Eco. Env. & Cons. 29 (October Suppl. Issue) : 2023; pp. (S224-S230)
Copyright@ EM International
ISSN 0971–765X

DOI No.: http://doi.org/10.53550/EEC.2023.v29i05s.039

Multivariate Analysis in Upland Cotton (*Gossypium hirsutum* L.) using Principal Component and Cluster Analysis

K. Mohan Vishnuvardhan*, B. Venkata Ravi Prakash Reddy, D. Lakshmikalyani, M. Sivaramakrishna, K. Sudheepthi, K. Amarnath[1] and N.C. Venkateswarlu

[1]Regional Agricultural Research Station, Nandyal, Andhra Pradesh, India

(Received 9 April, 2023; Accepted 30 May, 2023)

ABSTRACT

The experiment material comprising 17 cotton genotypes with a view to study of genetic parameters for different yield and yield parameters. The analysis of variance revealed the existence of significant differences among the traits studied. The findings of correlation coefficient studies revealed that seed cotton yield established strong positive correlation with lint yield (0.9827) followed by plant height (0.6405) and number of bolls per plant (0.4717). The results of principal component analysis revealed that, 4 Principal Components (PCs) were established with Eigen value greater than 1.00 which accounted for 83.9 % of the total variation for discriminating the lines. From principal component analysis, PC1 showed highest amount of variance (33.1%) with mostly related to traits like boll weight, seed index, lint index and halo length indicated the importance of these traits in relation to yield enhancement. Cluster analysis classified the genotypes into five clusters among which cluster I was largest with eight genotypes followed by cluster III and cluster IV with four and three genotypes respectively indicating the versatility of the genotypes of these clusters in the exploitation of heterosis.

Key words: *Correlation, Principal component analysis, Cluster analysis, Cotton.*

Introduction

Cotton is the most important renewable natural fibre used in textile industry. Cotton is primarily grown for fibre, oil and feed (livestock) and plays crucial role in boosting a nation's economy, hence commonly known as "White gold"(Adeela *et al.,* 2021). It belongs to the genus *Gossypium*, which consists of five allo-tetraploid and 45 diploid species. Among them only four species are cultivated worldwide comprising of two diploids and two tetraploids also called old world and new world species respectively (Ulloa *et al.,* 2006). India is the world's leading producer of cotton, surpassing China with a production of 362 lakh bales and a productivity of 510 Kg ha^{-1} cultivated over an area of 120.69 lakh hectares (AICRP on Cotton Project Coordination report, 2022-23). However, due to erratic climatic conditions combined with biotic and abiotic stress, cotton yields were declined from the past few years. Therefore, development of a variety with increased seed cotton yieldwith superior fiber quality is the need of the hour. Broadening genetic base and exploitation of genetic diversity of cultivated species aids in crop improvement.

Being a quantitative trait, seed cotton yield mainly depends on different contributed traits hence, could be increased by considering positive contribution of these yield traits. A through picture about nature and magnitude of crop performance

and its associated traits with yield is fundamental for a breeder to combine those favorable traits at the same time eliminating the limiting factors to the yield. Nature and magnitude of genetic variance relies on different statistical methods used for assessment. Biometrical techniques like principle component analysis (PCA), correlation analysis and cluster analyses have been repeatedly used identify the genetic diversity in different genotypes (Brown-Guedira *et al.,* 2000). Principal component analysis (PCA) was used to identify redundancy of the genotypes with similar characters and their elimination (Adams, 1995). PCA also illustrates the significance of major contributors towards total diversity at each axis of differentiation (Jarwar *et al.,* 2019). Principal component analysis in association with cluster analysis was accomplished to find the similarity among the genotypes for the traits and their placement into different clusters (Brown, 1991; Jian *et al.,* 2006; Qiaoling and Zhe, 2011). Principal component analysis (PCA) and cluster analysis are therefore two important statistical programs that aid in selecting elite genotypes. Therefore, information about the correlations of traits is of immense importance to the plant breeders for the development of improved lines. In this context, the intension of the present investigation isto appraise the genetic diversity in yield and its component traits in cotton genotypes and to analyze the associations among them.

