Electronic Journal of Plant Breeding

Research Article

Evaluation of desi cotton (*Gossypium arboreum* L.) germplasm using qualitative descriptors and principal component analysis

A. Krishnamoorthi[1], S. Hari Ramakrishnan[2], N. Premalatha[3], N. Manikanda Boopathi[4] and V. Thiruvengadam[1*]

[1]Department of Plant Genetic Resources, Centre for Plant Breeding and Genetics,
[2]Agricultural Research Station, Kovilpatti,
[3]Department of Cotton, Centre for Plant Breeding and Genetics,
[4]Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, TNAU, Coimbatore, Tamil Nadu, India
*E-Mail: thirugen@gmail.com

Abstract
In the present study, a total of 155 desi cotton accessions (*Gossypium arboreum* L.) were evaluated using six qualitative traits in an augmented design. The distribution of genotypes categorized under the selected traits and their states were computed for all the genotypes. Principal Component Analysis (PCA) identified three principal components accounting for 56.91% of the variability in the *arboreum* germplasm with PC 1 contributing 20.51% and PC 2 contributing 18.62%. The results of the eigen vectors of six qualitative traits indicated that the traits namely anther colour, plant body colour and boll shape could serve as distinguishable morphological markers for characterization of *G. arboreum* genotypes. Correlation studies revealed significant positive association among the traits under study. Scatter plot between PC 1 and PC 2 showed the convex of the hull occupied by the genotypes namely DWDA 1701, NDLA 3086, JLA 0610, CAN 1038, RG5 21, HD 544, PA 778, GAM 261, FDK2 95, K 12 and CIsa 90 presenting their diverse nature among the genotypes. The clustering analysis separated 155 *G. arboreum* accessions into five major clusters and maximum genotypes were placed in Cluster V. Combined results of PCA and cluster analysis selected DWDA 1701, GAM2 60, TAS 130, CISA 1793 and RG 690 as promising genotypes with highest degree of divergence for future desi cotton breeding research.

Key words
Desi cotton, qualitative traits, principal component analysis, scatter plot, cluster analysis

INTRODUCTION

Cotton is the king of fiber crops and is commonly known as the "White gold". It is the main economic and industrial crop of many countries supplying a prime raw material for textile mills. China, India, United States and Pakistan are the leading cotton producing nations accounting for more than 70% of the world production. There are more than 50 species belonging to the genus *Gossypium*. The cultivated species of cotton comprises of two AD genome tetraploids (2n=4x=52) namely *G. hirsutum* and *G. barbadense*, and two A genome diploid (2n=2x=26) namely *G. arboreum* and *G. herbaceum*. Although *G. hirsutum* cultivars contributed about 95% of current world production of 118 million bales and produce superior quality fiber with higher lint yield, yet they are susceptible for many biotic and abiotic stresses. In this context, diploid cotton such as *G. arboreum* would be the most appropriate choice for cotton breeders as it is known to exhibit desirable traits such as suitability to low input management practices,

adaptability to extreme environmental conditions like drought and salinity, disease and insect pest resistance etc., Though lacking in superior fiber quality standards required for textile industry, diploid cotton still produce short stable fibers which are highly suited for medical or surgical purposes.

Over the last two decades, the area under diploid cotton cultivation in India is on the decline primarily due to the development of high yielding Upland cotton varieties and hybrids and also the introduction of *Bt* transgenic cotton. During 2000, the area of *arboreum* and *herbaceum* had reduced to 1.46 m. ha and 0.95 m. ha, respectively. At present, only less than 3 per cent of area is occupied by the desi cotton (Kranthi, 2015). Furthermore, the crop breeding research in desi cotton is very limited in India that resulted in a slow progress in varietal development (Vafaie-Tabar *et al.*, 2003). Therefore, focused research efforts are needed to revitalize the diploid cotton improvement for broadening the available gene pool for specific traits. This could be made possible with the proper augmentation and characterization of diploid cotton germplasm resources for various agro-morphological traits.

