Principal component and cluster analysis of yield and fibre quality traits in *Gossypium barbadense* L. germpalsm accessions

N. PREMALATHA*, M. KUMAR AND L. MAHALINGAM

Department of Cotton, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore – 641 003

Email: npremlatha@gmail.com

Abstract: The present study was conducted to evaluate the genetic divergence among 160 *Gossypium barbadense* germplasm accessions for sixteen biometric characters. The principal components analysis was used in maintaining and utilizing genetic resources which were divided the total variance into its components. Six components were found to possess Eigen value more than one. The PC1, PC2, PC3, PC4, PC5 and PC6 contributes 2.72, 2.23, 1.88, 1.71, 1.24 and 1.03 per cent of variability. These six PCs contributed 72.1 per cent in total variability amongst the accessions. The maximum contribution to genetic divergence was by the lint index followed by bolls and seed index. The genotypes were grouped into ten clusters. Among the ten clusters, cluster IV grouped with a maximum of 109 accessions followed by cluster II with 14 accessions, cluster III with 9 accessions, cluster VII and IX with 7 accessions each, cluster VI with 6 accessions, cluster V with 5 accessions and remaining clusters were solitary. Based on cluster analysis, the genotypes *viz.,* Barbados, BCS 59-10, 32/2R, CBS 201, TCB 209 and Giza 45 from different clusters and may be useful for further breeding programme to develop high yielding with good quality cultivars.

Key words: Cotton (*Gossypium barbadense* .L), cluster analysis, germplasm, genetic diversity, PCA

Cotton is the world's most utilized natural textile fibre and also important cash crop of India, which accounts for 60 per cent of total foreign exchange earnings through export of lint and value added cotton products. Many of the current high yielding cotton varieties do not possess the fibre quality desired by the textile industry. One of the ways to improve fibre quality and cotton yield is to transfer genes from *Gossypium barbadense* cultivars into high yielding *Gossypium hirsutum* cultivars. Successful breeding programme depend on good knowledge about genetic diversity, which exists in a crop germplasm. Variation in germplasm collections has been utilized for identifying desirable genotypes to enhance yield improvement. Evaluation of germplasm and quantification of genetic diversity are indispensable for a pragmatic use of plant genetic resources. Broadening genetic base and exploitation of genetic diversity is a prerequisite of a successful breeding program in crop

improvement (Shakeel *et al.,* 2015 and Ali *et al.,* 2016). Information of genetic diversity within and among closely related genotypes is essential for a cogent use of germplasm (Govindaraj *et al.,* 2015) It allows the selection of desirable parents for crossing to develop a genetically diverse population for selection. Principal component analysis (PCA) and cluster analysis is being employed by the researchers to find out the similarity among the genotypes for different traits and grouping the traits into different clusters and it helps in transforming the interdependent traits into a set of independent traits. Among all other biometrical techniques, PCA has the edge to assign only one group to each genotype at the same time and also depicts the importance of major contributor toward total diversity at each axis of differentiation. Principal component analysis was carried out to transform the inter dependent traits into a set of independent traits as well as to reduce the dimensionality of the data structure. Principal component analysis and

cluster analysis was accomplished by the researchers to find the similarity among the genotypes for the traits and their placement into different clusters. The cluster analysis verifies sufficient diversity for any group of genotypes and this diversity is important in selecting useful genotypes. The current study was aimed to access the genetic diversity and grouping of the accessions using Principal Component Analysis (PCA) and cluster analysis.

MATERIALS AND METHODS

The field experiment was conducted with the 160 *Gossypium barbadense* germplasm accessions during *kharif* 2018-2019 at Department of Cotton, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore in a randomized block design with two replications. Each genotype was sown in one row of 6 m length with spacing of 90 x 45 cm. Recommended package of practices and plant protection measures were adapted to raise a good crop. Observations on days to flowering, days to 50 per cent flowering, plant height, number of sympodia, boll weight, number of bolls, seed index, lint index, ginning outturn, seed cotton yield/plant were recorded on five randomly selected plants. Fibre quality characters *viz.,* fibre length (mm), fibre strength (g/tex), micronaire value (µg/inch), elongation percentage and uniformity ratio were analyzed under high volume instrument (HVI) at Department of Cotton, CPBG, TNAU, Coimbatore. The data were analyzed based on Agglomerative cluster analyzing using the software TNAUSTAT.