Materials and Methods

Experimental material for the present investigation comprised of 17upland cotton genotypes including 15 advanced lines namely, NDLH 2091-2, NDLH 2092-3, NDLH 2094-3, NDLH 2094-4, NDLH 2095-1, NDLH 2095-2, NDLH 2097, NDLH 2099-1, NDLH 2099-3, NDLH 2102, NDLH 2104, NDLH 2106-1, NDLH 2106-5, NDLH 2107-1 and NDLH 2107-3developed at Regional Agricultural Research Station (RARS), Nandyal, Andhra Pradesh India along with two check entries *i.e.*, Srirama and Jaadoo. The experiment was carried out adopting Randomized Block Design with three replications following 60 × 30 cm spacing at Research Farm, Regional Agricultural Research Station, Nandyal, Andhra Pradesh during *kharif*, 2022. All need based plant protection measures were taken up during the experimental period. Observations were recorded on five randomly selected plants for 10 characters namelydays to 50 % flowering, plant height (cm), number of bolls per plant , boll weight (g), seed index (g), lint index (g), ginning percentage, halo length (mm), lint yield (kg/ha) and seed cotton yield (Kg/ha). Seed index was calculated by weighing 100 healthy seeds. Ginning percentage and Lint index was calculated by using formula suggested by Ghule *et al*. (2013).

$$\text{Ginning percentage (\%)} = \frac{\text{Weight of lint}}{\text{Weight of seed cotton}} \times 100$$

$$\text{Lint Index} = \frac{\text{Weight of 100 seeds X ginning \%}}{100 - \text{ginning \%}}$$

Pearson correlation coefficient and multivariate analysis techniques (PCA and Wards cluster analysis) was calculated for the 17 cotton genotypes using STAR (Statistical tool for Agricultural Research) 2.0.1 software (Gulles *et al.,* 2014). The freely available 64-bit version of the R studio statistical software R version was used to obtain correlation figure.

Results and Discussion

The results on analysis of variance (ANOVA) for yield and yield component traits revealed highly significant differences among the genotypes for all the characters studied (Table 1), indicating the existence of sufficient variation among the genotypes and therefore opportunity for plant breeder to undertake further breeding activities like hybridization program. Pearson correlation coefficient analysis

Table 1. Analysis of variance for yield and yield component traits in cotton (*Gossypium hirsutum* L.)

Source of variation	df	Days to 50% Flowering	Plant height (cm)	Number of bolls per plant	Boll weight (g)	Seed index	Lint index	Ginning percentage	Halo length (mm)	Lint yield (kg/ha)	Seed cotton yield (kg/ha)
Replication	2	2.90	258.84	3.35	0.13	0.13	0.03	11.614	0.29	475.30	16740
Entries	16	4.75**	470.62**	51.12**	0.20*	10.21**	2.63**	9.66*	42.14**	42766**	383700**
Error	32	1.59	109.21	3.08	0.10	0.05	0.014	4.70	0.80	858.0	17702

was performed to study the association among yield and yield related traits in 17 cotton genotypes. The results were presented in Table 2 and Figure 1. Seed cotton yield was found to be positively and significantly correlated with the traits, number of bolls per plant (r = 0.4717*), lint yield (0.9827*) and plant height (0.6405*). These results were in correspondence with Gowda *et al.* (2022) for number of bolls per plant, Mudhalvan *et al.* (2021) for lint yield and Manan *et al.* (2022) for plant height. Similarly, with regard to inter character associations, plant height exhibited positive significant association with lint yield (0.649**). The trait seed index reported positive significant association with lint index. Seed cotton yield exhibited non significant positive association with Days to 50 % Flowering, boll weight (g), ginning percentage (%) and halo length (mm). Similar results were reported by Reddy *et al.* (2015) for days to 50 % Flowering and lint index, Jyoti *et al.* (2021) for boll weight and halo length Rajamani (2016) for seed index and Mudhalvan *et al.* (2021) for ginning percentage.

