J. Farm Sci., 39(1): (1-5) 2026

RESEARCH PAPER

Assessment of correlation and principal component analysis for yield and fibre quality traits in cotton (*Gossypium hirsutum* L.)

K. SAINATH[1*], J. M. NIDAGUNDI[1], P. H. KUCHANUR[1], S.MUNISWAMY[1], M. K. MEENA[2] AND M. LAKSHMIKANTH[3]

[1]Department of Genetics & Plant Breeding, [2]Department of Crop Physiology,
[3]Department of Soil Science & Agril. Chemistry, College of Agriculture, Raichur
University of Agricultural Sciences, Raichur - 584 104, India
*E-mail: sainathkarbhari98@gmail.com

(Received: January, 2026 ; Accepted:March,2026)

DOI: 10.61475/JFS.2026.v39i1.01

Abstract: Cotton (*Gossypium* spp.) is one of the world's most important fibre crops, contributing significantly to the global textile industry. The present study was conducted to assess correlation and principal component analysis (PCA) among 100 cotton germplasm lines, which are maintained under the All India Coordinated Research Project (AICRP) on Cotton, MARS, Raichur, during *kharif* 2023. The material was evaluated in an augmented block design with five standard checks at College of Agriculture, Bheemarayanagudi. Correlation analysis showed that seed cotton yield had a highly significant and positive association with lint yield, number of bolls and boll weight, suggesting that these traits can serve as reliable selection criteria in yield improvement programmes. Principal component analysi revealed that the first four principal components accounted for 73.16% of the total variability, with the first component contributing 30.08%. Lint yield, seed cotton yield, seed index, number of bolls and boll weight were identified as the major contributors to total genetic divergence. The study highlights substantial genetic diversity between germplasm lines, which can be effectively utilized in cotton breeding programmes aimed at enhancing yield and fibre quality.

Key words: Correlation, Germplasm line, Genetic improvement, Principal component analysis, Phenotypic variability

Introduction

Cotton is mainly grown in the Asia (~70%) followed by the Americas, which has nearly 20% of the global cotton cultivation. Africa grows nearly 6% of the total world cotton cultivation while Europe grows less than 2% of it. The fiber obtained from cotton is a raw material for the textile industry. Textile is one of the biggest industries of the world that turn cotton fiber into a variety of products. Out of 54 *Gossypium* species, only four are cultivated, *i.e. G. hirsutum, G. barbadense, G. herbaceum and G. arboreum.* The former two are tetraploid whereas the latter two are diploid. The evolutionary history of cotton is very interesting and includes trans oceanic mating of parental species as diploid progenitors of tetraploid cotton belongs to different continents. Tetraploid species contain AADD genome and were discovered from the new world (America) in which the A genome originated from Africa and Asia and the D genome has new world origin (Jabran *et al.,* 2019).

The world population is increasing day by day; therefore, it is necessary to increase the productivity of crops to meet the requirement of textile industry. The utilization of various breeding tools is one method to meet the demand of textile industry (Farooq *et al.,* 2014). Genetic improvement in seed cotton yield has been mainly brought through selection and hybridization of the available germplasm. Seed cotton yield is a complex trait and depends on many morphological and physiological traits which contribute to the yield.Understanding the genetic basis of important yield contributing traits is the pre-requisite and information about their relationship must be available to cotton breeders. All of the yield related traits are correlated with each other in a way that increases or decreases in one trait directly affects others. So, estimation of phenotypic correlations among these traits are helpful to initiate the breeding programs.

Correlation provides the information of association among different traits being helpful in selection of plants with a good combination of yield and quality related traits (Salahuddin *et al.,* 2010). The association among yield related traits play a significant role toward improvement and to produce promising genotypes with high yield and quality. Correlation analysis also predicts direction of relationship among desirable plant traits. (Nawaz *et al.,* 2019).

The amount and nature of attainable genetic variability between the genotypes has adequate scope to develop in successful breeding program for upgrading of various attributes (Ahsan *et al.,* 2015). Principal component analysis (PCA) was used to analyze the variability in the germplasm, identify the plant characteristics that contribute to increased diversity and ascertain the relative contributions of the different characters to the total variability in the germplasm. In order to assess the relationships and variability among various germplasms for their usage in upcoming cotton development programs, principal component analysis is also utilized. (Ullah *et al.,* 2022). Thus, focusing this present study is being designedto check the genetic diversity amongst different accessionto find the magnitude of correlation between different traits.

