Electronic Journal of Plant Breeding, 6(4): 1019-1026 (Dec- 2015)

ISSN 0975-928X

**Research Article**

# Multivariate Analysis in Upland Cotton (*Gossypium hirsutum* L.)

**K. Bayyapu Reddy*, V. Chenga Reddy, Lal Ahamed M., T. C. M. Naidu and V. Srinivasarao**
Department of Genetics and Plant Breeding, Agricultural College, Bapatla 522101, Andhra Pradesh.
*E-mail: drkaipu3@gmail.com.

(Received: 14th May 2015; Accepted: 13th Jul 2015)

## Abstract

An experiment was conducted with 63 genotypes of cotton to assess the genetic diversity for seventeen characters at Regional Agricultural Research Station, Lam Farm, Guntur, Andhra Pradesh. The 63 genotypes were grouped into eight clusters based on hierarchial cluster analysis. Among the clusters, cluster I was the largest with 13 genotypes followed by cluster VI with 11 genotypes. Cluster VII had minimum intra cluster Euclidean$^2$ distance value while, the inter cluster Euclidean$^2$ distance was highest between the clusters VI and VIII indicating the usefulness of the genotypes of these clusters in the exploitation of heterosis. In principal component analysis, first seven principal components with eigen value more than one contributed 84.00% towards the total variability. $PC_1$ contributed maximum towards the total variability (23.799).

**Key words**
Hierarchical cluster analysis, principal component analysis, cotton.

## Introduction

Upland Cotton (*Gossypium hirsutum* L.) is a predominant species of cotton cultivated mainly for its lint in more than 80 countries of world. By virtue of its wider adaptability, it is grown in irrigated as well as in rainfed conditions. In India, it occupies about 75 per cent of the total cotton area of the country with a contribution of 85 per cent to the national production. The varieties/hybrids of this species played a significant role in achieving self sufficiency in cotton production of the country. In cotton improvement programmes, identification of desirable genotypes for the exploitation of heterosis is of prime importance and normally, divergence studies were carried out by using the Mahalanobis D$^2$ analysis and cluster analysis. Cluster analysis is of prime importance because of in depth analysis and formation of sub clusters within the clusters.

The prime objective of the present study was to analyze the genetic diversity among the 63 genotypes of cotton and to classify the genotypes into different groups and exploiting them in the breeding programmes for yield and quality traits improvement.

## Materials and methods

The present study was conducted with 63 genotypes obtained from all over India in randomized block design with three replications following 120 x 60 cm spacing during *kharif*, 2012-13 at Regional Agricultural Research Station, Lam Farm, Andhra Pradesh. Each plot consisted of three rows of 6 m length and observations were recorded on ten randomly selected plants from each genotype per replication for 17 characters *viz*., plant height (cm), no. of monopodia plant$^{-1}$, no. of sympodia plant$^{-1}$, no. of bolls plant$^{-1}$, boll weight (g), relative chlorophyll content, seed index (g), lint index (g), seed cotton yield plant$^{-1}$ (g) and lint yield plant$^{-1}$ (g) along with the traits, days to 50% flowering, ginning out turn (%), 2.5% span length (mm), micronaire value ($10^{-6}$g/inch), bundle strength (g/tex), uniformity ratio and elongation (%) which were recorded on plot basis. Agglomerative hierarchical clustering technique (Ward's minimum variance) as given by Anderberg (1993) was followed for clustering of genotypes while principal component analysis was performed as per Jackson (1991).

## Results and discussion

The analysis of variance (ANOVA) revealed highly significant differences among 63 genotypes for 17 characters indicating the existence of sufficient variability among the genotypes for characters studied (Table 1). The 63 genotypes were grouped into 8 clusters and the distribution of genotypes into 8 clusters was at random (Table 2 and Figure 1). The random distribution of genotypes indicated absence of parallelism between geographical and genetic diversity. Similar results are also reported by Altaher and Singh (2003), Rajamani and Mallikarjuna Rao (2009), Vijaya Lakshmi *et al.* (2009), Venkateswarulu *et al.* (2010), Haritha and Lal Ahamed (2013), Asha *et al.* (2013) and Tulasi *et al. (*2014).