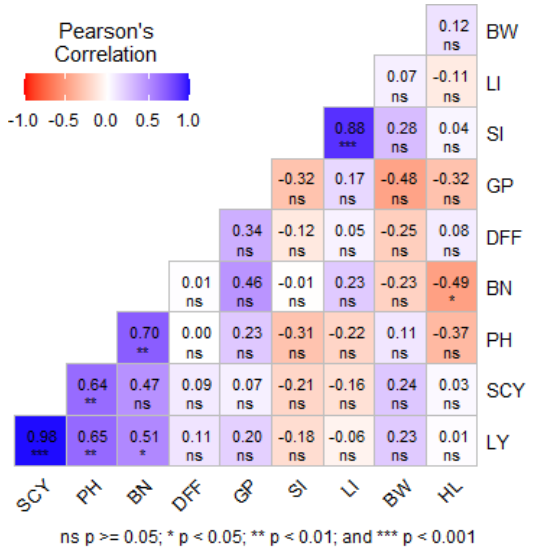

Fig. 1. Correlation matrix plot for yield and yield component traits of 17 genotypes of cotton.
BW: boll weight, LI: Lint index, SI: Seed index, GP: Ginning percentage, DFF: Days to 50% flowering, BN: Boll number, PH: Plant height, SCY Seed cotton yield, LY: lint index.

Principal component analysis

Principal component analysis revealed that, out of 10 principal components, four components had extracted Eigen value of more than onewhich accounted for 83.9 % of the total cumulative variation for discriminating the lines (Table 3). Scree plot exhibited the variance percentage associated with all principle components, as represented by a graph between the eigen values and principal components. PC 1 contributed highest variability of 33.17 per cent with eigen value of 3.31% while minimum variability was noticed in PC 9 and PC 10 with declining eigen values (Figure 2). Characteristics of each principal component were determined on the basis of estimated factor loadings. The results on Eigen vector loading values pertaining to 10 morphological traits of 17 cotton genotypes is presented in Figure 4. The characters, namely seed index (0.231), halo length (0.198), lint index (0.108) and boll weight (0.051) explained maximum variance in PC 1 component. The second principal component (PC 2) contributed to19.8 per cent of total variance. The characters namely halo length (0.3383), seed cotton yield

Table 2. Correlation matrix for yield and yield component traits in Cotton(*Gossypium hirsutum* L.)

	DF	PH	NBPP	BW	SI	LI	GP	HL	LY	SCY
DF	1.0000	0.0015	-0.016	0.2774	-0.1522	-0.0308	0.2512	0.026	0.0794	0.0796
PH		1.0000	0.6988**	0.0787	-0.3108	-0.2246	0.2303	-0.3716	0.649**	0.6405*
NBPP			1.0000	-0.2754	-0.010	0.2393	0.4794	-0.5068	0.5158*	0.4717*
BW				1.0000	0.2644	0.0557	-0.4591	0.1117	0.207	0.2157
SI					1.0000	0.8849*	-0.3151	0.0421	-0.1797	-0.2134
LI						1.0000	0.158	-0.1165	-0.064	-0.1603
GP							1.0000	-0.3175	0.1993	0.0655
HL								1.0000	0.0100	0.0300
LY									1.0000	0.9827*
SCY										1.0000

Note: DF = Days to 50% flowering; PH = Plant height (cm); NBPP = Number of bolls per plant; BW = Boll weight (g); SI = Seed index; LI = Lint index; GP = Ginning percent (%); HL = Halo length (mm), LY = Lint yield (kg/ha); SCY = Seed cotton yield (kg/ha)

(0.150), boll weight (0.136), lint yield (0.081), days to 50 % flowering (0.025) and plant height (0.0153) explained maximum loadings in this second component (PC2). It was observed that the traits namely boll weight (0.572), seed index (0.386), seed cotton yield (0.312), lint yield (0.296), lint yield (0.209), halo length (0.180) and plant height (0.091) contributed maximum variance in third principal component (PC 3). Further PC 3 contributed to 19.5 per cent of total variance. The fourth principal component was characterized by 12.0 per cent contribution towards the total variability. Characters namely plant height (0.218), boll weight (0.147) and number of bolls per plant (0.110) reported maximum variance in this component. The fifth principal component (PC 5) with eigen value nearly one contributed to a total variability of 7.0 per cent. The characters namely days to 50 % flowering (0.662), boll weight (0.394), plant height (0.151) and seed index (0.040). The biplot picture obtained from the first and second principal components represents that variables are super imposed as vector (Figure 3). The biplot exhibited that as a whole halo length, seed index, boll weight, seed cotton yield and lint index contributed maximum towards variation in the genotypes studied. Biplot also revealed strength of correlation