Effective management of plant genetic resources depends on the ability to distinguish the genotypes and to identify elite lines from the vast germplasm resources (Laurentin, 2009) which needs the basic characterization information. Characterization and evaluation of germplasm is a broad based approach for identification of accessions with desirable traits to be employed directly as cultivars or trait donors for using in the crop improvement programs by the plant breeders (Zada *et al*., 2013). Besides deciphering the genetic diversity of crop germplasm utilizing quantitative traits, variations present in the germplasm collections could also be carried out by employing qualitative traits. Unlike quantitative traits, the qualitative traits are widely preferred for morphological characterization as they are relatively less influenced by the environment and clearly categorized into discrete phenotypic classes. Furthermore, DUS (distinctiveness, uniformity and stability) characterization of the genotypes with these morphological traits assumes significance in fulfilling the legal aspects of "The Protection of Plant Varieties and Farmers' Right Act (PPV&FR Act) 2001" especially on issues such as plant variety registration and protection of plant breeder's rights.

Therefore, it is essential to identify key diagnostic traits of different genotypes that allows the breeder to select genotypes with distinct characters. In cotton, a number of works have earlier reported the identification of desired genotypes using various leaf, stem and floral morphological traits (Iqbal *et al.,* 2006; Sangwan *et al.,* 2008; Ranjan *et al.,* 2014; Manivanan *et al*., 2018). As new germplasm accessions from different regions of the country are routinely been incorporated in various breeding trials every year, it becomes imperative to characterize those resources continuously to facilitate the successful identification of promising lines for future crop

breeding research. Hence, the present study was carried out with the objectives to identify the traits suitable for DUS testing and to analyze the genetic diversity existing in the *G. arboreum* accessions through morphological characterization using qualitative descriptors.

MATERIAL AND METHODS
A total of 150 diverse accessions of *G. arboreum* along with five standard check varieties namely LD1019, LD949, DLSA17, K11, and K12 were obtained from Agricultural Research Station, Kovilpatti and used for the present study. The germplasm accessions were raised in the Experimental Farm of Department of Cotton, Centre for Plant Breeding and Genetics, TNAU, Coimbatore during Kharif 2019 by adopting augmented block design. A single row of each accession was planted with a row length of 3 m and row to row spacing of 90 cm. All agronomical practices were followed as recommended. Observations were recorded on six qualitative traits namely plant body colour, petal colour, anther colour, leaf shape, hairiness and boll shape. Five plants per accession were randomly selected for recording the data. Scores were given for each trait as per the DUS (Distinct, Uniform, Stability) guidelines using IBPGR (1985) revised manual (**Table 1**).

Pearson correlation coefficient was computed for six qualitative traits and correlation matrix was prepared for analyzing their relationship. Principal component analysis was carried out with six qualitative traits to dissect out the relative importance of different traits in capturing the genetic variation. The factor of these traits was subsequently used to determine the contribution of each factor towards variation. The standardized values were used to perform PCA using PAST 3 (Hammer *et al.,* 2001). A scatter plot was drawn from the Eigen values associated with a components or factors in descending order versus the number of components. Dissimilarity matrix based on Euclidean distance was calculated by using the DARwin 5. The cluster analysis for grouping of desi cotton germplasm was performed using agglomerative UPGMA hierarchical method (Sneath and Sokal, 1973).

RESULTS AND DISCUSSION
The DUS testing of genotypes/cultivars using qualitative traits is very important for all the plant species to recommend plant cultivar registration and also ensure cultivar identity protection. The advantages of using morphological traits lie in the fact that they act as simple markers which are directly visible and identified directly in the field. Nevertheless, only few such qualitative traits are available for DUS characterization in many of the plant species studied (Kruskal, 1978). Furthermore, reports available on the DUS examination of desi cotton genotypes are very scanty.