Among the clusters, cluster I was the largest with13 genotypes followed by clusters VI (11 genotypes), IV (10 genotypes), V and VII (8 genotypes), VIII (5 genotypes) and II and III (4 genotypes) each. The average intra- and inter-cluster Euclidean$^2$ distance were estimated based on Ward's minimum variance method and was represented in Table 3 and Figure 2. Of the 8 clusters formed, cluster VII had minimum intra cluster Euclidean$^2$ distance value of 94.708 followed by clusters IV (116.70), VI (134.86), III (137.07), VIII (164.14), I (196.76), V (272.49) and II (288.22) indicating larger variability within the cluster II compared to other clusters. The inter cluster Euclidean$^2$ distances ranged from 187.03 (between clusters IV and VII) to 867.172 (clusters VI and VIII). The genotypes from the clusters VI and VIII can be exploited in the breeding programmes for heterosis as these two clusters are highly divergent.

The cluster II, showed high mean values for most of the yield contributing traits like number of monopodia plant$^{-1}$ (1.781), number of bolls plant$^{-1}$ (47.208), boll weight (4.893), relative chlorophyll content (1.202), 2.5% span length (30.882), bundle strength (23.787), seed cotton yield plant$^{-1}$ (228.622) and lint yield plant$^{-1}$ (77.808) (Table 4). The genotypes of the cluster II (MCU 5, HYPS 152, G COT 16, NDLH 1938) could be utilized in cotton yield improvement programmes to incorporate the important yield improvement traits. High mean values for lint index (4.790), ginning out % (39.382) and micronaire (5.014) were recorded in cluster III. Cluster VI recorded high mean values for number of sympodia plant$^{-1}$ (18.173) and seed index (10.648). So, the genotypes from these clusters can be used for cotton fibre improvement programmes.

In the present study, the first seven principal components with eigen values more than one contributed 84.00 per cent towards the total variability. The 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 5. It was therefore inferred that the essential features of data set had been represented in the first seven principal components. The first principal component contributed maximum variability (23.799). The characters, number of monopodia plant$^{-1}$ (0.255), number of bolls plant$^{-1}$ (0.302), boll weight (0.254), 2.5% span length (0.339), micronaire (-0.309 ), bundle strength (0.314), fibre elongation % (-0.226), seed cotton yield plant$^{-1}$ (0.351) and lint yield plant$^{-1}$ (0.316) contributed maximum variance in the first principal component ($PC_1$) and signifying their importance in divergence.

Results of cluster analysis based on PCA scores were presented in the 2D scattered diagram (Figure 3). The genotypes falling in same cluster were present closer to each other in scattered diagram there by confirming the results of cluster analysis. Utilization of principal component analysis combined with clustering of Ward's minimum variance method in genetic divergence studies in cotton was supported by Rajamani and Mallikarjuna Rao (2009), Vijaya Lakshmi *et al.* (2009), Venkateswarulu *et al.* (2010), Haritha and Lal Ahamed (2013), Asha *et al.* (2013) and Tulasi *et al. (*2014).

Grouping of these genotypes revealed a single concept of non-correspondence of genetic divergence and geographic diversity. Hierarchical cluster analysis gave an advantage of identifying sub-clusters of major groups at different levels so that each small group can be critically analyzed and exploited in the breeding programme for incorporation of desirable traits.

## References

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

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

Asha, R., Lal Ahamed, M., Ratna Babu, D and Anil Kumar, P. 2013. Multivariate analysis in upland cotton (*Gossypium hirsutum* L.). *Madras Agric J.,* **100(4-6)**:333-335.

Haritha, T and Lal Ahamed, M. 2013. Multivariate analysis in upland cotton (*Gossypium hirsutum* L.). *Crop Res.,* **46(1, 2&3)**: 217-222.

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

Rajmani, S and Mallikarjuna Rao, Ch. 2009. Genetic divergence in upland cotton (*Gossypium hirsutum* L.*). The Andhra Agric J.,* **56(2)**: 181-185.

Tulasi, J., Lal Ahamed, M., Samba Murthy, J.S.V and Ashoka Rani, Y. 2014. Multivariate analysis in upland cotton (*Gossypium hirsutum* L.). *J of Cotton Res and Development*., **28(2)**: 191-194.

Venkateswarulu, K., Chenga Reddy V., Samba Murthy, J.S.V., Srinivasa Rao, V., Panduranga Rao, C., Siva Reddy, K.V and Sateesh Babu, J. 2010. Multivariate analysis of genetic diversity in upland cotton (*Gossypium hirsutum* L.). *The Andhra Agric J.,* **57 (3)**: 240-247.

