# Genetic diversity studies in some advanced lines of *Gossypium hirsutum* L. for yield and quality related attributes using cluster and principle component analysis

Jehanzeb Farooq, Muhammad Rizwan, Iram Sharif, Sadaf Saleem, Shahid M. Chohan, Riaz A. Kainth

Cotton Research Station, AARI, Faisalabad, Pakistan. Corresponding author: J. Farooq, jehanzeb1763@hotmail.com

**Abstract**. To study the genetic variability in 20 different genotypes of cotton, a preliminary yield trial was designed at Cotton Research Station, Faisalabad, Pakistan. The results of cluster analysis revealed that cluster 3 comprised of genotypes that have average yield of 3,802 kg/ha and good fiber quality traits while cluster 2 genotypes showed good potential along with staple length of 30 mm. The study of principal components revealed that out of 9 PCs 3 components exhibited eigen value greater than 1. These three PCs contributed 69.2% of the total variability amongst the genotypes assessed for yield, agronomic and fiber traits. The remaining components contributed only 30.8% towards the total variability. The traits like yield, sympodia per plant and bolls per plant showed considerable positive factor loadings on PC I while on PC II, maximum positive loading was shown by plant height, Ginning Out Turn (GOT) and fiber strength and maximum negative loadings were exhibited by fiber length. The diversification of genotypes as shown by the results of cluster and principle component analyses would be helpful to obtain superior parents that can be utilized in future breeding strategy and superior genotypes may also be tested in provincial and national yield trials due to a combination of superior yield and quality traits.
**Key Words**: multivariate analysis, variability, cotton, genotypes, Pakistan.

**Introduction**. Cotton is an important cash and single most grown fiber crop in Pakistan which provide major oil demands of the country (Farooq et al 2014). It is grown in more than 80 countries for its fiber and edible oil (Shakeel et al 2011). It has major role in financial stability of the country. Cotton provides raw material to the textile, oil and livestock industries as cotton lint, crude oil and cotton seed cake respectively (Ali et al 2009). In Pakistan 75% of population is directly or indirectly dependent on cotton sector for employment.

Globally four kinds of cotton species are cultivated having two diploids (Old world) and two tetraploid (New world) species. Genetic diversity in any crop is the baseline of breeding program as abrupt fluctuation in climate is responsible for unforeseen effects on plant. Therefore, narrow genetic base is one of the reasons of crop failure. Various researchers reported narrow genetic base in cultivated cotton (McCarty et al 2005). Cotton breeders and growers showed concern on less diverse genotypes which are responsible for yield reduction and quality deterioration (Rathinavel 2017). Genetically distant genotypes are the prerequisite to widen the genetic base in hybrid breeding program (Shakeel et al 2015).

Utilization of prevailing genetic diversity in term of morphological descriptor leads toward successful hybridization program and construction of evolutionary relationship (Zada et al 2013). Genetic variability in term of morphological and quality parameters has been studied by many scientists for improvement of cotton genotypes (Li et al 2005). Yield improvement needs better understanding about crop nature, genotype behavior and association of agronomic traits with yield to overcome yield limiting factors (Latif et al 2015). High yielding cotton genotypes could be selected on the base of morphological

descriptors: sympodial branches, monopodial branches, bolls per plant, boll weight, seed index and Ginning Out Turn (GOT) (Shazia et al 2010).

Principal component analysis is used mostly by the breeders to assess genetic diversity, which helps in exploration of promising genotypes for future breeding programme (Rehman et al 2015). Among all other biometrical techniques, PCA has the edge to assigned only one group to each genotype at the same time it also depicts the importance of major contributor toward total diversity at each axis of differentiation (Sharma 2006).

The present study was designed to investigate the extent of genetic diversity in eighteen genotypes with two check varieties in terms of yield related parameters by utilizing morphological markers. This study is helpful in classifying available genotypes into different clusters depending upon yield related attributes. The information generated through this research will assist in development of high yielding cotton varieties.

## Material and Method

***Plant material and site characteristics***. Eighteen genotypes along with two check varieties were evaluated for preliminary yield testing during crop season 2016-2017 at experimental area of Cotton Research Station, Faisalabad, Punjab, Pakistan which is present at an altitude of 184 m at 31° 21′ 52″ N 72° 59′ 40″ E with average rainfall of 300 mm. The experiment was sown on 1st May to evaluate genotypes for yield, agronomic and fiber parameters. The traits like sympodia and bolls per plant, plant height, boll weight, seed cotton yield, fiber length, fiber strength, fiber fineness and GOT were studied.