RESULTS AND DISCUSSION

Analysis of variance for different characters revealed high significant variations among the genotypes taken under study, ensuring high degree of genetic variability among the germplasm taken under study. Principal component analysis (PCA) explains the divergence of any germplasm. The phenotypic value of each trait measures the importance and contribution of each component to total variance, whereas, each coefficient of proper vectors indicates the degree of contribution of every original variable with which each principal component is associated. Higher the coefficients, regardless of the direction (positive or negative), the more effective they will be in discriminating between accessions. The Principal component analysis was done and the components having eigen values of above one were extracted and presented in the Table 1. Morphological traits like days to flowering, days to 50 per cent flowering, plant height, number of sympodia, number of bolls, boll weight, seed index, lint index, upper half mean length, micronaire value, fibre strength, uniformity ratio, elongation percentage, seed cotton yield and ginning outturn were subjected to principal component analysis (PCA) which revealed these first six components contributed 72.12 per cent of variability among 160 germplasm entries. These components were retained based on the Scree plot and threshold Eigen value greater than one. The Eigen value for PC1, PC2, PC3, PC4, PC5 and PC6 were 2.72, 2.23, 1.88, 1.71, 1.24 and 1.03 respectively (Table 1).

The first principal component contributed to 18.1per cent of total variability. Number of bolls (0.506) contributed the maximum variation in case of PC I followed by single plant yield/plant (0.493), sympodia (0.467), plant height (0.295), days to 50 per cent flowering (0.247) and lint index (0.177) and whereas fibre strength (-0.103), elongation percentage (-0.035) and uniformity ratio (-0.013) contributed negatively to this component. The second principal component contributed to 14.84 per cent of total variability. Higher variation to this PC II was contributed by fibre strength (0.467) followed by UHML (0.396) and

Table 1. Principal component analysis for *Gossypium barbadense* accessions with respect to various characters

Traits	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Days to flowering	0.234	-0.298	0.539	-0.210	-0.083	0.050
Days to 50 per cent flowering	0.247	-0.304	0.540	-0.185	-0.060	0.047
Plant height (cm)	0.294	-0.015	-0.195	-0.020	-0.154	-0.137
Monopodia	0.467	0.003	-0.150	0.190	-0.044	-0.118
Sympodia	0.506	0.166	-0.045	0.166	-0.092	0.012
Bolls	0.021	0.254	0.159	0.142	0.585	0.170
Boll weight (g)	0.177	0.176	-0.208	-0.613	0.072	-0.161
Lint index	0.032	0.280	0.118	-0.485	0.251	-0.363
Seed index	0.170	-0.128	-0.359	-0.313	-0.177	0.192
Ginning outturn (%)	0.493	0.020	0.016	0.231	0.173	0.034
Seed cotton yield (g)	-0.008	0.396	0.270	0.076	-0.021	-0.144
Upper half mean length (mm)	-0.103	0.468	0.200	0.004	-0.183	-0.077
Fibre strength (g/tex)	0.077	0.097	-0.018	-0.193	0.246	0.754
Micronaire value (µg/inch)	-0.013	0.234	0.177	0.073	-0.479	0.098
Uniformity ratio (%)	-0.035	0.\347	0.012	-0.163	-0.398	0.365
Eigen value	2.72	2.23	1.88	1.72	1.25	1.03
Variability (%)	18.10	14.84	12.55	11.45	8.33	6.86
Cumulative (%)	18.10	32.94	45.49	56.93	65.26	72.12

elongation percentage (0.347). The third principal component accounted to 12.55 per cent of total variability. Higher variation to PC III was contributed by days to 50 per cent flowering (0.540) followed by days to flowering (0.539) and UHML (0.269).

The fourth component contributes to 11.45 per cent of total variability higher variation to the PC4 was contributed by single plant yield/plant (0.231) followed by sympodia (0.190) where as lint index (-0.613), seed index (-0.485), ginning outturn (-0,313) and days to flowering (-0.210) contributes negatively to this component. The fifth principal component contributes to 8.33 per cent of total variability and the higher variability was contributed by boll weight (0.586) followed by seed index (0.251) and micronaire (0.249) while uniformity ratio (-0.479), elongation percentage (-0.398) and fibre strength (-0.183), contributes negatively. The sixth principal component contributed to 6.86 % of total variability where the higher variation was contributed by micronaire value (0.754) followed by elongation percentage

(0.365), ginning outturn (0.192) and boll weight (0.170). PCA revealed the relative contribution of yield and fibre quality traits in creating the variability. Sun *et al.*, (2019) provided a detailed phenotypic diversity description of a population using PCA analysis representing a wide range of upland cotton germplasm. Hence selection and utilization of genotypes based on these traits would be more effective in the selection for hybridization programme.