In the present study, a total of 155 desi cotton genotypes were evaluated for six qualitative traits namely plant body colour, petal colour, anther colour, leaf shape, hairiness and boll shape and the frequencies for DUS descriptors

Table 1. Morphological characterization of desi cotton (*G. arboreum)* accessions for important qualitative traits

Descriptor	Category	No. of accessions	Frequency (%)
Plant body colour	Green	51	32.90
	Greenish purple	96	61.93
	Red	8	5.16
	Yellow	51	32.90
Petal colour	white	17	10.96
	Light Yellow	76	49.03
	Lavender	3	1.93
	Creamy	8	5.16
Anther colour	Yellow	67	43.22
	Creamy	88	56.77
Leaf shape	Entire	27	17.41
	Lobed	128	82.58
Hairiness	Glabrous	5	3.22
	Short hair	150	96.77
Boll Shape	Oval	33	21.29
	Round	10	6.45
	Conical	112	72.25
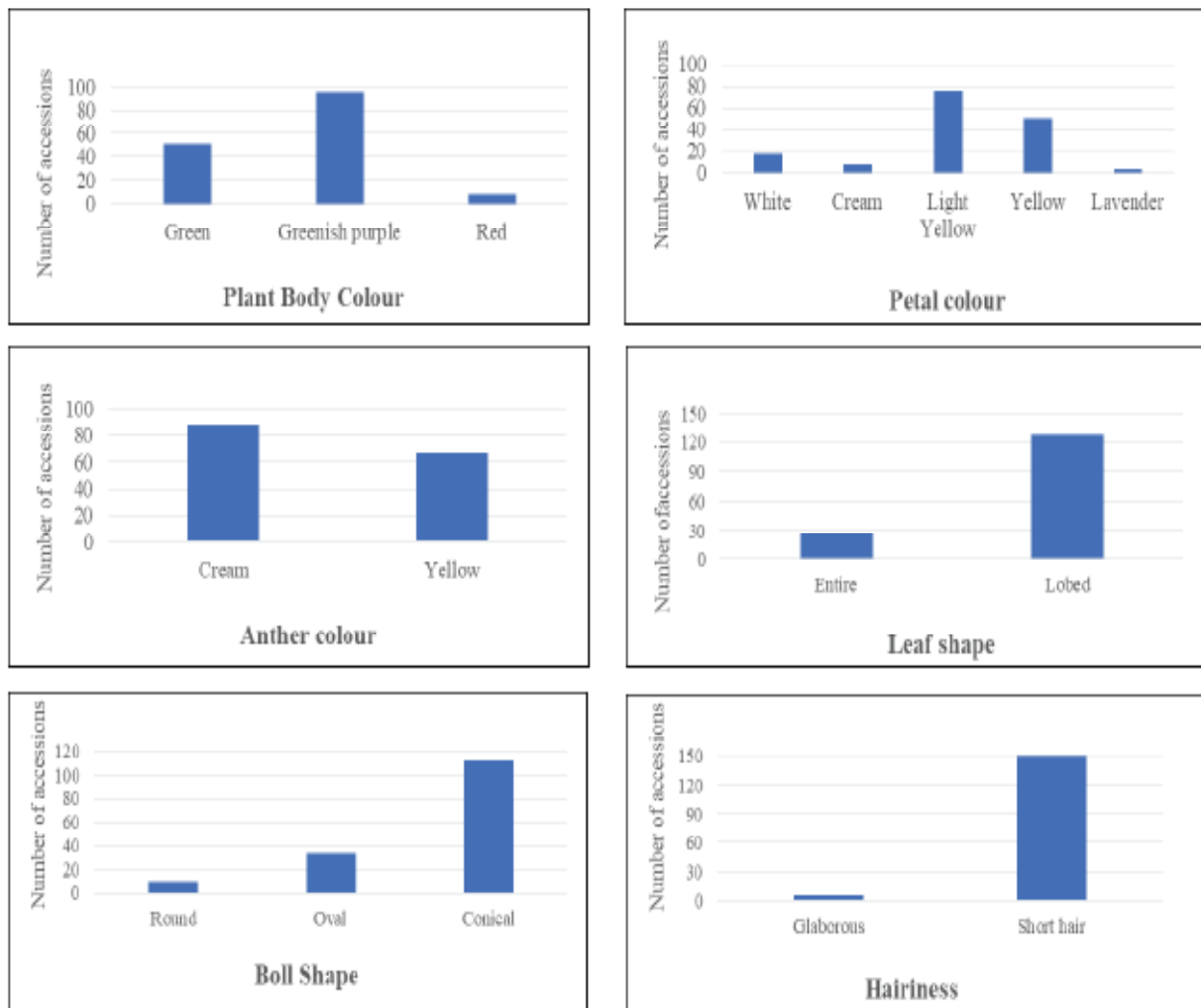
Fig. 1. Frequency distribution of 155 desi cotton germplasm accessions for various morphological traits

