Open Access

Phytopathogenomics and Disease Control., *2024, 3(2)267-275*
ISSN (Online):2957-5842; ISSN (Print):2957-5834
DOI: https://doi.org/10.22194/Pdc/3.1035
https://societyfia.org/journal/PDC

Society for Innovative Agiculture

Assessment of Genetic Diversity in *Bt* Cotton Germplasm using Multivariate Analysis

Hafza Anam[1], Amir Shakeel[1,*], Asif Saeed[1], Azeem Iqbal Khan[1], Muhammad Jabran[2], Shahid Iqbal[3], Amjad Abbas[3] and Muhammad Amjad Ali[3*]

[1]Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan; [2]State Key Laboratory for Biology of Plant Diseases, Institute of Plant Protection, CAAS, 100193, Beijing, China; [3]Department of Plant Pathology, University of Agriculture, Faisalabad, Pakistan
*Corresponding author's e-mail: amjad.ali@uaf.edu.pk; amirpbg@uaf.edu.pk

The development of cotton production and quality is always challenging due to the narrow genetic base of present cotton germplasm, which highlights the urgent need to fully utilize the available germplasm resources to address the problems with food security. Therefore, the present study was intended to evaluate genetic diversity in fifty genotypes for yield and quality related traits at the experimental field of the University of Agriculture, Faisalabad, Pakistan. The trial was conducted using a randomized complete block design (RCBD) with three replications. The genetic diversity of 50 cotton genotypes was evaluated using multivariate analysis. The results revealed that genotypes were genetically different from each other for all characters under study. In the factor analysis, the first 5 factors from 14 factors showed eigen value >1 and contributed 72.13 % to the total variability among 50 genotypes. The first factor contributed 24.72% towards total variability which is the highest value by a single factor. The outcome of factor analysis suggested that CRS-456 (36.55%), VH-282 (6.8%), VH-235 (4.3%) and FH-113 contributed maximum to the variability in germplasm. In cluster analysis using K-mean clustering and Ward's method, genotypes present in Cluster-II and cluster-III showed the greatest diversity for traits such as plant height, sympodial branches, bolls per plant, seeds per boll, seed cotton yield, lint index, and GOT%, genotypes from different groups could be selected for improvement in breeding for yield, quality, and other economic attributes in cotton crops. The genetic diversity, heritability, and moderate to high genetic advance suggest that the germplasm has potential for use in future breeding to produce high-yielding cotton varieties.

Keywords: Cotton, yield traits, factor analysis, cluster analysis, Bt varieties.

INTRODUCTION

Upland cotton (*Gossypium hirsutum* L.) is widely cultivated and is a profitable fiber crop cultivated in over 80 countries globally (Shahzad *et al.*, 2019). Cotton is considered as the backbone of Pakistan's economy being a major source for earning of foreign exchange (Rashid *et al.*, 2016). It contributes 1.0 % to GDP and 5.1% to the value addition in agriculture. However, Bt-cotton was introduced in Pakistan in 2010 to control Lepidopteron pests. Bt-cotton has been developed through transfer of a gene Cry1AC from soil bacterium *Bacillus thuringiensis*. It produces Cry proteins which are toxic to Lepidopteron pests (Sanahuja *et al.*, 2011). Bt cotton is very popular among farmers because it has the potential to increase productivity, savings on pesticides and labor as compared to non-Bt-cotton (Malik and Ahsan, 2016). Accordingly, Pakistan is holding 5th biggest producer of cotton regarding good quality of fiber after China, USA, India and Brazil (Arshad *et al.*, 2022). The area under cotton cultivation during 2020-21 decreased by 17.4% compared to the previous year, from 2,517 million hectares to 2,079 million hectares. The yield of cotton crop also decreased by 22.8% to 7.064 million bales, compared to 9.148 million bales from the previous year (Pakistan Economic Survey 2020-21). In 2015, the production was 10.074 million bales, which is a decrease of 27.8% compared to the previous year's production of 13.960 million bales. The decrease in cotton production was primarily caused by an infestation of pink bollworm, high levels of rainfall, and to a lesser extent, losses from leaf curl virus (CLCuV) (Rajani *et al.*, 2019).