Material and methods

A core set of 100 cotton germplasm lines contrasting for yield and yield attributing traits were utilized in study which were preserved at All India Co-ordinated Research Project (AICRP) on Cotton, MARS, UAS, Raichur. The experimental material was planted during *kharif* 2023 on medium black soil at Experimental block, College of Agriculture, Bheemarayanagudi which is geographically located at 16.43° N, 76. 47° E with an altitude of 412 m above mean sea level (MSL) in the North eastern dry zone (Zone No. II) of Karnataka. The germplasm line were evaluated in augmented block design along with five checks (BGDS-1033, BGDS-1063, SCS-1061, SCS-1062 and SCS-793).

Morphological traits like plant height, number of sympodia per plant and number of bolls per plant were recorded. Yield and yield attributing traits *viz.,* boll weight, seed cotton yield per plant, lint yield per plant, seed index, lint index and ginning out turn recorded along with quality parameter of the cotton fibrewere recorded after ginning like upper half mean length, fibre strength, micronaire value, uniformity index and fibre elongation. The above observations were recorded on randomly selected five plants per plot and the mean was considered for statistical analysis.

Correlation and principal component analysis (PCA) are important statistical tools used to study relationships among traits and to analyze genetic variability among germplasm lines. Correlation analysis measures the degree of association between yield and its component. This helps breeders to perform indirect selection for yield through strongly associated characters. When many traits are involved and show multicollinearity, PCA is applied to reduce them into a few uncorrelated principal components using eigenvalues and eigenvectors derived from the covariance or correlation matrix. Statistical software such as R (packages like *agricolae*, *FactoMine R*), SPSS (Bivariate Correlation and Principal Component extraction) and GEN STAT (widely used for breeding trials) were used. These software tools facilitate accurate computation, visualization (correlation matrix and scree plot) and interpretation of results for effective crop improvement programmes.

Results and discussion

Descriptive statistics

Trait-wise analysis revealed a wide range of variation among the studied genotypes (Table 1). Plant height (PH) varied from 79.00 to 162.00 cm with a mean of 113.15 cm, while the number of sympodia per plant (NS) ranged from 8.20 to 20.00 with a mean of 13.40. The number of bolls per plant (NB) showed considerable variation (8.40 to 43.00, mean 20.88). Boll weight (BW) averaged 3.69 g, ranging between 2.43 g and 5.45 g, whereas lint yield (LY) exhibited wide variability (4.79 to 76.71 g, mean 27.51 g). Seed index (SI) and lint index (LI) recorded mean values of 8.35 and 4.37, with ranges of 6.30 to12.90 and 2.57 to 6.79, respectively. Ginning outturn (GOT) ranged from 26.10 to 42.08 % with mean of 34.34%. Seed cotton yield (SCY) showed substantial variability (17.65 to 205.28 g, mean 79.30 g). Fibre quality traits such as upper half mean length (UHML), fibre strength (STR and Micronaire value (MIC) recorded mean values of 26.06 mm, 26.31 g/tex and 4.03, with ranges of 19.25 to 32.38 mm, 17.10 to 35.20 g/tex and 2.43 to 5.83, respectively. The wide range in most traits indicates considerable genetic variability among the genotypes, providing scope for selection and improvement in breeding programs.

Table 1. Descriptive statistics of different trait in cotton for yield and fibre quality traits.

	Mean	Minimum	Maximum
PH	113.15	79.00	162.00
NS	13.40	8.20	20.00
NB	20.88	8.40	43.00
BW	3.69	2.43	5.45
LY	27.51	4.79	76.71
SI	8.35	6.30	12.90
GOT	34.34	26.10	42.08
LI	4.37	2.57	6.79
SCY	79.30	17.65	205.28
UHML	26.06	19.25	32.38
STR	26.31	17.10	35.20
MIC	4.03	2.43	5.83

Plant height (PH), Number of sympodia per plant (NS), Number of bolls per plant (NB), Boll weight (BW), Lint yield (LY), Seed index (SI), Lint index (LI), Ginning outturn (GOT), Seed cotton yield (SCY), Upper half mean length (UHML), Fiber strength (STR) and Micronaire value (MIC).