Vijaya Lakshmi, G., Chenga Reddy, V., Panduranga Rao, C., Satish Babu, J and Srinivasulu, R. 2009. Multivariate analysis of genetic diversity in upland cotton (*Gossypium hirsutum* L.*). The Andhra Agric J.,* **56 (2)**: 8-15.

**Table 1. Analysis of variance for yield and yield components in cotton**

| Source | d.f. | Days to 50 % Flowering | Plant Height (cm) | No. of monopodia plant$^{-1}$ | No. of sympodia plant$^{-1}$ | No. of bolls plant$^{-1}$ | Boll weight (g) | Relative chlorophyll content | Seed index (g) | Lint index (g) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean squares | | | | | | | | |
| Replications | 2 | 1.973 | 44.638 | 0.099 | 4.680 | 16.218 | 0.104 | 0.021 | 0.211 | 0.044 |
| Treatments | 62 | 55.084** | 262.927** | 0.238** | 9.138** | 118.018** | 0.401** | 0.048** | 6.611** | 2.146** |
| Error | 124 | 0.785 | 15.858 | 0.047 | 1.633 | 6.492 | 0.073 | 0.008 | 0.074 | 0.021 |

| Source | d.f. | Ginning out turn (%) | 2.5% Span length (mm) | Micronaire value ($10^{-6}$ g/inch) | Budnle Strength (g/tex) | Uniformit y Ratio | Elongation (%) | Seed cotton yield plant$^{-1}$ (g) | Lint yield plant$^{-1}$ (g) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean squares | | | | | | | |
| Replications | 2 | 3.352 | 1.989 | 0.058 | 4.180 | 8.638 | 0.057 | 252.02 | 37.600 |
| Treatments | 62 | 36.132** | 21.727** | 1.255** | 10.898** | 10.438** | 0.089** | 3528.364** | 482.594** |
| Error | 124 | 1.176 | 0.765 | 0.021 | 1.454 | 3.070 | 0.020 | 160.730 | 14.475 |

****** Significance at 1% level, d.f = Degrees of freedom

**Table 2. Clustering of 63 cotton genotypes by Ward's minimum variance method**

| Cluster No | No. of genotypes | Name of genotype(s) |
|---|---|---|
| I | 13 | ARBH 2411, GJHV 500, MR 786, RHC 0811, L 808, RAH 103, ARBH 2004, NA 1325, L 804, L 389, L 604, BS 37, L 755 |
| II | 4 | MCU 5, HYPS 152, G COT 16, NDLH 1938 |
| III | 4 | BS 49, SCS 1002, RAH 1004, BS 51 SPI |
| IV | 10 | H 1442, CPD 867, F 2310, GJHV 09/02, CCH 11-1, TSH 301, CNH 44, CCH 2623, BS 27, L 762 |
| V | 8 | GSHV 162, BS 47, RS 2670, TCH 1741, L 799, L 761, KH 1101, PUSA 5760 |
| VI | 11 | GJHV 440, L 1008, CCH 11-2, LH 2170, L 770, L 765, L 1101, L 801, TCH 1740, SURABHI, L 788 |
| VII | 8 | HS 289, TCH 1728, F 2296, LH 2256, CA 105, RS 2669, GSHV 159, TSH 0250 |
| VIII | 5 | GISV 272, CNH 50, GJHV 08/70, CSH 3088, SHS 2-4 |

**Table 3. Average intra and inter- cluster Eucledian$^2$ values among the eight clusters in 63 cotton genotypes**

| Cluster No | I | II | III | IV | V | VI | VII | VIII |
|---|---|---|---|---|---|---|---|---|
| I | 196.767 | 429.697 | 396.841 | 250.36 | 397.126 | 358.674 | 278.077 | 396.713 |
| II | | 288.227 | 666.596 | 565.169 | 559.163 | 515.616 | 726.957 | 831.612 |
| III | | | 137.073 | 229.817 | 340.321 | 417.668 | 386.966 | 488.164 |
| IV | | | | 116.701 | 335.329 | 210.920 | 187.033 | 471.652 |
| V | | | | | 272.491 | 501.952 | 487.625 | 553.824 |
| VI | | | | | | 134.862 | 397.580 | 867.172 |
| VII | | | | | | | 94.708 | 311.329 |
| VIII | | | | | | | | 164.141 |