and states computed for each trait are presented in **Table 1 and Fig 1**. Among the genotypes studied for plant body colour, 96 genotypes represented greenish purple category (61.93%), 51 genotypes were green colour (32.9%) and eight genotypes (CISA1793, DW1601, PA839, ARBAS138, CJSA101, LD0995, RG385, CNA1039) expressed red colour with frequency of 5.16%. Five distinct groups were observed for petal colour trait with a higher frequency of 49.03% for light yellow colour in 76 genotypes followed by yellow colour 51 genotypes were observed with 32.9% frequency. On the other hand, 17 genotypes had white colour, eight genotypes (DSV1202, AKA2006, PA852, CSA1105, RG763, CNA2030, PA841, CNA2036) had creamy colour and three genotypes (SV385, POB4, TAS130) had lavender colour with frequencies of 10.96%, 5.16% and 1.93%, respectively.

In case of anther colour, 67 genotypes had shown yellow colour and 88 genotypes had shown creamy colour with the respective frequencies of 43.22% and 56.77%. Two categories of leaf shape were observed such as entire and lobed with the proportion of 17.41% and 82.58%, respectively. For the trait hairiness, majority of the genotypes belonged to short hair class (150), while only five genotypes (NDLa2005, CNA1054, PA835, JLA0610, PA255) belonged to glabrous class. Among the genotypes, three classes of boll shape namely oval (33), round (10) and conical (112) were observed. Similar work on the characterization of desi cotton genotypes based on the morphological characters such as plant body colour, colour of petals, leaf shapes and anther colours was carried out by Khangura *et al.*, (2014). Recently, Manivannan *et al.* (2018) studied 816 accessions of Asiatic cotton and reported variations in many qualitative traits namely purple green stem colour (98.7%), lobbed leaf (75%), yellow petal colour (98.9%) and conical boll shape (56%)

among the genotypes. Multivariate analytical techniques are important tools for classifying the germplasm and analysing the genetic relationships among breeding material (Mohammadi and Prasanna, 2003). Principal component analysis (PCA) is one of the multivariate statistical techniques employed to identify the minimum number of components that explain maximum variability out of the total variability present in the genotypes. In the present study, PCA analysis of qualitative data derived from 155 genotypes were used to identify determinants of qualitative trait variability. Eigen values of the correlation matrix revealed six principal components accounting for a cumulative variation of 100% (**Table 2**). As the Eigen values greater than one are considered to be significant (Hair *et al.*, 1998), out of six principal components derived in this study, first three PCs having Eigen value > 1 explained 56.91% of the total multivariate variation present in the desi cotton germplasm. It was also seen that PC1 and PC 2 contributed 20.5% and 18.62% of the variation, respectively. This indicated the contribution of many traits with high level of correlation in explaining the genetic variability present in the *G. arboreum* collections (Manivannan *et al.*, 2018).