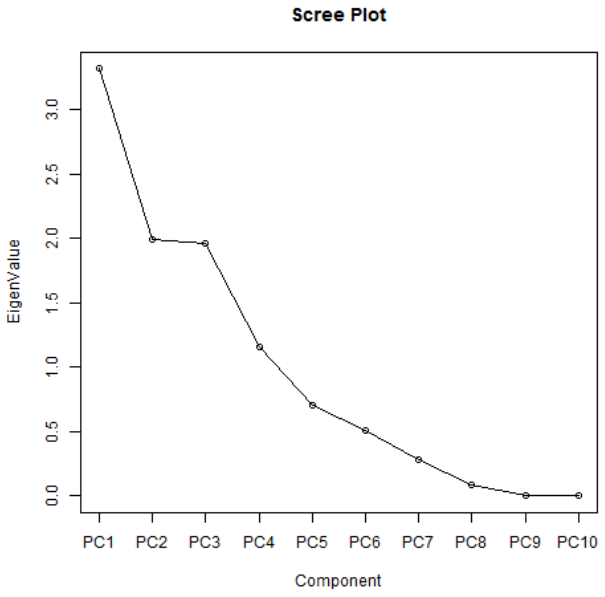

Fig. 2. Scree plot of PCA depicting eigen values and component numbers

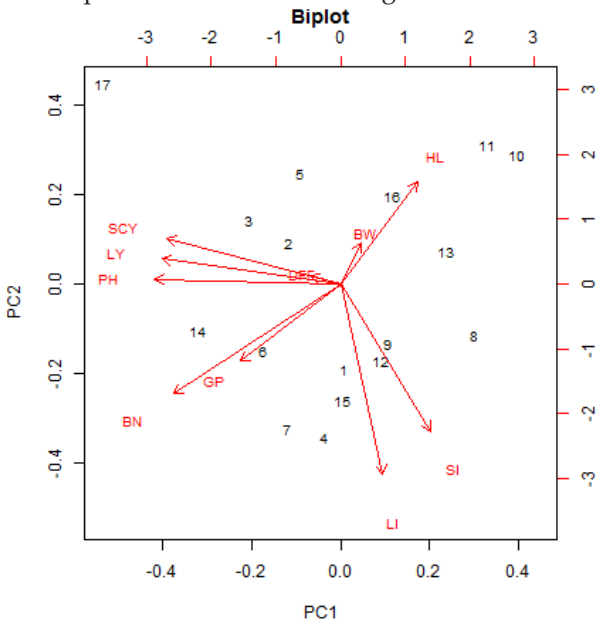

Fig. 3. Biplot of the first two PCAs showing relation among different traits in cotton.

Table 3. Eigen values, proportion of the total variance represented by first five principal components, cumulative percent variance and component loading of different traits in cotton (*Gossypium hirsutum* L.)

PC1	PC2	PC3	PC4	PC5	
Eigen value (Root)	3.316	1.989	1.955	1.154	0.703
%Var.Exp.	33.1	19.8	19.5	11.5	7.0
Cum.Var.Exp.	33.17	53.06	72.62	84.16	91.19
Days to 50 % Flowering	-0.076	0.025	-0.286	-0.657	0.662
Plant height (cm)	-0.478	0.0153	0.091	0.218	0.151
Number of bolls per plant	-0.427	-0.363	-0.029	0.11	-0.151
Boll weight (g)	0.051	0.136	0.572	0.147	0.394
Seed index	0.231	-0.488	0.386	-0.15	0.04
Lint index	0.108	-0.632	0.209	-0.241	-0.059
Ginning percentage (%)	-0.257	-0.2571	-0.4125	-0.1459	-0.2227
Halo length (mm)	0.198	0.3383	0.18	-0.516	-0.534
Lint yield (kg/ha)	-0.458	0.081	0.296	-0.256	-0.114
Seed Cotton yield (kg/ha)	-0.445	0.150	0.3122	-0.228	-0.060

