Multivariate Analysis of Genetic Diversity in Cotton
(*Gossypium hirsutum* L.)

G Vijaya Lakshmi, V Chenga Reddy, C Panduranga Rao, J Satish Babu and R Srinivasulu
Department of Genetics and Plant Breeding, Agricultural College, Bapatla 522 101, Andhra Pradesh

ABSTRACT

Genetic divergence was assessed among 72 genotypes of cotton based on 18 traits using Mahalanobis' D^2 statistic, cluster analysis and principal component analysis. Eight and nine clusters were obtained for Mahalnobis' D^2 statistic and cluster analysis, respectively. In PCA seven principal components were identified, out of which first five principal components explained 76.5% of variability in cotton. The principal component analysis (PCA) enabled loading of similar type of variables on a common principal component.

Key words : Cluster Analysis, Cotton, D^2 analysis, Principal Component Analysis

Cotton occupies a peerless position among the commercial corps of India. Knowledge of genetic divergence of the germplasm lines is important for the breeders in hybridization programme. Mahalanobis' D^2 analysis, cluster analysis and principal component analysis are potent tools for estimating it. The PCA technique has been applied in cotton by Brown (1991). An attempt was, therefore made to study genetic divergence in 72 germplasm lines of tetraploid cotton through multivariate analysis.

MATERIAL AND METHODS

The present study was carried out with 72 genotypes of cotton (*Gossypium hirsutum*) (Table1) obtained from different cotton growing zones of India. The experiment was conducted in randomized complete block design with three replications during *kharif* 2006. The inter - and intra-row spacing adapted was 120 X 60 cm. Each plot consisted of two rows of 6m length and observations were recorded on five competitive plants from each genotype per replication for characters *viz*., plant height (cm), number of monopodia per plant, number of sympodia per plant, number of bolls per plant, boll weight (g), seed index (g), lint index (g) elongation (%), fibre quality index, count strength product lint yield per plant (g) and seed cotton yield per plant. Days to 50% flowering, ginning out turn (%), 2.5% span length (mm), micronaire (10^{-6} g/in), bundle strength (g/tex) uniformity ratio were recorded on plot basis. D^2 analysis (Mahalanobis, 1928), hierarchial cluster analysis (Anderverg, 1993) and Principal component analysis (Jackson, 1991) were carried out using SPSS computer programme.

RESULTS AND DISCUSSION

The analysis of variance revealed highly significant difference among the genotypes for 18 characters. This difference indicated the existence of significant amount of variability among the genotypes for the characters studied. On the basis of D^2 analysis and cluster analysis, the 72 genotypes were grouped into 8 and 9 clusters, respectively (Table 2, 3 and fig1).

The variation in the composition of individual cluster with regard to the number of genotypes indicted the presence large amount of diversity in the a population. On the basis of D^2 and cluster analysis, the scattering of genotypes from the same geographic region to different clusters might be due to the heterogeneity, genetic architecture of the population, past history of selection, developmental traits and degree of general combining ability (Murthy and Arunachalam, 1966). The results have clearly indicated that there is no parallelism between the geographic diversity and genetic diversity in cotton in accordance with Altaher and Singh (2003).

On the basis of D^2 analysis the maximum intra-cluster distance was observed in cluster III (41.99) and minimum intra- cluster distance was observed in cluster IV, V, VI, VII and VIII (0.00) while on the basis of cluster analysis maximum intra-cluster distance was observed in cluster IX (163.37) and this minimum intra-cluster distance was observed in cluster V (42.66).

Inter-cluster distance was least between cluster I and cluster IV (30.14) and highest between cluster II and V (143.8) on the basis of D^2 analysis while, it was least between V and cluster VI (61.93) and maximum between cluster III and cluster IX

Table 1. Cotton (*Gossypium hirsutum* L.) genotypes name and source.