***Experimental design***. For each entry, plot size measured 4.54 × 3.03 m, comprising four rows set 75 cm apart. Distance between plants within rows was 30 cm. Normal agronomic, cultural practices (irrigation, weeding, hoeing, and fertilizer applications), and plant protection measures were adopted as and when required. The experiment was designed in randomized complete block fashion with three replications.

***Measurement of traits***. For measuring the traits, the 10 undamaged and true to type plants were selected in each line and marked for identification. Data for sympodia per plant and bolls per plant was measured in numbers by counting the number of fruiting and vegetative branches. Plant height was measured in centimeters from the base to the tip of the plant. Boll weight was measured in grams by collecting 25 bolls from base, top and middle of each tagged true to type plant. The weight of collected bolls was divided with number of collected bolls to get the weight of single boll. Seed cotton was picked in two pickings. 1st pick was done after 120 days and 2nd pick was done at maturity and recorded as kg/plot and extrapolated in kg/hectare. The harvest was weighed with the help of electrical balance and final averages were made from each tagged plant. Cleaned and dry samples of seed cotton were weighed and then ginned separately with single roller electric ginning machine. The lint obtained from each sample was weighed and GOT % was calculated by the following formula:

GOT (%) = Weight of lint / Weight of seed cotton × 100

Fiber characteristics such as staple length, fiber fineness of each guarded plant were measured by using spin lab HVI-1000. This computerized instrument provides us a true profile of raw fiber. It measures the most important characters such as staple length (mm) and fiber fineness (µg/inch) within a quick period of time according to international trading standards.

***Statistical analysis***. The average data of all the traits were subjected to basic statistics and PCA using statistical software packages of Minitab version 16.1 and STATISTICA version 8.1 (Sneath & Sokal 1973). First two principal components were plotted against

each other to find out the patterns of variability among genotypes and association between different traits using portable Minitab version 17.

**Results and Discussion**. The results of basic statistics of various studied traits showed that there is considerable genetic variability present in 20 cotton genotypes/varieties for 9 different studied traits. The minimum, maximum range of means of different traits is given in Table 1.

Table 1

Basic statistics for various traits of 20 cotton genotypes

| Trait | Minimum | Maximum | Mean±S.E. | Std. deviation | Variance |
|---|---|---|---|---|---|
| Sympodia per plant | 18.333 | 25.000 | 22.250 ± 0.483 | 2.160 | 4.665 |
| Bolls per plant | 25.00 | 52.33 | 35.90±1.77 | 7.93 | 62.90 |
| Plant height (cm) | 126.67 | 185.00 | 163.82±3.22 | 14.38 | 206.73 |
| Yield (kg/ha) | 2773.5 | 3801.6 | 3339.6±64.0 | 286.0 | 81802.9 |
| Fiber length(mm) | 27.207 | 30.907 | 28.717±0.220 | 0.982 | 0.965 |
| Fiber strength (g/tex) | 25.967 | 32.500 | 28.973±0.410 | 1.835 | 3.366 |
| Fiber fineness (mic) | 4.1200 | 5.4533 | 4.8443±0.0734 | 0.3284 | 0.1079 |
| Boll weight (g) | 3.3667 | 4.1000 | 3.8208±0.0443 | 0.1982 | 0.0393 |
| GOT % | 38.033 | 42.167 | 39.610±0.291 | 1.302 | 1.696 |