among characters studied. These results are in accordance with (Iqbal and Rahman, 2017) for boll weight, Saeed *et al.* (2014) for plant height Rathinavel (2018) for seed index, Jarwar *et al.* (2019) and Sarwar *et al.* (2021), Zafar *et al.* (2021) for seed cotton yield.

Cluster analysis

Cluster Analysis is the assignment of a set of observations into subsets called clusters based on the similarity of the observations in the same cluster. The results of Ward's cluster analysis revealed that a total of 17 genotypes were grouped in five clusters. The cluster I, being the largest comprised of 8 genotypes followed by cluster III and cluster IV comprising 4 and 3 genotypes respectively (Table 5 & Fig 5). The cluster group II and V have single genotype each. When compare the mean values of clusters for

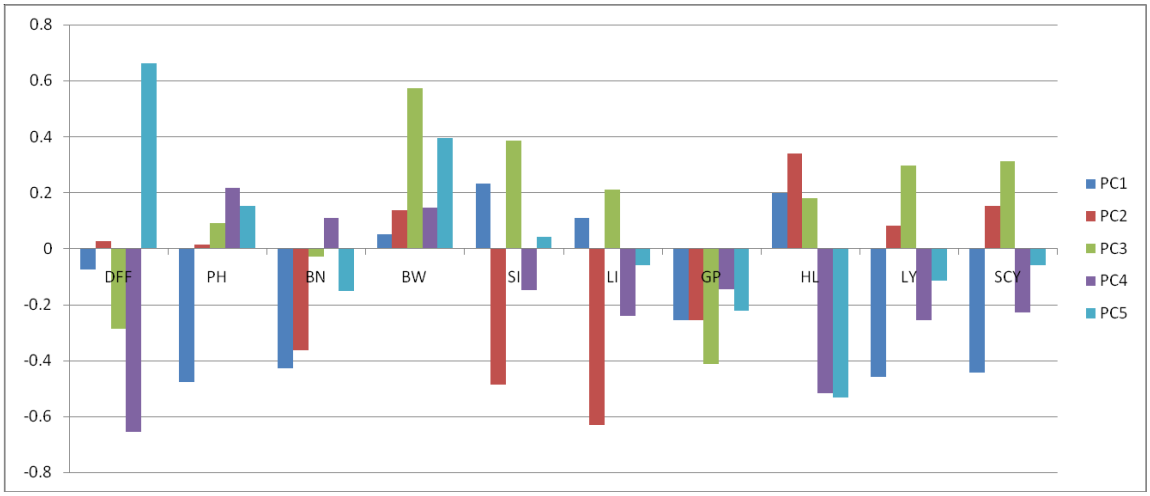

Fig. 4. Eigen vector loading values of 10 morphological traits for 17 upland cotton genotypes

Table 4. Mean values of clusters for 10 morphological traits in 17 genotypes of upland cotton

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
DF	60.12	61	58.75	61.67	61
PH	107.5	115	99.25	86	130
NBPP	22	17	17	14	21
BW	3.0	2.9	3.3	2.9	3.5
SI	11.5	7.8	13.1	11.53	9.8
LI	6.3	4.4	6.3	5.8	4.7
GP	35.46	36.2	32.38	33.4	32.5
HL	22.65	22.2	26.85	28.7	27.2
LY	389	270	321	270.67	719
SCY	1093.8	745	921.25	808.67	2212

Note: DF = Days to 50% flowering; PH = Plant height (cm); NBPP = Number of bolls per plant; BW = Boll weight (g); SI = Seed index; LI = Lint index; GP = Ginning percent (%); HL = Halo length (mm), LY = Lint yield (kg/ha); SCY = Seed cotton yield (kg/ha)