Correlation coefficient analysis

The association between various plant characteristics is measured *via*., correlation coefficient analysis. Determining the phenotypic relationship between traits offers important information for initiating successful breeding initiatives. When two attributes are positively and significantly correlated, it means that improving one will benefit the other. Therefore, selection for one trait can simultaneously enhance other positively associated traits. In present study,the plant height exhibited significant and positive correlation with lint yield (0.305), fibre strength (0.302) and seed cotton yield (0.313) the similar findings were observed by Abdullah *et al.* (2016) Anjani *et al.* (2020) and Nawaz *et al.* (2019). The results clearly indicated that even though there is no difference in important traits like number of bolls, boll weight and no of sympodia they contributing indirectly to seed cotton yield *via* plant height because these traits are positive in their contribution. The number of sympodia exhibit significant and positive association with number of bolls (0.373). It demonstrates how an increase in sympodia causes an increase in nodes, which in turn causes an increase in bolls. The number of bolls exhibited significant and positive correlation with lint yield (0.745), ginning out turn (0.347), micronaire value (0.223) and seed cotton yield (0.736) which indicate that as number of bolls increase which directly contributing to seed cotton yield hence so, selection of number of bolls trait would be fruitful. The results were consistence with Abdullah *et al.* (2016) and Nikhil *et al.* (2018).While seed index (-0.219) exhibited negative and significant association with this trait and boll weight (0.121) exhibited negative and non-significant association with number of bolls as it demonstrates as number of bolls increase the boll weight decreases hence there should be balance between selection number of bolls and boll weight while improving the following traits. The boll weight exhibited significant and positive

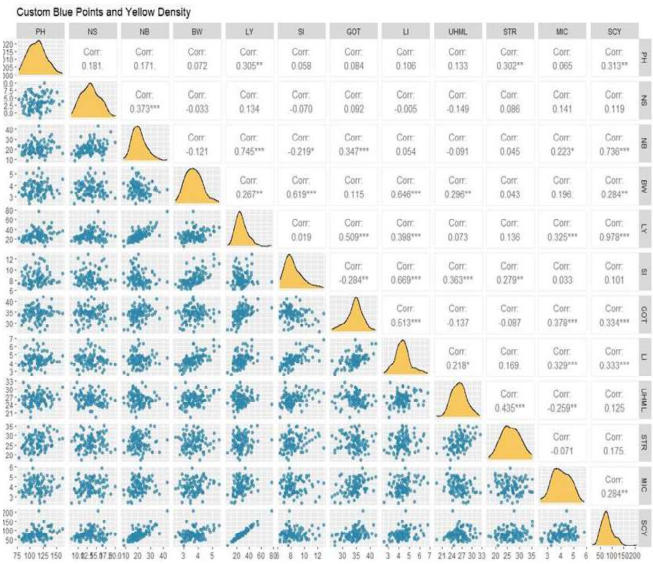

Fig. 1. Depiction of correlation plot for yield and fibre quality traits in cotton germplasm lines.

association with lint yield (0.267), seed index (0.619), lint index (0.646), upper half mean length (0.296) and seed cotton yield (0.284).The similar observations in accordance with Nikhil *et al.* (2018) Joshi *et al.* (2018) and Rizwan *et al*. (2022). The lint yield exhibited significant and positive correlation with trait ginning out turn (0.509), lint index (0.398), Micronaire value (0.325) and seed cotton yield (0.978). The similar result reported by Kumar *et al*. (2019). The seed index exhibited positive and significant correlation with lint index (0.669), upper half mean length (0.363) and fibre strength (0.279).While ginning out turn showed negative and non-significant association with this trait. The ginning outturn exhibited positive and significant correlation with lint index (0.513), micronaire value (0.378) and seed cotton yield (0.334). while negative and non-significant association with upper half mean length and fibre strength. The lint index exhibited positive and significant correlation with upper half mean length (0.218), micronaire value (0.329) and seed cotton yield (0.333). The upper half mean length (0.125) and fibre strength (0.175) showed positive and non-significant association with seed cotton yield, respectively. The micronaire value (0.284) showed positive and significant association with seed cotton yield. The following results are in correspondence with Manonmani *et al.* (2019), Asif *et al.* (2022), Abdullah *et al.* (2016), Raza *et al* , Zamir *et al* and Pradeep *et al.* (2014). The following results are mentioned in fig 1.