The characters contributing to maximum divergence were given greater emphasis on deciding the clusters for the selection of parents for further hybridization. The contribution of the various biometric traits towards the genetic variation is presented in the Table 2. Among the various traits, lint index (23.06 %) contributed maximum genetic variation followed by bolls (18.74 %) and seed index (16.06 %). Hence selection of parents based on the lint index and bolls/plant will generate more variability in breeding programme. Similarly the Scree plot for 16 quantitative traits is given in Fig 1. Scree plot explained the percentage of variation between the Eigen value and Principal components. It clearly

Table 2. Contribution of various quantitative traits towards genetic divergence

Characters	Times ranked Ist	Per cent contribution
Days to flowering	1070	0.84
Days to 50 per cent flowering	1	0.01
Plant height (cm)	1238	9.73
Sympodia	29	0.23
Bolls	2384	18.74
Boll weight (g)	456	3.58
Lint index	2933	23.06
Seed index	2043	16.06
Ginning outturn (%)	13	0.10
Seed cotton yield (g)	368	2.89
Upper half mean length (mm)	261	2.05
Fibre strength (g/tex)	8	0.06
Micronaire value (µg/inch)	1879	14.77
Elongation per cent	47	0.29
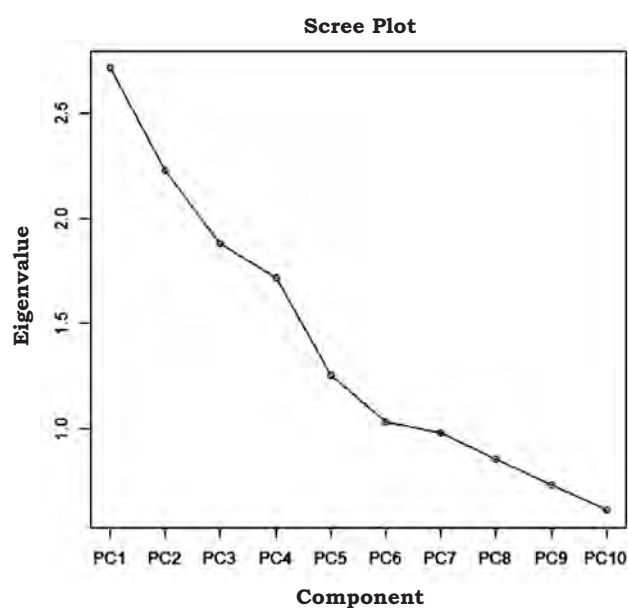
Scree Plot

*Fig 1. Scree plot of PCA for quantitative traits of germplasm

showed that the maximum variation was observed in PC1, PC2, PC3, PC4, PC5 and PC6.

Cluster analysis : Cluster analysis resulted in construction of a dendrogram on the basis of yield related traits and quality data by using Statistical tool for Agricultural Research software following Agglomerative Cluster method. The genotypes were grouped into ten clusters. Among the ten clusters, cluster IV grouped with a maximum of 109 accessions followed by cluster II with 14 accessions, cluster III with 9 accessions, cluster VII and IX with 7 accessions each, cluster VI with 6 accessions, cluster V with 5 accessions and remaining clusters are solitary. The dendrogram showed 10 clusters and the details of genotypes in each cluster are presented in the Table 3. Genotypes belonging to distant clusters may be used for exploiting the maximum genetic diversity. Khan *et al.,* 2010, Shakeel *et al.,* 2018 and Akter *et al.*, (2019) reported that cluster analysis was used to assess genetic diversity and to select genotypes for breeding programs to improve various traits. Grouping of accessions into different clusters aids for selection of genotypes were designed for future breeding program (Rathinavel, 2019).

The mean performances of 15 characters in ten clusters (Table 4) indicated considerable differences for all the characters between the clusters. Most of the characters showed distinct difference among the clusters. The cluster I had highest mean value for sympodia (15), bolls (22) and seed cotton yield/plant (102 g) and minimum cluster mean value for micronaire value (2.6 µg/inch). The results were in accordance with the findings of Sundar *et al.,* (2014). The cluster III had highest mean value for ginning outturn (36.3 %), lint index (5.2) and