Table 5. Clustering pattern of 17 cotton genotypes for yield and yield component traits using Wards Method

S. No.	Cluster Number	Number of genotypes	Name of Genotype(s)
1	Cluster I	8	NDLH 2091-2, NDLH 2092-3, NDLH 2094-3, NDLH 2094-4, NDLH 2095-2, NDLH 2097, NDLH 2107-1, NDLH 2107-3
2	Cluster II	1	NDLH 2095-1
3	Cluster III	4	NDLH 2099-1, NDLH 2099-3, NDLH 2106-5, SRIRAMA
4	Cluster IV	3	NDLH 2102, NDLH 2104, NDLH 2106-1
5	Cluster V	1	JAADOO

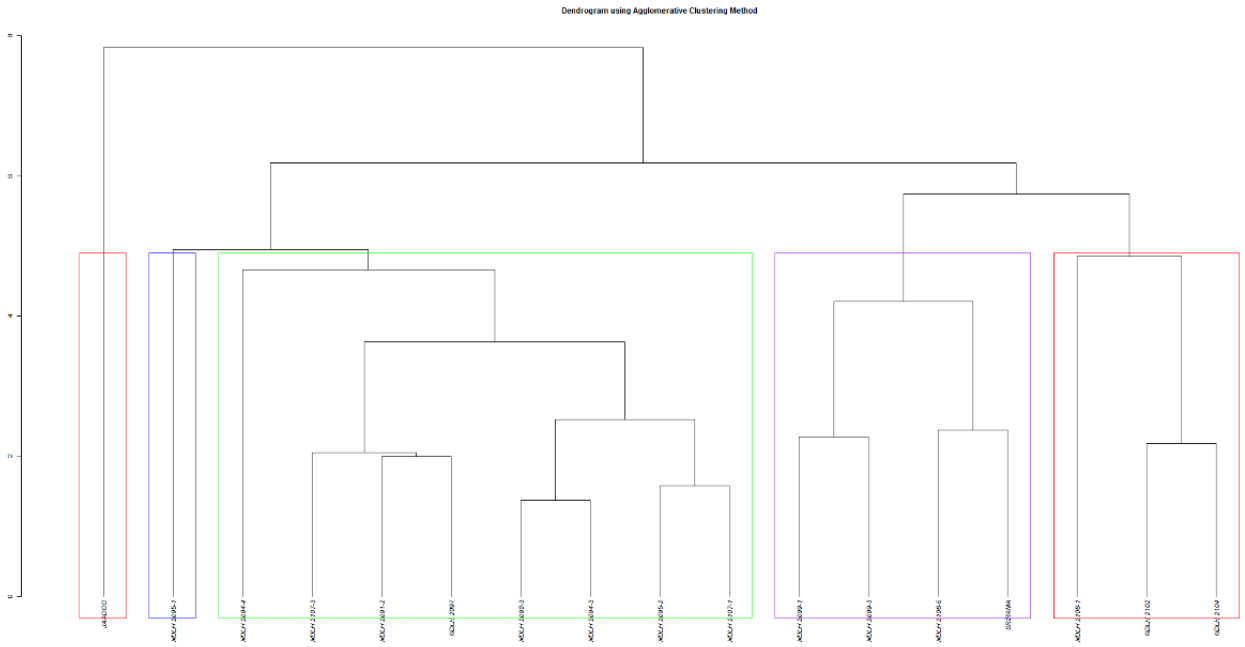

Fig. 5. Dendrogram of cotton genotypes resulting from cluster analysis using Ward's method based on standardized data of all the traits.

the studied traits, cluster V showed maximum values for seed cotton yield, lint yield, boll weight while cluster I showed higher values for number of bolls per plant and lint index. Similarly cluster II showed higher for ginning percentage, cluster III for seed index, lint index and for earliness while cluster IV for halo length. Similar results were observed by Shabbir *et al.* (2016); Farooq *et al.* (2017); Shakeel *et al.* (2018) and Jarwar *et al.* (2019). Based on cluster analysis the genotypes in cluster-IV may be utilized for incorporation of earliness traits. The clusters-V and I may be further exploited in breeding programs for the development of high yielding cotton genotypes with desirable fiber quality traits.