The results of the Eigen vectors of six qualitative traits revealed that PC1 was loaded positively with the traits such as petal colour (0.64), anther colour (0.50), leaf shape (0.41) and hairiness (0.10) (**Table 3**). Second principal component (PC 2) was correlated with plant body colour (0.73) and anther colour (0.38). The traits like hairiness (0.70), boll shape (0.56), anther colour (0.34) and plant body colour (0.26) showed considerable positive loadings on PC 3. The fourth component (PC 4) was positively loaded by major traits like leaf shape (0.72), boll shape (0.40), anther colour (0.24) and plant body colour (0.15). PC 5 was positively associated with

Table 2. Percent contribution of principal components analyzed for morphological traits

Principal components	Eigen value	Percentage of variance	Cumulative
1	1.23	20.50	20.50
2	1.11	18.62	39.12
3	1.06	17.78	56.91
4	0.93	15.56	72.47
5	0.86	14.33	86.81
6	0.79	13.18	100.00

Table 3. Principal components of important morphological traits

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Plant body colour	-0.02	0.73	0.26	0.15	0.40	0.46
Petal colour	0.64	-0.12	0.06	-0.35	-0.32	0.59
Anther colour	0.51	0.38	0.34	0.24	-0.35	-0.55
Leaf shape	0.41	-0.37	-0.13	0.72	0.37	0.13
Hairiness	0.10	-0.34	0.70	-0.33	0.50	-0.17
Boll shape	-0.39	-0.22	0.56	0.40	-0.48	0.30

hairiness (0.50), plant body colour (0.40) and leaf shape (0.37). The traits like petal colour (0.59), plant body colour (0.46), boll shape (0.30) and leaf shape (0.12) showed considerable positive loadings on PC 6. On perusal of the above results, it is evident that some of the characters under study such as anther colour, plant body colour and boll shape could conspicuously be used as significant morphological markers for characterization of *G. arboreum* genotypes. In the previous studies of Manivannan *et al*., (2018), boll shape, leaf shape and bract size contributed in higher order among five PCs in differentiating 816 *G. arboreum* genotypes.

Pearson correlation coefficient revealed significant associations among traits in desi cotton accessions (**Table 4**). Boll shape was significantly and positively correlated with plant body colour, petal colour, anther colour, leaf shape and hairiness. Hairiness exhibited significant association with plant body colour, petal colour, anther colour and leaf shape. Significant positive correlation was observed between leaf shape and plant body colour, petal colour and anther colour. Plant body colour was significantly positively correlated with petal colour. There was no negative correlation observed among the morphological traits studied.

Table 4. Simple correlation matrix for eight qualitative traits of 155 desi cotton accessions

	Plant body colour	Petal colour	Anther colour	Leaf shape	Hairiness	Boll shape
Plant body colour	1.00	0.56**	0.19	0.34**	0.75**	0.86**
Petal colour		1.00	0.11	0.30*	0.48**	0.21*
Anther colour			1.00	0.54**	0.78**	0.64**
Leaf shape				1.00	0.88**	0.66**
Hairiness					1.00	0.33**
Boll Shape						1.00