Table 3. Cluster composition of different genotypes

Cluster number	Number of germplasm accessions	Name of the accessions
Cluster 1	1	Barbados
Cluster 2	14	Barx Xu, CBS 201, EC 97618, EC 97634, EC 101783, EC 111248, SB S79, 22/2R, 3530, C 6002-3, SIA-9, CCB-1, CCB-4
Cluster 3	9	BCS 9-45, CBS148, CBS 202, EC 126597, EC 136450, EC 136452/B, Giza 45, DB-5, RAB 107, EC 136452/B
Cluster 4	109	BCS 9- 7 6, BCS 10-91, BCS 22- 73, CBS-34, EB(A), EC 9256, EC 9260, EC 97619, EC 97620, EC 97623, EC 97624, EC 97626 , EC 97627, EC 97631, EC 97632/4, EC 97633, EC 97635 , EC 97638, EC 97639 ,EC 97640, EC 98254, EC 10178 , EC 101786, EC 111265, EC 131979, EC 136451, EC 136452/A, EC 136453, Giza 12, Giza 7, Giza, G-122-A-45, Egyptian, EC 13758, ERB 4530, ERB 4492, Giza 1467, Giza 2767, Giza 14647, PSH, Pima S2, Pima SI, Orissa SI, 49, BS 66, Seaberry 528, Sea Broke, Sealand 339, SI and rews, SIV 135-6, SIV 135-8, SIV RB, Sudan G-45, Sudan G-55, Sujata, Tadla, Tadla (P), Tadla 2, 19/61, 19/2R 19/2, 18-4, 17/3A, 16/2W, 16/2R, 13-2, 4-Dec, 85/2, 83/1W, 82/2R, 76/3, 46/5W, 32/1W, 26/5W, 25-1-3, 6002-1, 5904-U , 5904-S, 5391-T, 3530 U, 3415 B, 3415 A, TCB 355 , TCB 472/5, TNB -1, TCB 371, TCB 372, TCB 377, TCB 404, Suvin, SIA-4, SIA-5, 7822-5, SB 1085-6, CCB-2, CCB-3, CCB-6, CCB-7, GSB 39, RHCB 01, NDGB 5
Cluster 5	1	BCS 59-10
Cluster 6	6	EC 9257, EC 97625, EC 97628, EC 111254, Giza 7(A), SB 1085-6
Cluster 7	7	EC 97632, EC 101785, ERB 4488, SBS (YF), 17/3B, 6002, CCB-5
Cluster 8	5	Pima 34, Menoubi, Monspi Serrate, IBSI, TCB 365
Cluster 9	7	K 3530, St.Kitts, Sudan G-5-S, TCB 472/4, ERB 13754, TCB 209, EC 9261(B)
Cluster10	1	32/2R

Table 4. Cluster mean for fifteen characters in 160 *Gossypium barbadense* L. germplasm accessions

Characters	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII	Cluster VIII	Cluster IX	Cluster X
Days to flowering	58	64	68	63	66	59	71	66	58	63
Days to 50 per cent flowering	66	70	75	69	71	65.6	77	72	63	68
Plant height (cm)	73.6	98.2	127.5	108.6	98.3	126.5	102.3	94.4	101.8	103.4
Sympodia	15	12	15	13	14	15	11	9	12	11
Bolls	22	16	20	15	15	19	15	8	14	12
Boll weight (g)	3.7	4.0	3.7	3.6	4.3	3.6	3.8	3.6	3.1	4.4
Lint index	4.2	3.8	5.2	4.7	4.7	3.8	5.2	4.2	4.0	3.5
Seed index	9.1	8.7	8.8	9.1	9.9	7.5	9.2	9.3	8.1	10.3
Ginning outturn (%)	32.8	30.1	36.3	33.5	34.0	32.8	36.2	31.7	33.3	27.6
Seed cotton yield (g)	102	65.9	82.2	57.1	69.0	75.5	60.9	30.8	46.1	56.0
Upper half mean length (mm)	35.2	34.8	32.0	33.0	35.3	31.4	32.6	32.6	32.1	34.8
Fibre strength (g/tex)	29.7	31.3	28.3	29.0	29.2	27.6	29.6	28.1	29.4	36.1
Micronaire value (µg/inch)	2.6	2.7	3.4	3.2	3.1	3.0	2.9	3.3	2.8	3.8
Uniformity ratio (%)	84.8	84.6	82.9	83.6	69.8	84.0	83.0	83.8	82.3	82.3
Elongation per cent	5.6	6.0	6.0	5.8	6.0	6.0	6.0	6.3	6.3	6.3

plant height (128 cm). The genotypes with maximum lint yield potential can be developed either in the form of hybrids or as transgresive segregants (Sharma *et al.*, (2016)). The cluster V recorded maximum mean value for Upper half mean length (35.3 mm) and seed index (9.9). The