Conclusion

Correlation analysis revealed seed cotton yield positive significance with traits number of bolls per plant, lint yield and plant height. Both the principal component analysis and the hierarchical cluster analysis confirmed the findings of each other. PCA is helpful in identifying those factors that influence the genetic variation in population. However, cluster analysis could effectively explain the characteristics of genotypes in various clusters. Selection of genotypes from clusters with high mean for the respective traits is suggested for utilization in hybridization programmes aimed at improvement of the respective traits.

Acknowledgement

The authors would like to express their sincere regards to Acharya N. G. Ranga Agricultural University, Lam, Guntur (AP), Indiafor providing necessary facilities to the research experiment conducted at Regional Agricultural Research Station, Nandyal

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.

Adeela, S., Zafar, M.M., Razzaq, A., Manan, A., Muhammad, H., Sunaina, S., Abdul, R., Huijuan, Mo., Muhammad, A., Maozhi, R., Amir, S. and Youlu, Y. 2021. Genetic variability for yield and fiber related traits in genetically modified cotton. *Journal of Cotton Research.* 4 (19): 1-10.

AICRP on Cotton Project Coordination report, 2022-23 presented at Cotton Annual group meeting at Punjab Agricultural University, Ludhiana, Punjab during 6th to 7th April, 2023. ICAR–Central Institute of Cotton Research, Nagpur, Maharashtra, India.

Brown, J.S. 1991. Principal component and cluster analyses of cotton cultivar variability across the US cotton belt.*Crop Science Society of America*, pp. 915-922.

Brown-Guedira, G.L., Thompson, J.A., Nelson, R.L. and Warburton, M.L. 2000. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. *Crop Science.* 40: 815-823. https://doi.org/10.2135/cropsci2000.403815x

Farooq, J., Rizwan, M., Sharif, I., Saleem, S., Chohan, S.M. and Kainth, R.A. 2017. Genetic diversity studies in some advanced lines of *Gossypium hirsutum* L. for yield and quality related attributes using cluster and principle component analysis. *AAB Bioflux.* 9(3): 111-118.

Ghule, P.L., Jadhav, J.D., Palve, D.K. and Dahiphale, V.V. 2013. Bt cotton and its leaf area index (LAI), ginning (%), lint index (g), earliness index and yield contributing characters. *International Research Journal of Agricultural, Economics & Statistics.* 4(1): 85-90.

Gowda, S.A., Katageri, I.S., Kumar, N.V.M. and Patil, R.S. 2022. Development and evaluation of India's first intraspecific *Gossypium barbadense* cotton recombinant inbred mapping population for extra-long staple fibre traits. *Journal of Genetics*. 101: 4(1-14).Https://doi.org/10.1007/s12041-021-01338-7.

Gulles, A.A., Bartolome, V.I., Morantte, R.I.Z.A., Nora, L.A., Relente, C.E.N., Talay, D. T., Caneda, A.A and Ye, G. 2014. Randomization and analysis of data using STAR (Statistical Tool for Agricultural Research). *Philippine Journal of Crop Science.* 39(1): 137.

Iqbal, M.A. and Rahman, M.U. 2017. Identification of marker trait associations for lint traits in cotton. *Frontiers in Plant Science.* 8(86): 1-20.

Jarwar, A.H., Wang, X., Iqbal, M.S., Sarfraz, Z., Wang, L., Qifeng, M.A. and Shuli, F. 2019. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality traits in cotton cultivars. *Pakistan Journal of Botany*. 51(3) : 1143-1148. https://doi.org/10.30848/ PJB2019-3(38).

Jarwar, H.A., Wang, X., Iqbal, S.M., Sarfraz, Z., Wang, L., Ma, Q. and Shuli, F.A. 2019. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality traits in cotton cultivars. *Pakistan Journal of Botany.* 51(3): 1143-1148.