Genotype	Source
L769	RARS, Lam
SCS 601	UAS, Siriguppa
GJHV 374	ARS, Gujarat
NH 615	Nanded
CA 100	Nagapur
CPD 812	Dharwad
CSH 3119	CICR, Sirsa
TCH 1706	CICR, Coimbatore
P 514	AICCIP
RS 2455	CICR, Coimbatore
ADB 134	ARS, ADilabad
AKH 2017	ARS, Akola
GSHV 152	ARS, Gujarat
L 604	RARS, Lam
Bihani 161	AICCIP
CNDTS 52	AICCIP
GTHV 02/45	NAU, Surat
HAGH 810	ARS, Hagari
CNHO 23	CICR, Nagpur
F 2052	PAU, Faridkot
RS 2461	RAU, Sri Ganga Nagar
RAH 216	UAS, Raichur
CNDTS 51	AICCIP
Surabhi	CICR, Coimbatore
RSA 2495	AICCIP
RAH 219	UAS, Raichur
F 2157	PAU, Faridkot
TSH 9812	PNAU, Serivelli Puttur
CSH 3114	CICR, Sirsa
HS 276	HAU, HIssar
LH 2111	PAU, Ludhiana
CCHLS 2	CICR, Coimbatore
BS 3	AICCIP
ARBH 813	UAS, Arbhavi
H 1300	HAV, Hissar
CCH 226	CICR, Coimbatore

Genotype	Source
RHC 2004	AICCIP
CSH 7	AICCIP
BS 144	AICCIP
RB 557	AICCIP
GISV 103	ARS, Gujarat
CCH 1831	CICR, Coimbatore
ADL 903	RARS, Lam
NDL 762	RARS, Nandyal
GISV 97/016	ARS, Gujarat
CNHO 12	CICR, Nagpur
LH 2076	PAU, Ludhiana
CPD 814	AICCIP
L 761	RARS, Lam
P 72-9-37	AICCIP
CCH 510	CICR, Coimbatore
GSHV 97/612	ARS, Gujarat
HAG 1055	ARS, Hagari
GSHV 01/1338	ARS, Gujarat
CPD 787	AICCIP
CPD 755	AICCIP
ARB 760	UAS, Arbhavi
ARB 2001	UAS, Arbhavi
ARB 784	UAS, Arbhavi
ADB 320	ARS, Adilabad
LRA 5166	CICR, Coimbatore
ADB 39	ARS, ADilabad
ARB 8906	UAS, Arbhavi
LPS 141	RARS, Lam
NDLH 1755	RARS, Nandyal
ADB 250	ARS, Adilabad
L 603	RARS, Lam
ARB 2005	UAS, Arbhavi
ADB 10050	ARS, ADilabad
L 804	RARS, Lam
L 778	RARS, Lam
L781	RARS, Lam

(389.8) on the basis of cluster analysis (Table 4 and 5). The hybridization among the genotypes of these culstern should give rise to heterotic hybrids is subsequent segregating generations.

D^2 analysis indicated that the number of bolls per plant (35.49%), plant height (10.13%), number of monopodia per plant (8.96%), number of sympodia per plant (8.8%), micronaire (8.61%), boll weight (7.47%), 2.5% span length (7.24%) and lint yield per plant (6.38%) contributed maximum to the total divergence in the material (Table 6). The results are in accordance with Karunakar Raju *et al.* (2005).

In the principal component method, first 7 principal components altogether explained 87.47% of the variability (Table 7). The first component explained 33.06% of total variability and remaining ones accounted for progressive lesser and lesser amount of variation.

The PCA scores of 72 cotton genotypes were plotted in graph to get the 2D and 3D scattered diagram (Fig.2 and Fig. 3). All the genotypes were plotted for PC_1, PC_2, and PC_3 which clearly indicated the grouping of genotypes of same clusters falling nearer to each other. For example the genotypes

Fig-1: Dendrogram showing relationship of 92 cotton (*Gossypium hirsutum* L.) genotypes in nine clusters

Ward's Minimum Variance Dendrogram

Mahalanobis Euclidean² Distance

Table 2. Clustering of 72 cotton (*Gossypium hirsutum* L.) genotypes by Ward's minimum variance method

Cluster No.	No.of genotypes	Name of genotype (s)
I	5	L 769, TCH 1706, Surabhi, BS 3, CCHLS 2
II	8	ADL 903,NDL 762, LH 2076, ARB 760, CCH 510, ADB 320, NDLH 1755, P 72-9-37
III	8	RS 2461, CCH 1831, ADB 134, RSA 2495, CPD 787, GISV 103, CPD 814, L 761
IV	8	ARB 2001, ADB 39, LRA 5166, L 781, L 804, ADB 10050, L 778, ADB 250
V	21	SCS 601, H 1300, RB 557, P 514, F 2157, ARBH 813, CPD 812, HAGH 810, BS 144, Bihani 161, CSH 3114, CSH 7, CNDTS 52, HAG 1055, HS 276, LH 2111, CCH 226, AKH 2017, CA 100, F 2052, CNDTS 51
VI	11	CSH 3119, RHC 2004, NH 615, CNHO 12, L 604, GTHV 02/45, GSHV 152, GSHV 97/016, GSHV 01/1338, CPD 755, GISV 97/612
VII	6	GJHV 374, RS 2455, CNHO 23, TSH 9812, RAH 216, RAH 219
VIII	3	ARB 784, LPS 141, ARB 8906
I X	2	L 603, ARB 2005