Plant height (PH), Number of sympodia per plant (NS), Number of bolls per plant (NB), Boll weight (BW), Lint yield (LY), Seed index (SI), Lint index (LI), Ginning outturn (GOT), Seed cotton yield (SCY), Upper half mean length (UHML), Fiber strength (STR) and Micronaire value (MIC).

Principle component analysis

Biometrical techniques *i.e.,*principal component analysis (PCA) have been repeatedly used to identify the genetic diversity in different genotypes (Brown-Guedira *et al.,* 2000). Estimation of genetic diversity through principal component analysis has made a way of recognition of phenotypic variability (Sarwar *et al.,* 2021). The variance is divided into its components for maintenance and exploitation of genetic diversity. Principal component analysis (PCA) is a useful technique to explore genotypes for successful breeding strategies (Nazir *et al.,* 2013). In present study, the total variance was divided into 12 components. Two-dimensional representation was illustrated by using the first two PCs. The first 4 PCs out of the total of the 12 PCs displayed eigen-values (> 1) and had maximum share to total variability. PC-I, II, III and IV had a share of 30.08, 20.81, 13.50 and 8.77 to total variability, respectively (Table 2).

PC-1 exhibited negative loading factor for all the traits plant height, number of sympodia, number of bolls, boll wight, lint yield, seed cotton yield, lint index, seed index, ginning out turn, upper half mean length, fibre strength and micronaire value while the PC-2 demonstrated positive loading factor for traits like number of sympodia, number of bolls, lint yield, seed cotton yield, ginning out turn and micronaire value and negative loading factor for plant height, boll wight, lint index, seed index, upper half mean length and fibre strength. The PC-3 exhibited positive loading factor for boll weight, seed index, ginning outturn, lint index and micronairevalue and negative loading factor for plant height, number of sympodia, number of bolls,

Table 2. Eigen Values, percent variation and cumulative variance percentage of Principal Components.

Principal Components	eigen value	Percentage of variance	Cumulative percentage of variance
PC1	3.61	30.08	30.08
PC2	2.49	20.81	50.90
PC3	1.62	13.50	64.40
PC4	1.05	8.77	73.17
PC5	0.86	7.21	80.39
PC6	0.73	6.14	86.53
PC7	0.65	5.48	92.02
PC8	0.47	3.96	95.98
PC9	0.32	2.73	98.71
PC10	0.14	1.21	99.92
PC11	0.07	0.05	99.98

Table 3 Principal component analysis of different traits of cotton.

	PH	NS	NB	BW	LY	SI
PC1	-0.202	-0.125	-0.343	-0.251	-0.48	-0.139
PC2	-0.013	0.191	0.351	-0.399	0.148	-0.534
PC3	-0.365	-0.148	-0.200	0.253	-0.088	0.062
PC4	-0.302	-0.739	0.076	-0.039	0.239	-0.178
	GOT	LI	SCY	UHML	STR	MIC
PC1	-0.301	-0.351	-0.462	-0.085	-0.128	-0.248
PC2	0.204	-0.324	0.105	-0.387	-0.233	0.115
PC3	0.313	0.311	-0.161	-0.358	-0.485	0.38
PC4	0.111	-0.069	0.237	0.276	-0.157	-0.304

Plant height (PH), Number of sympodia per plant (NS), Number of bolls per plant (NB), Boll weight (BW), Lint yield (LY), Seed index (SI), Lint index (LI), Ginning outturn (GOT), Seed cotton yield (SCY), Upper half mean length (UHML), Fiber strength (STR) and Micronaire value (MIC).