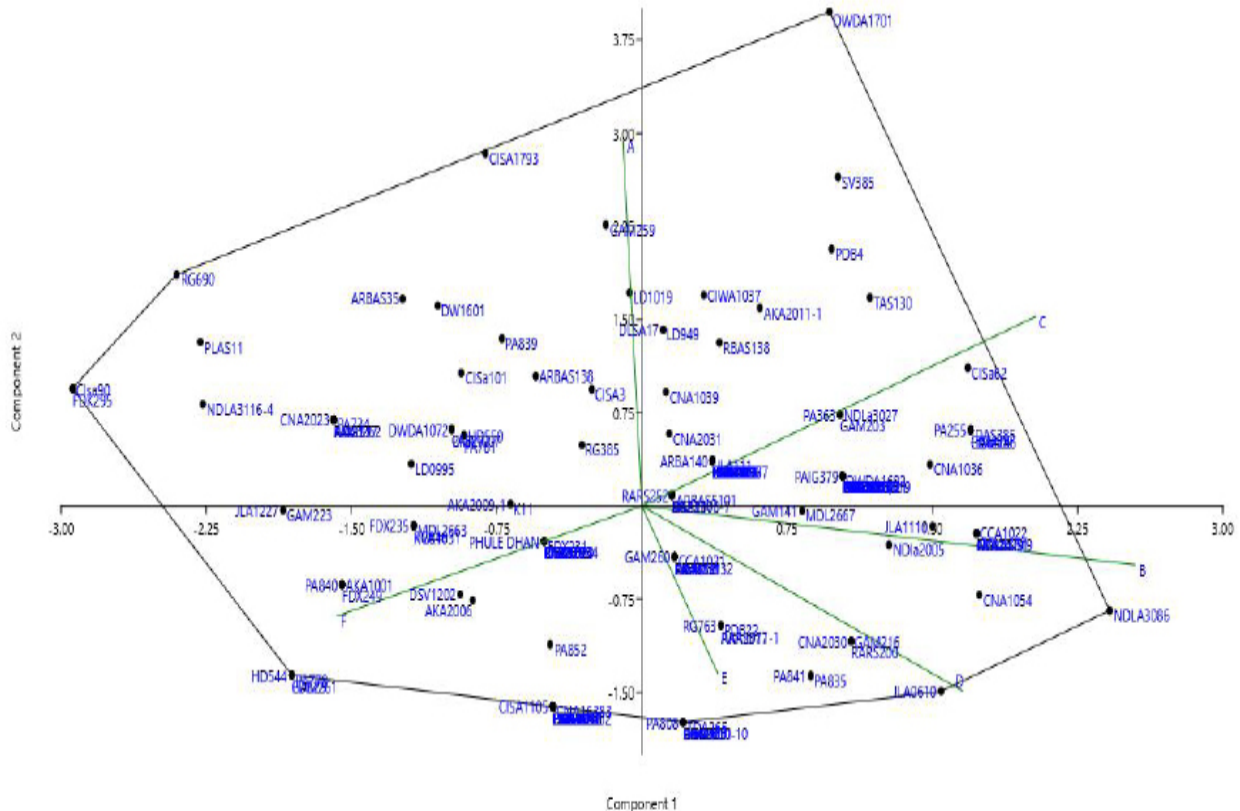** significant at 1%level ; * significant at 5% level

Fig 2. Scatter plot based on PC 1 and PC 2 of qualitative characters

Manivannan *et al*., (2018) reported both positive and negative correlation among many of the qualitative traits in asiatic cotton accessions.