Table 3. Clustering of 72 cotton (*Gossypium hirsutum* L.) genotypes by Tocher's minimum variance method

Cluster No.	No.of genotypes	Name of genotype (s)
I	41	Bihani 161, CSH 3114, LH 2111, CNDTS 52, CCH 226, CPD 812, P 514, HS 276, H 1300, SCS 601, RHC 2004, HAG 1055, HAGH 810, CSH 7, BS 144, ARBH 813, F 2157, AKH 2017, NH 615, CSH 3119, RB 557, LH 2076, CPD 755, GSHV 97/612, GSHV 512, GSHV 01/1338, CCH 510, NDL 762, CCHLS 2, BS 3, Surabhi, ARB 760, CNHO 23, L 604, GTHV 02/45, CPD 814, RSA 2495, CNHO 12, CPD 787, GISV 103, ADB 320
II	7	RS 2461, CCH 1831, ADB 134, TCH 1706, P 72-9-37, ADL 903, L761
III	19	LRA 5166, L 781, ADB 10050, L 778, ADB 250, L 804, ARB 2001, GJHV 374, NDLH 1755, ADB 39, RS 2455, TSH 9812, L 769, CNDTS 51 , RAH 216, RAH 219, F 2052, CA 100, ARB 784
IV	1	GISV 97/016
V	1	LPS 141
VI	1	ARB 8906
VII	1	L 603
VIII	1	ARB 2005

Table 4. Average intra- and inter-cluster D^2 values among eight cluster in 72 cotton (*Gossypium hirsutum* L.) genotypes

Cluster No.	I	II	III	IV	V	VI	VII	VIII
I	**17.39**	33.1	37.58	30.14	98.5	86.9	95.5	94.9
	(4.17)	(5.75)	(6.13)	(5.49)	(9.92)	(9.32)	(9.77)	(9.74)
II		**30.1**	59.4	48.3	143.8	127.5	117.1	117.9
		(5.49)	(7.71)	(6.95)	(11.99)	(11.29)	(10.82)	(10.86)
III			**41.99**	60.5	86.3	64.80	87.23	74.30
			(6.48)	(7.78)	(9.29)	(8.05)	(9.34)	(8.62)
IV				**0.00**	99.20	97.4	103.02	110.04
				(0.00)	(9.96)	(9.87)	(10.15)	(10.49)
V					**0.00**	34.23	34.45	89.87
					(0.00)	(5.85)	(5.87)	(9.48)
VI						**0.00**	51.4	55.95
						(0.00)	(7.17)	(7.48)
VII							**0.00**	54.46
							(0.00)	(7.38)
VIII								**0.00**
								(0.00)

Bold and diagonal values indicating intra-cluster distance , figures in parentheses are D values

Table 5. Average intra- and inter-cluster Eucledian^2 values among nine clusters in 72 cotton (*Gossypium hirsutum* L.) genotypes

Cluster No.	I	II	III	IV	V	VI	VII	VIII	IX
I	**46.54**	80.58	109.92	124.26	71.30	89.57	115.97	241.23	276.81
II		**54.59**	98.40	91.25	84.12	93.83	128.46	247.16	253.45
III			**77.71**	156.11	91.51	91.10	196.91	387.04	389.86
IV				**66.66**	107.46	152.83	127.61	204.45	225.89
V					**42.66**	61.93	87.50	245.08	275.87
VI						**44.82**	113.30	277.40	316.30
VII							**67.35**	145.96	212.61
VIII								**105.62**	184.39
IX									**163.39**

Bold and diagonal values indicating intra-cluster distance

Table 6. Contribution of different characters towards genetic divergence in 72 genotypes of cotton (*Gossypium hirsutum* L.)