Jian, C., Wei, L., Ruili, L. and Feng, L.F.W. 2006. Genetic diversity detected by cluster analysis North of Anhui major wheat cultivars. *Chinese Agricultural Science Bulletin.* 11: 031.

Jyoti, J.G., Valu, M.G. and Geeta, O.N. 2021. Correlation, path coefficient and D2 analysis study of seed cotton yield and fibre quality traits in American cotton (*Gossypium hirsutum* L.). *Journal of Pharmacognosy and Phytochemistry*. 10 (3): 222-230.

Manan, A., Zafar, M.M., Ren, M., Khurshid, M., Sahar, A., Rehman, A., Firdous, H., Youlu, Y., Razzaq, A. and Shakeel, A. 2022. Genetic analysis of biochemical, fiber yield and quality traits of upland cotton under high temperature. *Plant Production Science.* 25 (1): 105-119.

Mudhalvan, S., Rajeswari, S., Mahalingam, S., Jeyakumar, P., Muthuswami, M. and Premalatha, N. 2021. Causation studies for Kapas yield, yield components and lint quality traits in Mexican cotton (*Gossypium hirsutum* L.). *Environment Conservation Journal*. 22 (3): 357-363.

Qiaoling, W. and Zhe, L. 2011. Principal component analysis of F2 individual selection in uplandcotton (*Gossypium hirsutum* L.). *Journal of Henan Institute of Science and Technology (Nature Sciences Edition).* 5: 004.

Rajamani, S. 2016. Character association and path analysis on seed cotton yield and its component characters in cotton (*Gossypium* Spp.). *Journal of Cotton Research and Development.* 30(1): 16-21.

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.

Reddy, K.B., Reddy, V.C., Ahamed M.L., Naidu, T.C.M. and Srinivasarao, V. 2015. Multivariate Analysis in Upland Cotton (*Gossypium hirsutum* L.). *Electronic Journal of Plant Breeding.* 6(4) : 1019-1026.

Saeed, F., Farooq, J., Mahmood, A., Riaz, M., Hussain, T. and Majeed, A. 2014. Assessment of genetic diversity for Cotton leaf curl virus (CLCuD), fiber quality and some morphological traits using different statistical procedures in (*Gossypium hirsutum* L.). *Australian Journal of Crop Science.* 8(3): 442-447.

Sarwar, G., Nazir, A., Rizwan, M., Shahzadi, E. and Mahmood, A. 2021. Genetic diversity among cotton genotypes for earliness, yield and fiber quality traits using correlation, principal component and cluster analyses. *Sarhad Journal of Agriculture*. 37(1) : 307-314.

Shabbir, R.H., Bashir, Q.A., Shakeel, A., Khan, M.M., Farooq, J., Fiaz, S., Ijaz, B. and Noor, M.A. 2016. Genetic divergence assessment in upland cotton (*Gossypium hirsutum* L.) using various statistical tools. *Journal of Global Innovations in Agricultural and Social Sciences*. 4(2): 62-69. https://doi.org/10.22194/JGIASS/4.2.744.

Shakeel, A., Azhar, M.T., Ali, I., Ain, Q.U., Zia, Z.U., Anum, W. and Zafar, A. 2018. Genetic diversity for seed cotton yield parameters, protein and oil contents among various bt. cotton cultivars. *Int. J. Biosci.* 12(1): 242-251. https://doi.org/10.12692/ijb/12.1.242-251.

Ulloa, M., Steward, J.M.D., Garcia, E.A.C., Godoy, S.A., Gaytan, A.M. and Acosta, S.N. 2006. Cotton genetic resources in the Western States of Mexico: in situ conservation status and germplasm collection for ex situ preservation. *Genetic Resources and Crop Evolution.* 53(4): 653–668.

Zafar, M.M., Manan, A., Razzaq, A., Zulfqar, M., Saeed, A., Kashif, M., Khan, I.A., Sarfraz, Z., Mo, H., Iqbal, S.M., Shakeel, A. and Ren, M. 2021. Exploiting agronomic and biochemical traits to develop heat resilient cotton cultivars under climate change scenarios. *Agronomy.* 11 (1885): 1-14.