Khan et al (2017) utilized descriptive statistics, correlation and principal component analysis to assess the genetic variability in exotic lines of *Gossypium arboreum* for sympodial branches, boll weight, bolls per plant, seed cotton yield, lint percentage and micronaire values. The true utilization of genetic resources could be made by partition of the total variance into its different components. It also provides an opportunity for utilization of suitable germplasm in crop improvement for particular plant traits (Pecetti et al 1996). Germplasm screening and improvement for certain yield and quality related traits is crucial for breeding stress resilience high yielding cotton cultivars (McCarty et al 2005). Genetic variation exists among cotton genotypes for the sake of crop improvement in terms of yield and fiber quality related traits (Alishah 2001). The utilization of various biometrical techniques including principal component analysis and linkage cluster analysis for genotype characterization into different clusters based on their performance for yield related attributes have been reported in previous studies (Qiaoling & Zhe 2011). While studying clusters it was revealed that the most superior cluster in term of yield was cluster 3 with maximum height of plants, bolls per plant, boll weight, higher fiber strength, higher GOT % and lowest micronaire value. The cluster 1 was elucidated with superior fiber length and moderate seed cotton yield. Both cluster 1 and 3 genotypes may be combined to obtain new combinations that can perform even better than their parents due to the presence of high yielding and better quality parents in both these clusters. Some of these genotypes may also be recommended for provincial and national testing because they possess the quality traits that fit to the standards required for varietal approval in Pakistan. The cluster details and cluster membership is given in Table 2 and 3 respectively and dendrogram showing these clusters is given in Figure 1.

While studying 9 principal components (PCs) only 3 have eigen value more than 1. These three PCs contributed 69.2% of the total variability amongst the genotypes assessed for yield, fiber and agronomic traits (Table 4). However, the remaining components contributed only 30.8% towards the total diversity. The PC I contributed maximum towards the variability (30.5%) followed by PC II (22.1%) and PC III (16.7%). It was reported that 1st two principal components play key role in describing major variability contributor while investigating different traits (Saeed et al 2014).

Table 2

Cluster analysis of various yield and fiber quality traits in cotton

| Variable | Cluster-1 | Cluster- 2 | Cluster-3 |
|---|---|---|---|
| Sympodia per plant | 24 | 19 | 25 |
| Bolls per plant | 32 | 29 | 46 |
| Plant height (cm) | 174 | 157 | 185 |
| Yield (kg/ha) | 3329 | 2774 | 3802 |
| Fiber length (mm) | 30 | 29 | 28 |
| Fiber strength (g/tex) | 28 | 29 | 32 |
| Fiber fineness (mic) | 5.0 | 4.9 | 4.8 |
| Boll weight (g) | 4.0 | 4.1 | 4.2 |
| GOT % | 40 | 39 | 42 |

Table 3

Cluster membership of different genotypes

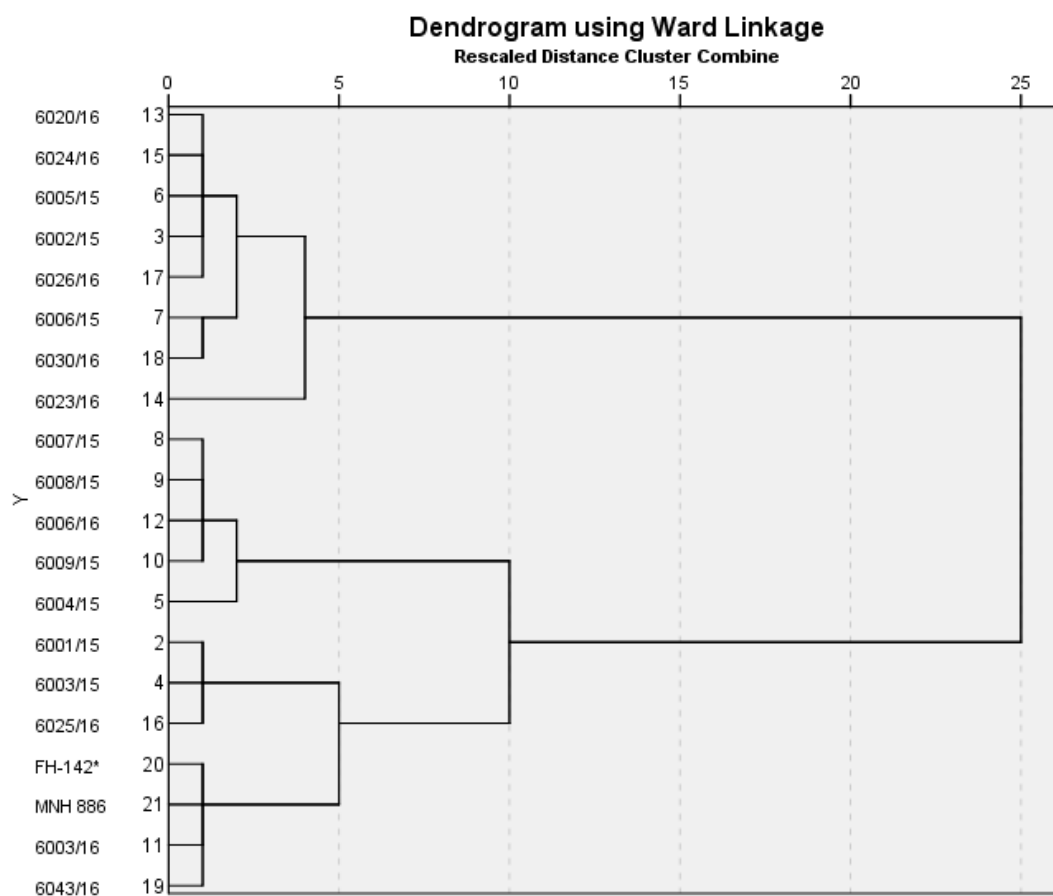| Cluster number | No. of genotypes | Genotype name |
|---|---|---|
| Cluster -1 | 7 | 6001/15,6003/15,6003/16,6025/16,6043/16,FH-142, MHH-886 |
| Cluster -2 | 8 | 6002/15,6005/15,6006/15,6020/16,6023/16,6024/16,6026,16,6030/16 |
| Cluster -3 | 5 | 6004/15,6007/15,6008/15,6009/15,6006/16 |