Bold and diagonal values indicate intra-cluster distance

**Table 4. Mean values of eight clusters from 63 cotton genotypes by Wards minimum variance method**

| | Days to 50 % Flowering | Plant Height cm | Monopodia/ Plant | Sympodia/ Plant | Bolls/ Plant | Boll Weight (g) | Chlorophyll Content | Seed Index | Lint Index | Ginning Out % | 2.5% Staple Length | Micronaire | Budnle Strength | Uniformity Ratio | Elongation % | Seed Cotton Yield/ Plant | Lint Yield/ Plant |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Cluster | 57.359 | 90.231 | 1.568 | 16.975 | 28.979 | 4.497 | 1.111 | 8.033 | 3.689 | 32.828 | 28.967 | 3.671 | 22.496 | 46.851 | 5.128 | 131.601 | 43.783 |
| 2 Cluster | 61.167 | 89.556 | 1.781 | 17.543 | 47.208 | 4.893 | 1.202 | 9.106 | 4.188 | 33.293 | 30.882 | 2.793 | 23.787 | 46.289 | 5.013 | 228.622 | 77.808 |
| 3 Cluster | 60.917 | 97.388 | 1.643 | 17.327 | 36.084 | 4.573 | 1.081 | 7.607 | 4.790 | 39.382 | 26.704 | 5.014 | 21.106 | 48.577 | 5.307 | 164.446 | 63.355 |
| 4 Cluster | 58.900 | 100.511 | 1.531 | 16.641 | 31.022 | 4.473 | 0.997 | 8.996 | 4.225 | 32.485 | 26.681 | 4.522 | 21.061 | 48.577 | 5.215 | 138.619 | 45.424 |
| 5 Cluster | 67.667 | 98.573 | 1.413 | 16.486 | 29.986 | 4.587 | 1.065 | 8.576 | 4.360 | 34.930 | 27.291 | 4.091 | 22.115 | 48.406 | 5.203 | 137.527 | 48.058 |
| 6 Cluster | 56.909 | 97.364 | 1.677 | 18.173 | 33.142 | 4.593 | 1.088 | 10.648 | 4.786 | 31.715 | 29.387 | 4.114 | 22.965 | 47.492 | 5.230 | 155.035 | 48.335 |
| 7 Cluster | 57.208 | 97.597 | 1.236 | 16.457 | 31.962 | 4.412 | 1.051 | 8.126 | 3.173 | 29.550 | 24.582 | 4.708 | 19.917 | 49.307 | 5.316 | 142.257 | 42.257 |
| 8 Cluster | 59.333 | 100.223 | 1.479 | 15.229 | 33.134 | 4.520 | 1.053 | 5.729 | 2.558 | 34.897 | 24.756 | 4.353 | 19.456 | 49.197 | 5.188 | 149.623 | 50.883 |

Note: Bold figures are minimum and maximum values

**Table 5. Eigen values, proportion of the variance represented by first seven principal components, cumulative per cent variance and component loading of different characters in cotton**

| | $PC_1$ | $PC_2$ | $PC_3$ | $PC_4$ | $PC_5$ | $PC_6$ | $PC_7$ |
|---|---|---|---|---|---|---|---|
| Eigene Value (Root) | 4.046 | 2.392 | 2.248 | 2.040 | 1.441 | 1.169 | 0.945 |
| % Var. Exp. | 23.799 | 14.069 | 13.224 | 11.999 | 8.474 | 6.877 | 5.561 |
| Cum. Var. Exp. | 23.799 | 37.868 | 51.092 | 63.092 | 71.566 | 78.442 | 84.004 |
| Days to 50 % Flowering | 0.116 | 0.174 | 0.505 | 0.009 | 0.162 | 0.081 | 0.327 |
| Plant Height (cm) | -0.091 | -0.025 | 0.329 | 0.220 | 0.225 | 0.188 | -0.641 |
| No. of monopodia plant$^{-1}$ | 0.255 | 0.078 | 0.319 | 0.080 | 0.178 | -0.230 | 0.198 |
| No. of sympodia plant$^{-1}$ | 0.161 | -0.177 | 0.096 | -0.040 | 0.252 | -0.550 | -0.471 |
| No. of bolls plant$^{-1}$ | 0.302 | -0.091 | -0.009 | 0.476 | -0.088 | 0.008 | 0.154 |
| Boll weight (g) | 0.254 | 0.175 | 0.204 | 0.208 | 0.035 | 0.250 | -0.031 |
| Relative chlorophyll content | 0.191 | 0.160 | -0.173 | -0.051 | -0.191 | 0.518 | -0.281 |
| Seed index (g) | 0.154 | -0.502 | -0.117 | -0.205 | 0.250 | 0.137 | 0.005 |
| Lint index (g) | 0.208 | -0.399 | 0.256 | -0.296 | -0.135 | 0.116 | -0.039 |
| Ginning out turn (%) | 0.072 | 0.171 | 0.401 | 0.022 | -0.509 | -0.111 | -0.201 |
| 2.5% span length (mm) | 0.339 | 0.234 | 0.046 | -0.174 | 0.266 | 0.030 | 0.101 |
| Micronaire value ($10^{-6}$ g/inch) | -0.309 | -0.302 | 0.203 | 0.204 | -0.024 | 0.042 | 0.203 |
| Bundle Strength (g/tex) | 0.314 | 0.269 | -0.225 | -0.144 | 0.348 | 0.023 | -0.033 |
| Uniformity Ratio | -0.199 | 0.286 | -0.232 | 0.391 | 0.121 | -0.205 | -0.081 |
| Elongation (%) | -0.226 | -0.147 | 0.105 | 0.280 | 0.448 | 0.352 | 0.043 |
| Seed cotton yield plant$^{-1}$ (g) | 0.351 | -0.174 | -0.182 | 0.299 | -0.170 | 0.139 | -0.081 |
| Lint yield plant$^{-1}$ (g) | 0.316 | -0.271 | -0.147 | 0.357 | -0.081 | -0.196 | 0.087 |