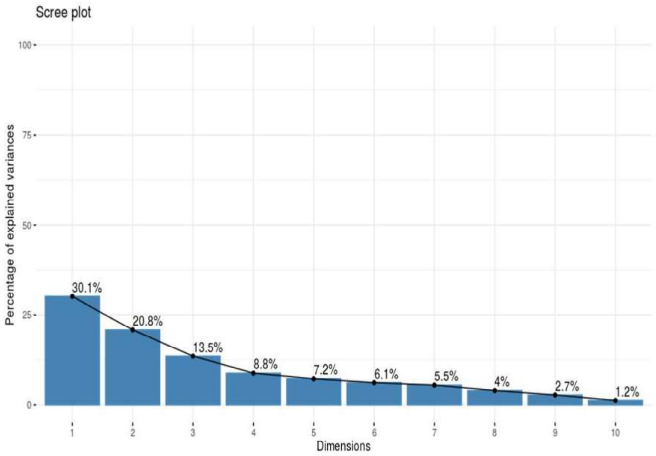

Fig 2. Scree plot showing Eigen values of Principal Components.

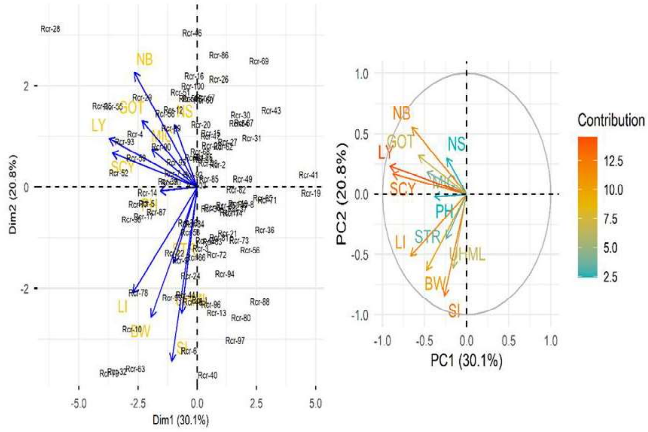

Fig 3. Biplot between Principal component 1 and Principal component 2 showing contribution of different variables.

lint yield, seed cotton yield, upper half mean length and fibre strength. The PC-4 demonstrated positive factor loading for number of bolls, lint yield, ginning out turn, seed cotton yield and upper half mean length and negative factor loading plant height, number of sympodia, boll wight, seed index, lint index, fibre strength and micronaire value (Table3).

Scree plot

Scree plot demonstrated the variance percentage in accordance with all principal components illustrated by a graph between the eigenvalues and principal components (Fig. 2). PC1 displayed the highest variability of 30.1% with the eigenvalue of 3.61. PC1 had maximum variability so the genotypes in PC1 should opt for selection. The similar results were according to Adeela *et al*., 2021. The following outcomes are mentioned in fig 2

Biplot

The PCA biplot provides a clear visual representation of trait relationships and helps evaluate germplasm performance based on their connection with key contributing traits. The length of each trait vector indicates its contribution to the total variation, with longer vectors showing a greater influence of the trait on economically important traits such as yield. Lint yield exhibited the longest vector length in the PCA biplot, indicating its major contribution to total genetic diversity, followed by seed cotton yield, seed index, number of bolls and boll weight where as plant height, number of sympodia, fibre strength and micronaire value showed minimum differences as they were close to the origin (Fig. 3). The similar result was according to Mangi *et al*., 2024

Conclusion

Correlation analysis showed that seed cotton yield had a strong and positive association with lint yield, number of bolls and boll weight, suggesting these traits as reliable selection criteria for yield enhancement. Principal component analysis further revealed that the first four components accounted for most of the total variability, with lint yield, seed cotton yield, seed index, number of bolls and boll weight being the major contributors. These results confirm the existence of considerable diversity among the evaluated germplasm lines, which can be effectively exploited for developing superior cotton varieties.

References

Abdullah M A, Muhammad Numan M N, Shafique M S, Awais Shakoor A S, Shamsur Rehman S R and Ahmad M I, 2016, Genetic variability and interrelationship of various agronomic traits using correlation and path analysis in cotton (*Gossypium hirsutum* L.). *Academic Journal of Agricultural Research,* 4(6), 315-318.

Adeela S, Zafar M M, Razzaq A, Manan A, Muhammad H, Sunaina S, Abdul R, Huijuan M, 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.

Ahsan M Z, Laghari M A, Bhutto H, Soomro A W, Majidano M and Malik T H, 2015, Genetic variability in different biometric traits of upland cotton genotypes. *International Journal of Biology and Biotechnology*, 12: 607-612.

Anjani A V, Padma J, Ramana Y and Satish Y, 2020. Evaluation of genetic parameters of agro-morpho-quality traits in American cotton *(Gossypium hirsutum* L.*).Electronic Journal of Plant Breeding*, 11(1): 279-282.