Figure 1. Dandrogram showing the allocation of genotypes in different clusters.

Table 4

Principle component analysis for various morphological traits in cotton

| Variable | PC I | PC II | PC III |
|---|---|---|---|
| Eigen value | 2.74 | 1.99 | 1.50 |
| % of total variance | 30.5 | 22.1 | 22.1 |
| Cumulative variance % | 30.5 | 52.6 | 69.2 |
| Factor loadings by various traits | | | |
| Variable | PC I | PC II | PC III |
| Sympodia per plant | 0.506 | -0.030 | -0.190 |
| Bolls per plant | 0.430 | -0.225 | -0.431 |
| Plant height (cm) | 0.098 | 0.512 | 0.395 |
| Yield (kg/ha) | 0.519 | 0.082 | -0.070 |
| Fiber length (mm) | 0.072 | -0.446 | 0.438 |
| Fiber strength (g/tex) | -0.266 | 0.406 | -0.225 |
| Fiber fineness (mic) | -0.172 | 0.203 | -0.573 |
| Boll weight (g) | 0.378 | 0.312 | 0.185 |
| GOT % | 0.179 | 0.421 | 0.109 |

The traits like fiber strength and fiber fineness showed considerable negative factor loadings on PC I while yield, sympodia per plant and bolls per plant had maximum positive loadings. Maximum positive loading on PC II was shown by plant height while maximum negative loadings were exhibited by fiber length. The PC III was elucidated by diversity among the genotypes for fiber length with maximum positive and fiber fineness with maximum negative loadings. PC analysis confirmed the extent of variation for the traits among the material studied which could be utilized in designing a breeding programme aimed at improving sympodia per plant, boll weight, bolls per plant and ultimately seed cotton yield (Nazir et al 2013). Genetic diversity for morphological traits has been estimated using principal component analysis, which leads to identification of phenotypic variability in cotton (Li et al 2008). Many researchers reported that prevailing diversity in colored cotton germplasm may offer wide scope in field of colored cotton characterization (Malik et al 2011; Ashokkumar & Ravikesavan 2011). The biplot drawn by principal component represents that variables are super imposed as vector (Figure 2). From origin, the length of vector describes the relative variability contribution made by each variable. The genotypes which fall away from the origin are considered more diverse and vice versa. These findings are also supported by the results of Saeed et al (2014), who observed leading role of PC1 and PC2 for diversity assessment of different genotypes. The biplot exhibited that as a whole plant height, fiber length, fiber strength and fiber fineness contributed maximum towards variation in the genotypes studied. Biplot also revealed strength of correlation among characters studied. The estimates of PC and cluster analysis in the studies of Shakeel et al (2015) revealed that the identification of suitable genotypes having good yield and better quality would be helpful in designing efficient breeding programmes. Principal component analysis (PCA) is more appropriate among these techniques because it reveals the significance of largest contributor to the total variability at each axis of differentiation (Sharma 2006).