PC= Principal component

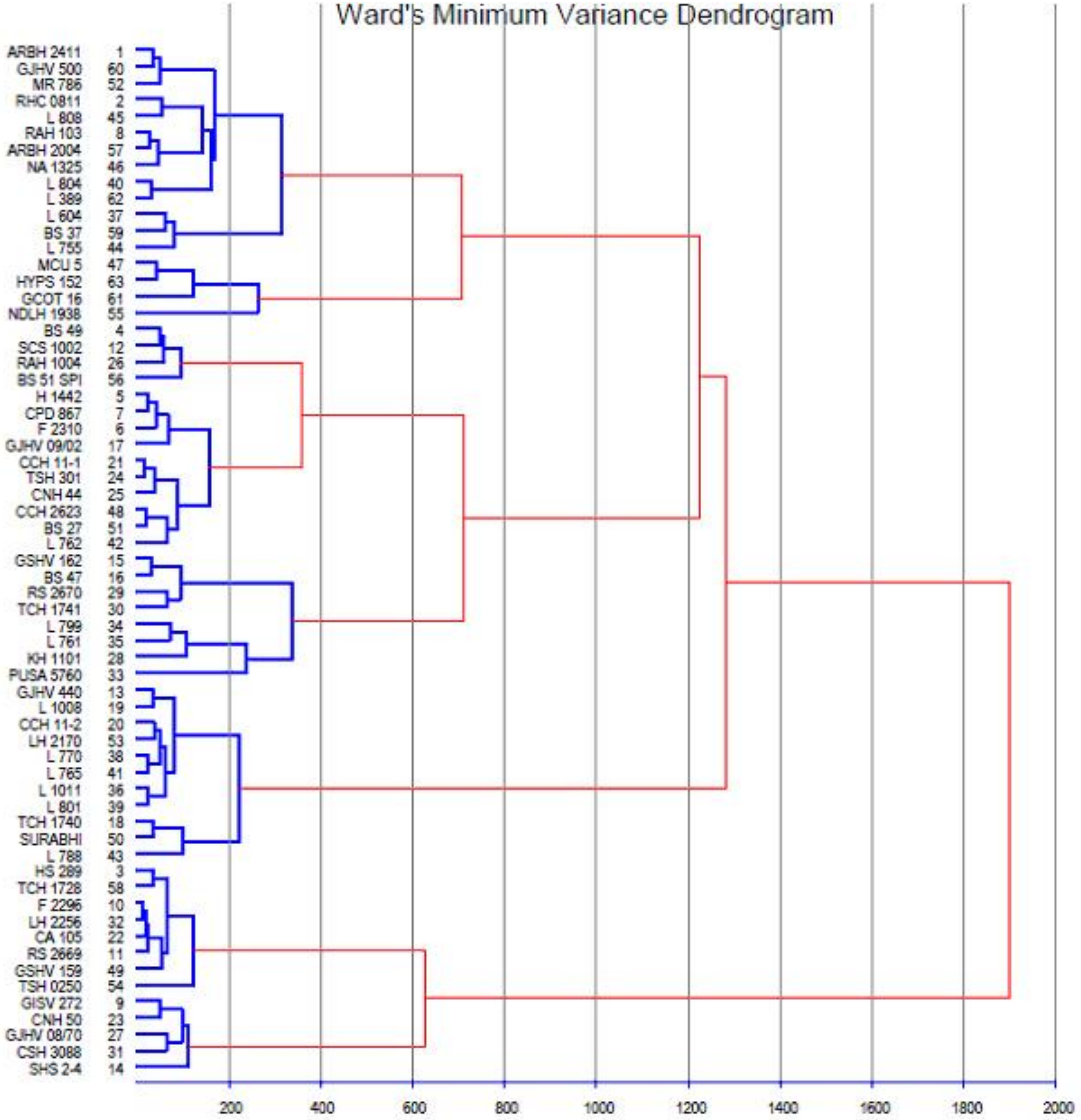

**Figure 1. Dendrogram showing relationship of 63 cotton genotypes grouped into eight clusters**