cluster IX recorded minimum mean value for days to flowering (58 days) and days to 50 per cent flowering (64 days). The cluster X recorded maximum mean value for boll weight (4.4 g) and cluster II had highest mean value for fibre strength (36.1 g/tex). The classification of germplasm accessions into different clusters, aids for selection of genotypes for future breeding program. Ammer *et al.,* 2019 reported that characters including seed cotton yield, bolls, staple length and seed index exhibited higher values among different clusters and can be further used as selection criteria. Based on cluster analysis, BCS 9-10 in cluster V had recorded maximum mean value for upper half mean length (35.4 mm) and Barbados in cluster I recorded highest seed cotton yield of 102 g/plant. Accession 32/2R in cluster X recorded highest boll weight of 4.4 g and fibre strength of 36.1 g/tex. These accessions are from different clusters which are in diverse in nature. The germplasm accessions which are in diverse nature may be useful for transfer of the desired gene for cotton fibre quality improvement, suitable for different ambiances. Cluster III had nine accessions *viz.,* BCS 9-45, CBS 148, CBS 202, EC 126597, EC 136450, EC 136452/B, Giza 45, DB 5, RAB 107, EC 136452/B had recorded maximum mean ginning outturn of 36.3 per cent.

On the basis of mean values of different clusters desirable lines for a particular trait may be selected from a specific cluster such as seed cotton yield (Cluster I), boll weight (XI), ginning outturn (VII), seed index (X), lint index (III, VII). By combining the genotypes with high seed cotton yield (cluster I) with genotypes having high ginning outturn (cluster VII), the genotypes with maximum lint yield potential can be developed either in the form of hybrids or as transgressive segregants. The hybrids with even better yield potential can be developed by crossing such lines from cluster I with the lines of cluster XI. The genotypes *viz.,* Barbados, BCS

59-10, 32/2R, CBS 201, TCB 209 and Giza 45 from different clusters and may be useful for further breeding programme to develop high yielding with good quality cultivars.

REFERENCE

Aktar, T., Islam, A.K.M.A., RasuL,M.G., Kundu, S., Khalequzzaman and Ahmed, J.U. 2019. Evaluation of genetic diversity in short duration cotton (*Gossypium hirsutum* L.). *J. Cotton Res.* **2** :1- 6.

Ali, I, Shakeel, A, Saeed, A, Nazeer, W, Zia, ZU, Ahmad, S, Mahmood, K and Malik W. 2016. Combining ability analysis and heteroitic studies for within – boll yield components and fibre quality in cotton. *J. Anim. Plant Sci.,* **26** : 156-62.

Ammer, Hussain Jarwar, Xiaoyan, Wang, Muhammad Shahid, Iqbal, Zareen, Safraz, Long, Wang, Qifeng, Ma and Fan, Shuli. 2019. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality triats in cotton cultivars. *Pak. J. Bot.,* **51** :1143-48.

Govindaraj, M.M, Vetriventhan, M., Srinivasan, M.,2015. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Gent.Res. Int.,* (*Review article*) pp. 1-15.

Rathinavel, K.2019. Principal component analysis with quantitative traits in extant cotton varieties (*Gossypium hirsutum* L.) and parental lines for diversity. *Curr. Agri. Res.* **6** : 54-64.

Shakeel, A., Azhar, M.T., Ali, T., Qurat-Ul-Ain, Zia,Z.U., Anum, W., Ammar, A and Zafar, A., 2018. Genetic diversity for seed cotton yield parameters, protein and oil contents among various *Bt.* Cotton cultivars. *Int. J. Biosci.,* **12** : 242-51.

Shakeel, A.I., Talib, M., Rashid, A., Saeed, K., Ziaf. M.F and Saleem.V. 2015. Genetic diversity among upland cotton genotypes for quality and yield related traits. *Pak. J. Agri. Sci.,* **52** : 73-77.

Sharma, Pankaj, Sohu R.S. and Pathak Dharminder. 2016. Assessment of genetic diversity in *Gossypium hirsutum* L. germplasm following Mahalanodis D^2 analysis. *J. Cotton Res. Dev.,* **30** : 1-5.

Sun Zhengwen, Wang Xinfen, Liu Zhengwen, Gu Qishen, Zhang Yan, Ll Zhikun, KE Huifeng, Yang Jun, Wu Jinhua, Wu Liqiang, Zhang Guiyin and MA Zhiying. 2019. Evaluation of the genetic diversity of fibre quality traits in upland cotton (*Gossypium hirsutum* L.) inferred from phenotypic variations. *J. Cotton Res.,* **2** :1-8.

Sundar, Rajeev, Patil, Shreekanth S., Ranganath, H.M., Srivalli, P., Manjula, S.M. and Pranesh, K.J. 2014. Genetic diversity analysis in cotton (*Gossypium hirsutum* L.) based on morphological traits and microsatellite markers. *Int. J. Basic Applied Biol.,* **1**: 19-22.