Character	% contribution to- wards divergence	Times ranked first
Days to 50% flowering	0.00	0
Plant heigtht	10.13	259
Number of monopodia per plant	8.96	229
Number of sympodia per plant	8.80	225
Number of bolls per plant	35.49	907
Boll weight	7.47	191
Ginning outturn	0.39	10
Seed index	2.70	69
Lint index	0.70	18
2.5% span length	7.24	185
Micronaire	8.61	220
Bundle strength	0.20	5
Uniformity ratio	0.08	2
Elongation(%)	1.53	39
Fibre quality index	0.31	8
Count strength product	1.02	26
Lint yield	6.38	163
Seed cotton yield per plant	0.00	0

Table 7. Eigen values, proportion of the total variance represented by first seven principal components, cumulative per cent variance and component loading of different Charcters in cotton (*Gossypium hirsutum* L.)

|	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7
Eigen Value (Root)	445.82	221.89	144.81	117.31	101.81	77.19	70.52
% Variance explained	33.06	16.45	10.74	8.70	7.55	5.72	5.23
Comulative variance explained	33.06	49.52	60.26	68.97	76.52	82.24	87.47
1 Days to 50% flowering	0.06	0.05	0.004	0.05	0.03	0.09	0.01
2 Plant heigtht	0.03	-0.35	-0.39	0.31	0.62	-0.09	0.01
3 Number of monopodia per plant	0.22	0.34	-0.07	0.64	0.05	-0.07	0.27
4 Number of sympodia per plant	-0.14	-0.38	-0.59	0.09	0.47	0.09	-0.19
5 Number of bolls per plant	0.88	-0.16	-0.14	-0.19	0.007	0.11	0.07
6 Boll weight	-0.08	0.38	-0.53	-0.39	0.05	-0.13	0.22
7 Ginning outturn	-0.13	0.03	-0.03	0.17	0.21	0.31	-0.07
8 Seed index	-0.20	0.07	-0.12	-0.26	0.39	-0.13	0.01
9 Lint index	-0.05	0.08	-0.19	0.16	0.002	0.02	-0.11
10 2.5% span length	0.04	0.22	0.08	0.16	0.27	0.23	-0.54
11 Micronaire	-0.13	-0.41	0.20	0.17	0.06	-0.41	0.36
12 Bundle strength	-0.15	0.04	-0.08	0.11	0.13	0.01	-0.10
13 Uniformity ratio	-0.02	-0.04	0.02	0.04	0.05	0.15	0.02
14 Elongation(%)	-0.009	-0.35	0.019	-0.24	0.03	0.09	-0.19
15 Fibre quality index	0.02	-0.22	0.01	0.04	0.19	0.05	-0.14
16 Count strength product	0.01	0.008	0.12	0.08	0.09	-0.07	-0.04
17 Lint yield	0.15	0.08	-0.05	0.05	0.18	-0.64	-0.55
18 Seed cotton yield per plant	0.07	0.06	0.06	-0.06	0.01	-0.35	-0.09