A scatter plot drawn using PC1 and PC2 factor scores showed clear pattern of grouping between the genotypes in the factor plane (**Fig. 2**). Convex of the hull occupied by the genotypes namely DWDA 1701, NDLA 3086, JLA 0610, CAN 1038, RG 521, HD 544, PA 778, GAM 261, FDK 295, K 12 and CIsa 90 showed highest point of diversity among all the genotypes. With the regard to the genotypes namely GAM 260, CCA 1021, AKA 1101, NDLa 3027, RARS 200, ARBAS 5101 and CAN 2031, they were clustered in the origin as a distinct group. In this study, the clustering pattern based on multivariate analysis of qualitative traits 155 *G. arboretum* accessions were grouped into five major clusters (**Fig. 3**). Cluster I was divided into two sub-clusters and comprising of two genotypes namely DW 1601 and GAM 259. The genotypes namely TAS 130, CISA 1793 and RG 690 were placed separately in the individual clusters II, III and IV, respectively. Cluster V was found to be the largest one with all remaining *arboreum* accessions grouped together including the check varieties. Grouping of germplasm accessions based on
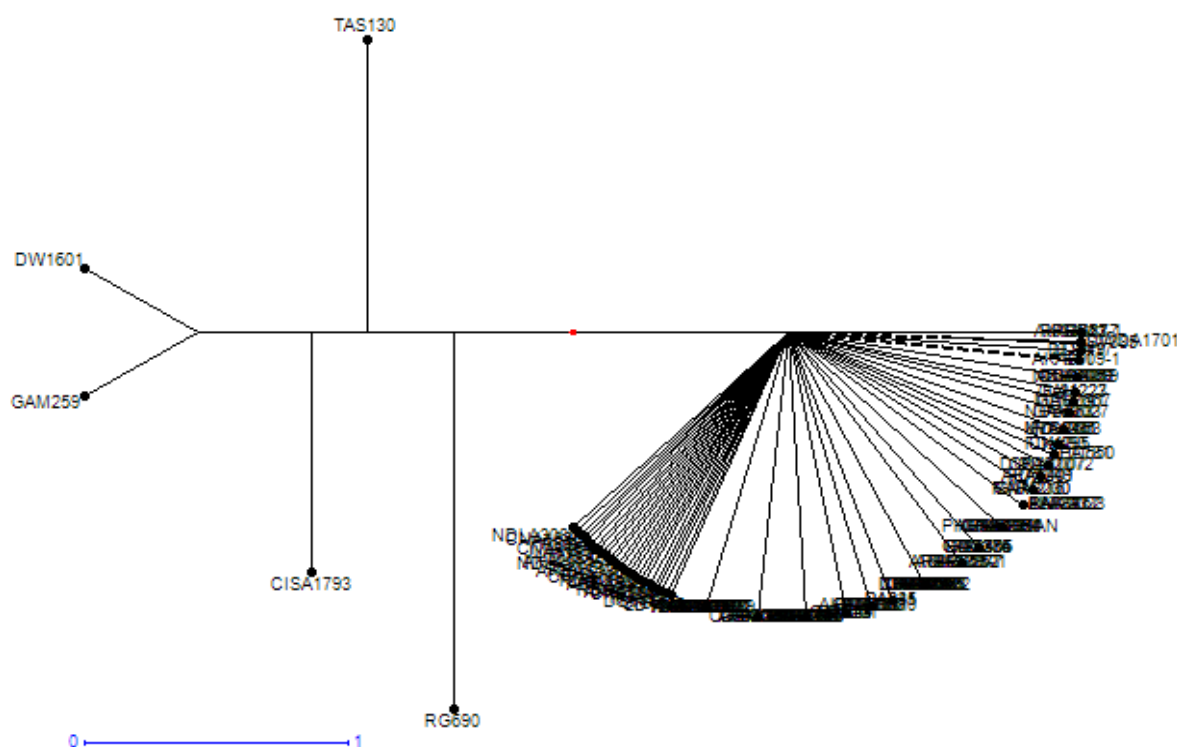

Fig 3. Clustering pattern of desi cotton germplasm accessions

qualitative descriptors using PCA and cluster analysis in cotton was also reported by Rathinavel (2017) and Manivannan *et al*., (2018)

From the data derived from both PCA and cluster analysis, no clear cut pattern of grouping could be observed in this study. However, few accessions such as DWDA 1701, GAM260, TAS 130, CISA 1793 and RG 690 were identified as genetically diverse genotypes in the present study and could be used as promising parental lines in the future desi cotton breeding programs.

The results on characterization of 155 *G. arboreum* accessions based on the qualitative traits indicate the possibility of identifying and utilizing distinguishable morphological traits for screening the genotypes/cultivars for varietal protection and genetic purity testing. This would greatly facilitate the cotton breeders to address the concerns associated with IPR issues. However, as cotton being an often-cross pollinated crop, field evaluation of genotypes should only be carried out with genetically pure homozygous genetic materials in order to ensure the precise selection of traits and genotypes to fulfill the needs of DUS characterization.

REFERENCES

Akatwijuka, R., Rubaihayo, P.R. and Odong, T.L. 2016. Genetic diversity among sorghum landraces of southwestern highlands of Uganda. *African Crop Science Journal.* **24**(2): 179-190. [Cross Ref]

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. 1998. Multivariate Data Analysis. Prentice Hall, New Jersey.

Hammer, O., David, A.T. Harper and Paul D Ryan. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia electronica* **4**(1)**:**9.