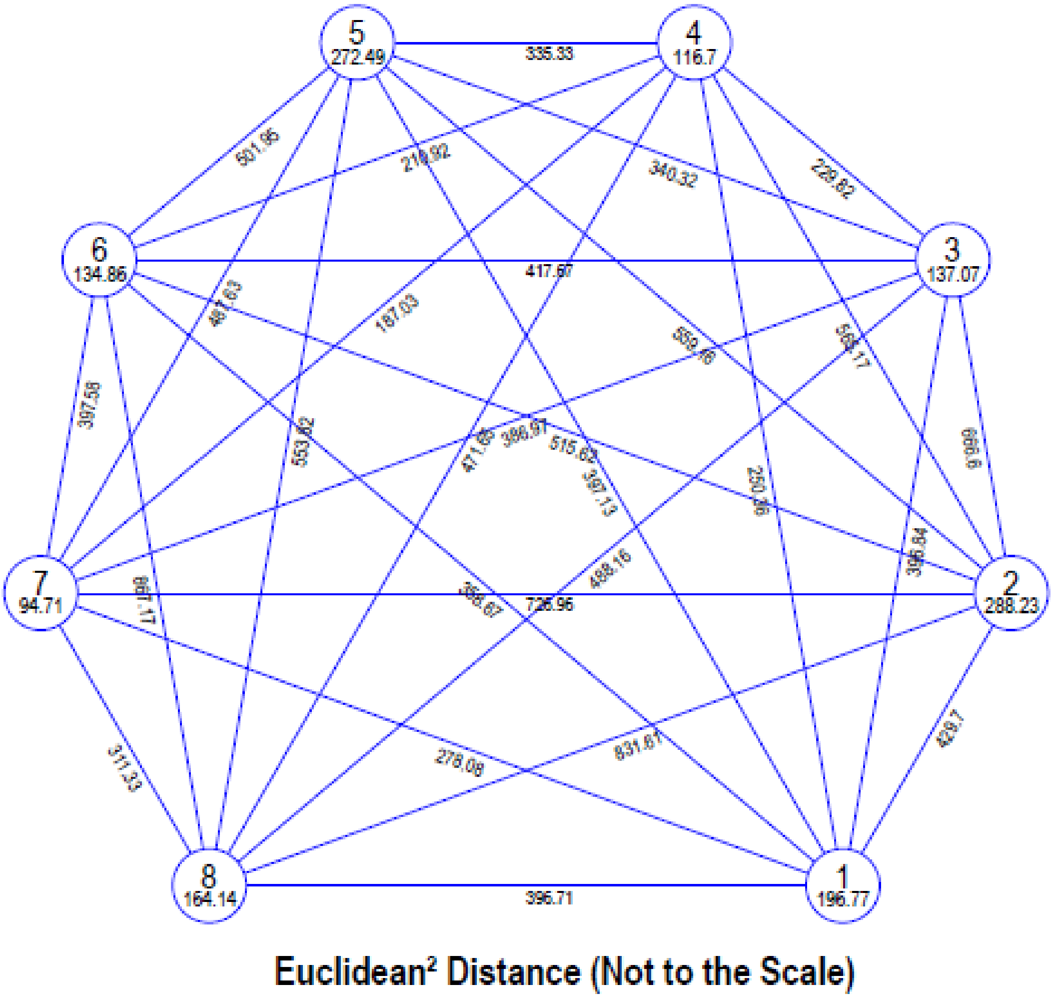

**Figure 2. Intra-and inter-cluster distance of 63 cotton genotypes in eight clusters based on Euclidean $^2$ distance.**