Asif M, Khan A A, Naseer Cheema H M, Khan S H and Iqbal Z, 2022, Genetic variability in diverse cotton germplasm for drought tolerance.*Pakistan Journal of Agricultural Sciences*, 59(1): 63-74.

Brown-Guedira G L, 2000, Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. *Crop Science*, 40: 815-823.

Farooq M A, Shakeel A, Atif R M and Saleem M F, 2019, Genotypic variations in salinity tolerance among *Bt* cotton. *Pakistan Journal of Botany*, 51(6): 1945-1953.

Jabran K and Chauhan B S (eds.), 2019, Cotton Production. John Wiley & Sons.

Joshi V and Patil B, 2018, Genetic variability and heritability study in F_2 population for yield, yield attributes and fibre quality traits in cotton *(Gossypium hirsutum* L.*). Journal of Pharmacognosy and Phytochemistry*, 7(4): 2816-2818.

Manonmani K, Mahalingam L, Malarvizhi D, Sritharan N and Premalatha N, 2019, Genetic variability, correlation and path analysis for seed cotton yield improvement in upland cotton *(Gossypium hirsutum* L.*).Journal of Pharmacognosy and Phytochemistry*, 8(4): 1358-1361.

Mangi N, Iqbal M S, Shuli F, Iqbal M T, Alharthi B, Jatoi G H, Sarfraz Z, Ma Q and Sun X, 2024, Empowering cotton breeding programs through the strategic exploration and exploitation of phenotypic diversity of genetic resources under climate change conditions. *Plant Stress*, 13: 100548.

Nawaz B, Naeem M, Malik T A, Muhae-Ud-Din G, Ahmad Q and Sattar S, 2019, Estimation of gene action, heritability and pattern of association among different yield related traits in upland cotton. *International Journal of Innovative Approaches in Agricultural Research*, 3(1): 25-52.

Nawaz B, Sattar S and Malik T A, 2019, Genetic analysis of yield components and fiber quality parameters in upland cotton. *International Multidisciplinary Research Journal*, 9(9): 13-19.

Nikhil P G, Nidagundi J M and Anusha H A, 2018. Correlation and path analysis studies of yield and fibre quality traits in cotton *(Gossypium hirsutum* L.*). Journal of Pharmacognosy and Phytochemistry*, 7(5): 2596-2599.

Pradeep S, Satish Y, Babu D R and Rao V S, 2014, Character association and path coefficient analysis for yield and yield component traits in upland cotton *(Gossypium hirsutum* L.*). The Journal of Research ANGRAU,* 42(3 & 4): 32-35.

Raza H, Khan N U, Khan S A, Gul S, Latif A, Hussain I, Khan J, Raza S and Baloch M, 2016, Genetic variability and correlation studies in F_4 populations of upland cotton. *Journal of Animal and Plant Sciences*, 26(4): 1048-1055.

Rizwan M, Farooq J, Farooq A, Farooq M, Sarwar G, Nadeem M, Riaz M, Shahid M R, Abbas H G and Sarwar M K S, 2022, Yield and related components of cotton *(Gossypium hirsutum* L.*)* affected by chlorophyll contents. *Pakistan Journal of Agricultural Research*, 35(1): 29-35.

Salahuddin S, Abro S, Kandhro M M, Salahuddin L and Laghari S, 2010, Correlation and path coefficient analysis of yield components of upland cotton (*Gossypium hirsutum* L.) sympodial. *World Applied Sciences Journal*, 8(8): 71-75.

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.

Ullah A, Shakeel A, Ahmed H G M D, Naeem M, Ali M, Shah A N, Wang L, Jaremko M, Abdelsalam N R, Ghareeb R Y and Hasan M E, 2022, Genetic basis and principal component analysis in cotton *(Gossypium hirsutum* L.*)* grown under water deficit condition. *Frontiers in Plant Science*, 13: 981369.

Zamir A, Farooq J, Mahmood A, Shahid M and Riaz M, 2013, Estimation of genetic diversity for CLCuV, earliness and fiber quality traits using various statistical procedures in different crosses of *Gossypium hirsutum* L. *Pakistan Journal of Botany*, 43(4): 2-9.