Iqbal, M., Hayat, K., Khan, R. S. A. and Sadiq, A. 2006. Correlation and path coefficient analysis for earliness and yield traits in cotton (*G. hirsutum* L.). *Asian Journal of Plant Sciences*, **5**(2): 341-344. [Cross Ref]

Johnson, R. A., and Wichern. 2002. Applied multivariate statistical analysis. Prentice-Hall, Upper Saddle River, NJ.

Khangura, R. S., Singh, L., Pathak, D., Gill, M. S. and Sohu, R. S. 2014. Assessment of genetic diversity in desi cotton (*Gossypium arboreum* L.). *Journal of Research*, **51**(2): 101-104.

Kranti, K.R. 2015. Desi Cotton - Returns. *Cotton Statistics & News .,***15:**1-4.

Kruskal, J. B. 1978. Factor analysis and principal components: Bilinear methods*: International Encyclopedia of Statistics, New York: Free Press.*

Kulkarni, V. N., Khadi, B. M., Maralappanavar, M. S., Deshapande, L. A. and Narayanan, S. 2009. The worldwide gene pools of *Gossypium arboreum* L. and *G. herbaceum* L., and their improvement. *Genetics and genomics of cotton*. pp. 69-97. [Cross Ref]

Laurentin, H. 2009. Data analysis for molecular characterization of plant genetic resources. *Genetic Resources and Crop Evolution,* **56**(2): 277-292. [Cross Ref]

Manivannan, A, Mohan, P., Saravanan, M and Waghmare, V.N. 2018. Morphological characterization of asiatic cotton (*G. arboreum*) germplasm of India. *Electronic Journal of Plant Breeding* **9**(2):673-681. [Cross Ref]

Mohammadi, S, A. and Prasanna, B.M. 2003. Analysis of genetic diversity in crop plants- Salient statistical tools and considerations. *Crop Sci*.,**43**:1234-1248. [Cross Ref]

Pearson, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* **2**(11): 559-572. [Cross Ref]

Ranjan, R., Sangwan, R.S., Siwach, S.S., Sangwan, O., Sah, M. and Singh, N. 2014. Studies on genetic divergence in *Gossypium arboreum* L. *J. Plant Breed. Genet.,* **2**(2): 57- 61.

Rathinavel, K. 2018. Principal Component Analysis with Quantitative Traits in Extant Cotton Varieties (*Gossypium hirsutum* L.) and Parental Lines for Diversity. *Current Agriculture Research Journal*, **6**(1): 54-64. [Cross Ref]

Sangwan, R., Siwach, S. and Kumar, R. 2008. Evaluation of Genetic Stock of Desi Cotton (*Gossypium arboreum*). *Indian Journal of Plant Genetic Resources,* **21(1)**: 68-71.

Sneath, P.H.A. and Sokal, R.R. 1973. Numerical taxonomy. *Freeman San Francisco*, 573 p.

Stanton, M. A., Stewart, J. M., Pervical, A. E. and Wendel, J. F. 1994. Morphological diversity and relationships in the A-genome cottons, *Gossypium arboreum* and *G. herbaceum*. *Crop science,* **34**(2): 519-527. [Cross Ref]

Stoskopf, N. C., Tomes, D. T., Christie, B. R. and Christie, B. R. 2019. *Plant breeding: theory and practice*: CRC Press. [Cross Ref]

Vafaie-Tabar, M., Chandrashekaran, S., Singh, R., and Rana, M. 2003. Evaluation of genetic diversity in Indian tetraploid and diploid cotton (*Gossypium spp.)* by morphological characteristics and RAPDs. *Ind. J. Genet*., **63**(3):230-234.

Zada, M., Zakir, N., Rabbani, M. A. and Shinwari, Z. K. 2013. Assessment of genetic variation in Ethiopian mustard (*Brassica carinata* A. Braun) germplasm using multivariate techniques. *Pak. J. Bot,* **45**(S1): 583-593.