As it is depicted in scoreplot and biplot, the genotypes 6007/15, 6004/15, 6008/15, 6006/16, 6043/16, MNH-886 and FH-142 have good potential for yield, sympodia, boll weight and bolls per plant (Figure 2 & 3). While considering fiber length, the genotypes 6007/15, 6003/16, 6025/16, 6026/16 and 6043/16 have good potential. The genotypes 6004/15, 6009/15, 6026/16, 6008/15, MNH-886 and FH-142 have good potential for GOT percentage. The results obtained using principle component and cluster analysis provided excellent information which may be exploited to obtain genotypes that can be utilized in future breeding programmes.
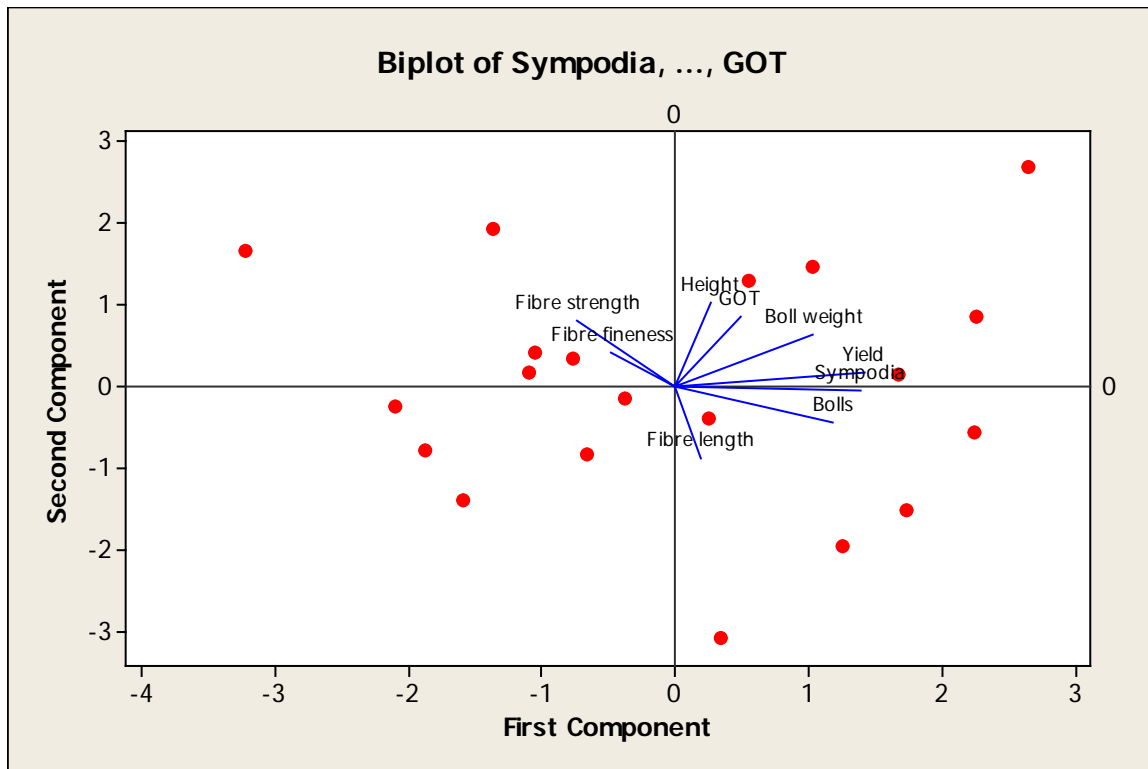
**Figure 2.** Biplot between PC-1 and 2 showing contribution of various traits in variability among different genotypes.
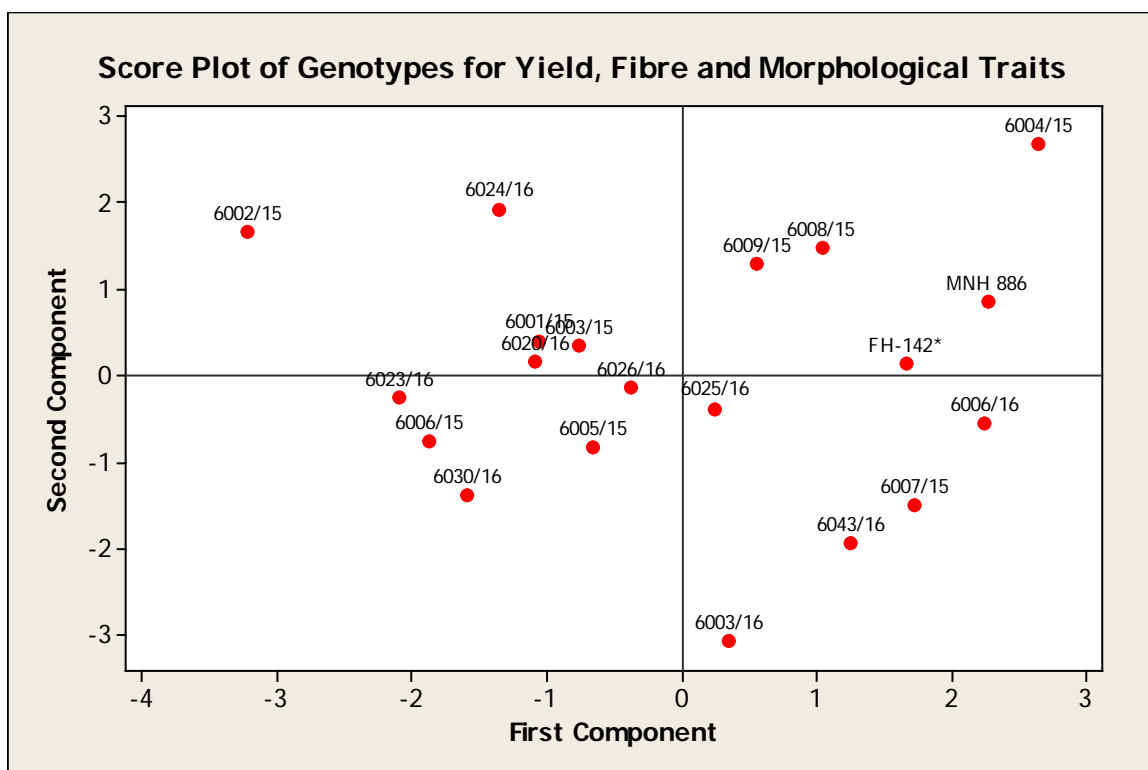