PC = principal component

2D Plot
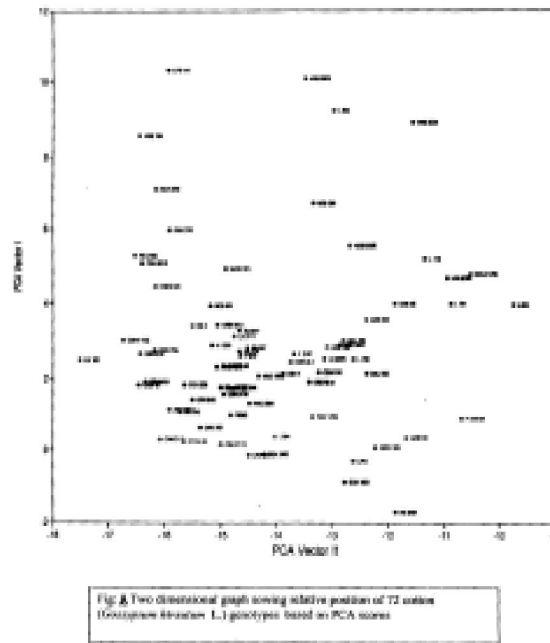
Fig 2. Two dimensional graph showing relative position of 72 Cotton (*Gossypium hirsutum* L.) genotypes based on PCA scores.
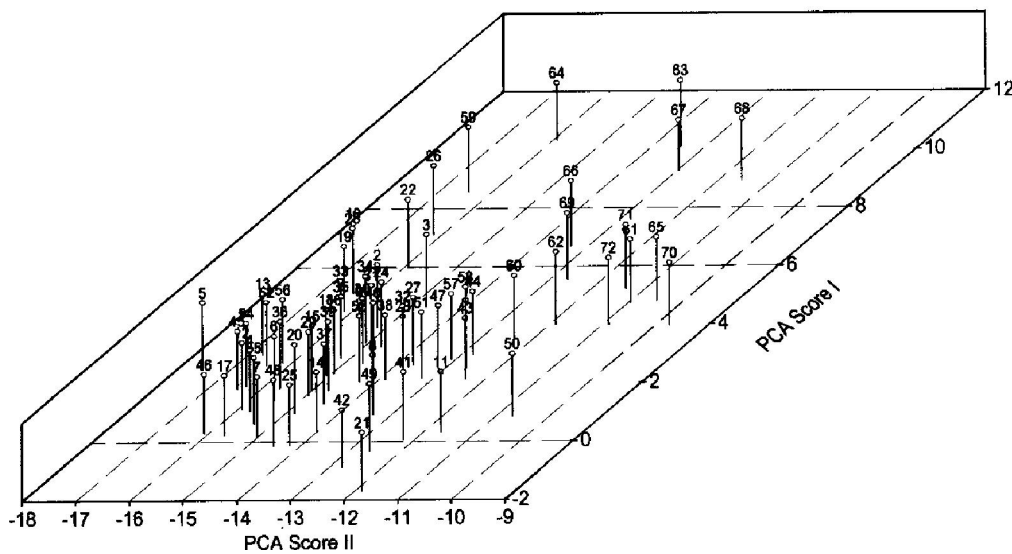
3D Plot

Fig 3. Three dimensional graph showing relative position of 72 Cotton (*Gossypium hirsutum* L.) genotypes based on PCA scores.

RS 2455 and TSH 9812 of cluster VII fall nearer to each other on the positive axis of PC_1. While genotypes ARB 2005 (Cluster IX) and L 603 (Cluster IX) fall away from each other indicating their specificity in clustering. The PC_1 reflected high component loading of important yield contributing character ie., number of bolls per plant while PC_2 reflected significant loading of boll weight. PC_4 reflected positive significant loadings of plant height and number of monopodia per plant.

Agglomerative cluster analysis revealed that wide genetic distance between cluster III and XI followed by cluster III and VIII. Therefore during hybridization programme, selection of parents from these clusters will produce superior segreagants.

The results of hierarchial cluster analysis and PCA confirmed the findings of each other. Genotypes belonging to a common cluster have fallen nearer to each other and vice-bersa and this confirms the results of cluster analysis.

All the three methods of grouping revealed a single concept of non-correspondence of genetic and geographical diversities. All the three methods of classifying genotypes into different groups are equally useful but hierarchial cluster analysis gave an additional advantage of identifying sub-cluster of the major groups at different levels so that each small group can be critically analysed.

LITERATURE CITED

Altaher A F and Singh R P 2003b. Genetic diversity studies in upland cotton (*Gossypium hirsutum* L.). Journal of the Indian Society for Cotton Improvement 28(3): 158-163.

Anderberg M R 1993. Cluster analysis for application. Academic Press, New York.

Brown J S 1991. principal component and cluster analysis of cotton cultivars variability across the U.S cotton belt. Crop Science 31: 915-922.

Jackson J E 1991. A user's guide to Principal component .John Wiley and sons New York.

Karunakar Raju G B , Chenga Reddy V and Panduranga Rao C 2005. Multivariate analysis of genetic diversity in upland cotton . The Andhra Agricultural Journal , 52 (1 & 2) 66-74.

Mahalanobis P C 1928. A statistical study at Chinese head measurement. Journal of Asiatic Society of Bengal 25: 301-307.

Murleedhar S. Aski and Murthy J S V 2005. Mahalanobis' D^2 and Principal component analysis of genetic diversity in tetraploid cotton. The Andhra Agricultural Journal 52 (1&2) : 99-104.

Murthy B R and Arunachalam V 1966. The nature of divergence in relation to breeding systems in some crop plants. Indian Journal of Genetics and Plant Breeding 26:188-198.