield plant-1, days to 50 per cent silking and cob girth. Similar results were reported earlier for plant height, by Sandeep *et al*. (2017); days to 50 per cent tasselling, days to 50 per cent silking and for kernel yield per plant by Mounika *et al*. (2018). Further, the PCA scores for 49 maize inbreds in the first three Principal Components were computed and principal components I, II, and III were considered as three axes (X, Y and Z) and squared distance of each inbred from these three axes were calculated and are presented in Table 4. These three PCA scores for 49 inbreds were plotted in graph to get two-dimensional scatter diagram presented in Fig. 1. A perusal of these results revealed the need for hybridization between CML1505, CML28707 and PDML19-39 inbred lines with ADL8070 and PDML15 inbreds for production of heterotic and high yielding hybrids due to their

diversity.

REFERENCES

Adams, M. W. 1995. An estimate of homogeneity in crop plants with special reference to genetic vulnerability in dry season. *Euphytica.* **26:** 665-679.

Hassan, S. M. F., Iqbal, M. S., Rabbani, G., Naeem-ud-Din, Shabbir, G., Riaz, M. and Noorka, I. R. 2013. Correlation and path analysis for yield and yield components in sunflower (*Helianthus annus* L.). *African J. Biotechnology.* **12(16):** 1968-1971.

Jupp, V. 2006. The Sage Dictionary of Social Research Methods. SAGE Publications India Pvt. Ltd, New Delhi, India. PP.114-115.

Kumar Saurabh Singh, Suneetha, Y., Sandeep Raja, D and Srinivas, T. 2020. Principal component analysis for yield and quality traits of coloured rice (*Oryza sativa* L.). *The Pharma Innovation J.* **9(7):** 456-462.

Mounika K., M. Lal Ahamed and Nafeez Umar Sk. 2018. Principal component and cluster analysis in inbred lines of maize (*Zea mays* L.). *International J. Current Microbiology and Applied Science.* **7(6):** 3221-3229.

Panse V.G and Sukhatme P.V. 1967. Statistical Methods for Agricultural Workers. Indian Council of Agricultural Research, New Delhi. PP.103-108.

Ray, K., Dutta, J., Banerjee, H., Biswas, R., Phonglosa, A and Pari, A. 2014. Identification of principal yield attributing traits of Indian mustard [*Brassica juncea* (L.) Czernj and Cosson] using multivariate analysis. *The Bioscan.* **9(2):** 803-809.

Sandeep, S., Bharathi, M and Narsimha Reddy, V. 2017. Principal component analysis in inbreds of maize (*Zea mays* L.) *International J. Tropical Agriculture.* **33(2):** 213-216.

Sudeepthi, K., Srinivas, T., Ravi Kumar, B.N.V.S.R., Jyothula, D.P.B. and Umar, S.N. 2020. Genetic diversity studies for yield and yield component traits using principal component analysis in rice (*Oryza sativa* L.). *International J. Pure and Applied Bioscience.* **8(3):** 228-235.

Suryanarayana, L., Reddi Sekhar, M., Ratna Babu, D., Ramana, A.V and Srinivasa Rao, V. 2017. Cluster and principal component analysis in maize. *International J. Current Microbiology and Applied Sciences.* **6(7):** 354-359.

Banfield, C. F. 1978. Principal component analysis for genstat. *J. Stat. and Computer Simulation.* **6:** 211-222.

Jackson, J. E. 1991. *A User's Guide to Principal Components.* John Wiley and Sons Inc., New York.

Ehrenberg, A. S. C. 1985. *Data Reduction Analysis and Interpreting Statistical Data.* John Wiley and Sons Inc., New York.