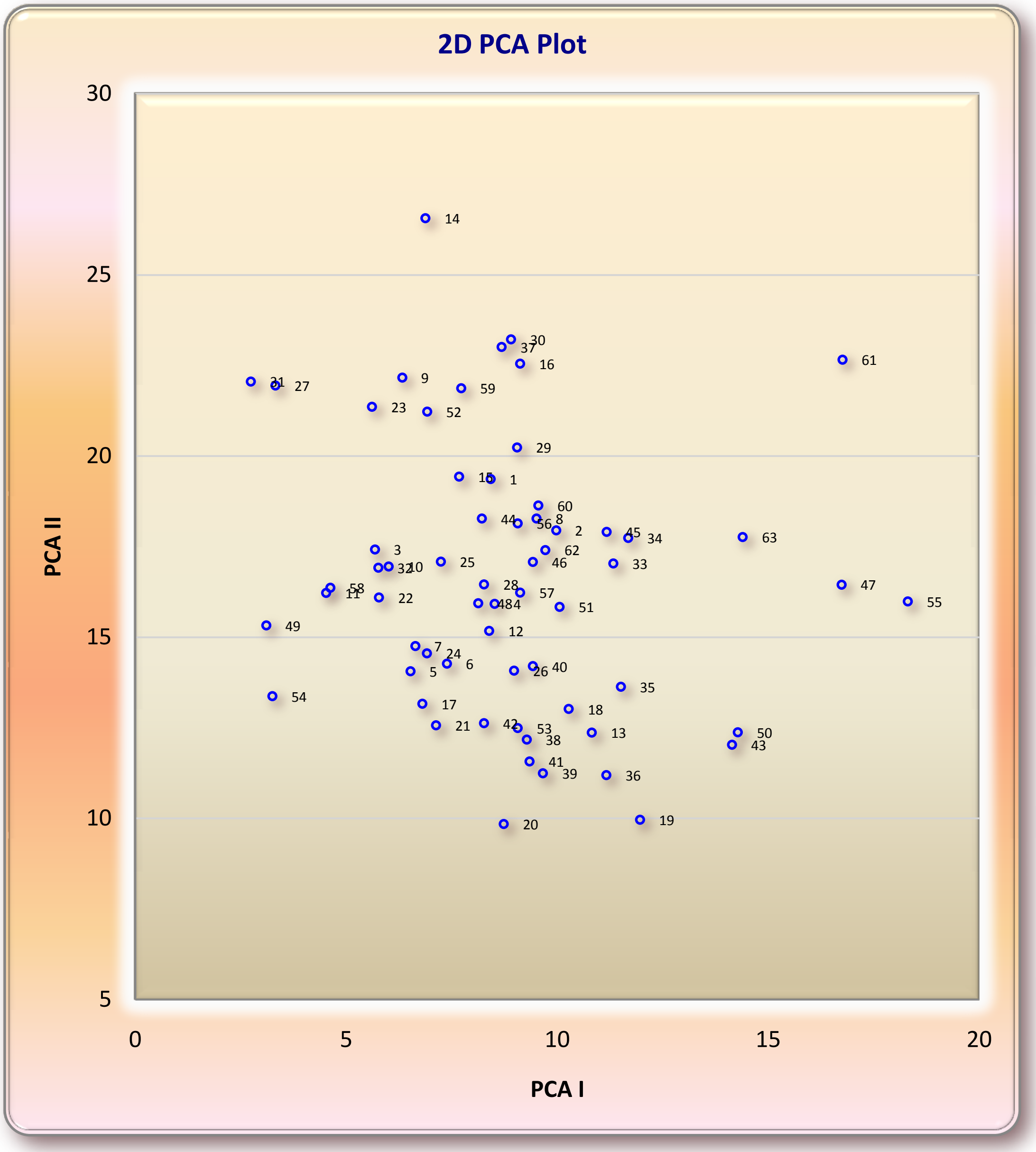

**Figure 3. Two dimensional graph showing relative position of 63 cotton genotypes based on PCA scores (Genotype numbers are as per the cluster diagram).**