**Figure 3.** Score plot between PC-1 and 2 showing contribution of various traits in variability among different genotypes.

**Conclusions**. Morphological and quality parameters determine general varietal behavior in its growing environment. The classification of genotypes into different clusters provides

instant glance for selection of parents in breeding program. Diverse genotypes could be a potential source for transfer of desirable genetic combination for yield improvement. Dendrogram construction give raise three different clusters/clades, within cluster genotypes are more similar to each other showing the ancestral relationship while between clusters they become quite dissimilar to each other. The genotypes 6023/16, 6004/15, 6006/15 were found more diverse while 6007/15, 6004/15, 6008/15, 6006/16, 6043/16, MNH-886 and FH-142 showed maximum yield potential which suggest their utilization in future breeding programs.

## References

Ali M. A., Abbas A., Younas M., Khan T. M., Hassan H. M., 2009 Genetic basis of some quantitative traits in upland cotton. Plant Omics Journal 2(2):91-97.

Alishah E., 2001 Study of morphological traits and genetic variation in different genotypes of upland cotton (*Gossypium hirsutum* L.) in Iran. Seed and Plant Improvement Journal 17:44-60.

Ashokkumar K., Ravikesavan R., 2011 Morphological diversity and per se performance in upland cotton (*Gossypium hirsutum* L.). Journal of Agriculture Science 3(2):107-113.

Farooq J., Anwar M., Riaz M., Farooq A., Mahmood A., Shahid M. T. H., Shahid M. R., Ilahi F., 2014 Correlation and path coefficient analysis of earliness, fiber quality and yield contributing traits in cotton (*Gossypium hirsutum L.*). J Anim Plant Sci 24(3):781-790.

Khan M. I., Haq H. A., Ullah K., Arshad M., Majid A., 2017 Genetic diversity and correlation studies for cotton leaf curl disease (CLCuD), fiber & yield related attributes in exotic lines of *Gossypium arboreum* L. Am J Plant Sci 8:615-624.

Latif A., Bilal M., Hussain S. B., Ahmad F., 2015 Estimation of genetic divergence, association, direct and indirect effects of yield with other attributes in cotton (*Gossypium hirsutum* L.) using biplot correlation and path coefficient analysis. Tropical Plant Research 2(2):120–126.

Li Z., Wang X., Yan Z., Guiyin Z., Wu L., Jina C., Ma Z., 2008 Assessment of genetic diversity in glandless cotton germplasm resources by using agronomic traits and molecular markers. Front Agric China 2:245-252.

Li R. I., Long Z., Chen J., 2005 Comparison of fiber quality characteristics of different varieties of hybrid cotton. J Hunan Agric Uni 31:492-pages.

Malik W., Iqbal M. Z., Khan A. A., Noor E., Qayyum A., Hanif M., 2011 Genetic basis of variation for seedling traits in *Gossypium hirsutum* L. Afr J Biotechnol 10:1099-1105.

McCarty J. C., Jenkins J. N., Wu J., 2005 Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement. Crop Sci 44:1231-1235.

Nazir A., Farooq J., Mahmood A., Shahid M., 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. Vestnik Orel GAU 43(4):2-9.

Pecetti L., Annicchiario P., Damania A. B., 1996 Geographic variation in tetraploid wheat (*Triticum turgidum* spp. *Turgidum* convar. Durum) landraces from two provinces in Ethiopia. Genet Resour Crop Evol 43:395-407.

Qiaoling W., Zhe L., 2011 Principal component analysis of F2 individual selection in upland cotton (*Gossypium hirsutum* L.). J. Henan Inst Sci Technol (Natural Sci Ed) 5:004- pages.

Rathinavel K., 2017 Exploration of genetic diversity for qualitative traits among the extant upland cotton (*Gossypium hirsutum* L.) varieties and parental lines. Int J Curr Microbiol App Sci 6(8):2407-2421.

Rehman S. U., Abid M. A., Bilal M., Ashraf J., Liaqat S., Ahmed R. I., Qanbar G., 2015 Genotype by trait analysis and estimates of heritability of wheat (*Triticum aestivum* L.) under drought and control conditions. Basic Res J Agri Sci Rev 4(4):127-134.

Saeed F., Farooq J., Mahmood A., Riaz M., Hussain T., 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. Aust J Crop Sci 8(3):442-447.

Shakeel A., Farooq J., Ali M. A., Riaz M., Farooq A., Saeed A., Saleem M. F., 2011 Inheritance pattern of earliness in cotton (*Gossypium hirsutum* L.). Aust J Crop Sci 5:1224-1231

Shakeel A., Talib I., Rashid M., Saeed A., Ziaf K., Saleem M. F., 2015 Genetic diversity among upland cotton genotypes for quality and yield related traits. Pak J Agric Sci 52(1):73-77.

Sharma J. R., 2006 Statistical and biometrical techniques in plant breeding. New Age International Publishers, 432 pp.

Shazia S., Saifullah A., Kandhro M., Salahuddin L., Laghari S., 2010 Correlation and path coefficient analysis of yield components of upland cotton (*Gossypium hirsutum* L.) sympodial. World Appl Sci J 8:71-75.

Sneath P. H. A., Sokal R. R., 1973 Numerical taxonomy: the principles and practice of numerical classification. W. F. Free-Man and Co. San Francisco, 573 pp.

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

Authors:
Jehanzeb Farooq, Cotton Research Station, AARI, Pakistan, Faisalabad, 38060, e-mail: Jehanzeb1763@hotmail.com
Muhammad Rizwan, Cotton Research Station, AARI, Pakistan, Faisalabad, 38060, e-mail: muhammadrizwan29@gmail.com
Iram Sharif, Cotton Research Station, AARI, Pakistan, Faisalabad, 38060, e-mail: iramsharif695@yahoo.com
Sadaf Saleem, Cotton Research Station, AARI, Pakistan, Faisalabad, 38060, e-mail: sadafsaleempbg@gmail.com
Shahid Munir Chohan, Cotton Research Station, AARI, Pakistan, Faisalabad, 38060, e-mail: cshahidmunir@yahoo.com
Riaz Ahmad Kainth, Cotton Research Station, AARI, Pakistan, Faisalabad, 38060, e-mail: cbcrsfsd@gmail.com