Anam, H., A. Shakeel, A. Saeed, A.I. Khan, M. Jabran, S. Iqbal, A. Abbas and M.A. Ali. 2024. Assessment of Genetic Diversity in *Bt* Cotton Germplasm using Multivariate Analysis. Phytopathogenomics and Disease Control 3:267-275
[Received 15 Feb 2024; Accepted 2 Apr 2024; Published 18 Jun 2024]

To achieve sustainable cotton production with better tolerance against various environmental and biotic stresses, it is necessary to develop higher yielding varieties with superior fiber quality (Khan *et al.,* 2017). The selection procedure for yield is complicated due to the interaction between genetic and climatic factors. Yield is controlled by multiple genes that are significantly impacted by various climatic attributes. Information about the relationship between yield and its related components enables breeders to select desirable genotypes (Abbas *et al.,* 2013; Farooq et *al.,* 2015; Iqbal *et al.,* 2023). In comparison to the average productivity of other cotton-growing nations, Pakistan's cotton production is low down. This is because of biotic and abiotic conditions including high temperatures, the disease cotton leaf curl virus (CLCuV), pest attacks, and insufficient production methods (Nawaz *et al*., 2019). Hence, several strategies, including boosting inputs, using pesticides, and genetically enhancing best cultivars, can be used to resolve this issue (Ahmad *et al.,* 2012). Diversity in germplasm not only provides opportunities for selecting desirable combinations of genotypes but also increases resistance against various biotic and abiotic stresses. Evaluation of cotton germplasm divergence is crucial for initiating breeding programs, characterizing genotypes, identification of outstanding breeding material, and envisaging the inheritance patterns and heterosis level necessary for achieving breeding goals (Riaz *et al*., 2019). In the literature, different approaches are used for assessment of diversity like multivariate analysis D^2 statistics, principal component analysis, cluster analysis, factor analysis and principal coordinate analysis (Arifuzzaman *et al.,* 2020; Jabran *et al*., 2021). Similarly, factor analysis is also widely used in the cereal crop improvement and considered as a most satisfactory statistical method for the assessment of genetic diversity and character associations simultaneously among cotton accessions. In addition, it is helpful in clustering of germplasm accession based on deviations in morphological characteristics instead of the geographic origin of the genotypes (Khan *et al.,* 2017). Cluster analysis is a technique that categorizes similar genotypes into the same groups based on different yield and fiber traits under study (Khan *et al.,* 2015).

Therefore, to improve yield, fiber quality, and insect resistance, it is essential to have a comprehensive understanding of the breeding potential of Bt and the genetic diversity of cotton genotypes. The aim of the present research was to determine the nature and significance of genetic diversity present in Bt-cotton germplasm for morphological and quantitative yield and fiber attributes using multivariate analysis based on factor and cluster analysis approaches. The outcomes could be applied to the production of new cultivars and to the enhancement of current cotton genotypes.

MATERIALS AND METHODS

***Experimental Materials*:** The study was carried out to assess the genetic diversity in 50 Bt-cotton genotypes for yield and quality related traits. The genotypes were collected from different cotton research institutes of Punjab, during 2015-2016 given in (Table 1).

***Research methods*:** The field trial was conducted in three replications using a randomized complete block design (RCBD) in a field setting at the Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad (UAF). Each of the 50 plots in the replications was randomly assigned a genotype. The spacing between plants and rows was maintained at 75 cm and 30 cm, respectively. All recommended cultural and field practices were followed, including the application of fertilizer, irrigation, and plant protection.

***Data recording for different yield and fiber quality traits*:** At maturity, data were recorded for different morphological characters including plant height (cm), number of monopodial branches, number of sympodial branches, number of bolls per plant, boll weight (g), number of seeds per boll, seed cotton yield per plant (g), ginning out turn (GOT%), fiber length (mm), fiber fineness, fiber strength and uniformity ratio from five plants from each genotype per replication. The height of individual plant was recorded in 'cm' from the first node of cotyledon to apical bud.

Monopodial or indirect fruit bearing branches were situated at the bottom of the main stem. Sympodial or direct fruit bearing branches appeared as small from top and large towards bottom of stem. Bolls was picked from five plants of each genotype per replication. The total seed cotton yield of a plant was divided by bolls of that plant to obtain boll weight for individual plant.

The harvested bolls were weighed in grams with the help of electronic balance to obtained seed cotton yield. The seed cotton was ginned to obtain lint. Lint weight was recorded with the help of electronic balance and then GOT% was calculated according to Ali and Khan (2007). The seed index was determined in grams by weighing 100 seeds. Lint index was obtained as lint weight obtained from 100 seeds. Fiber characteristics like fiber length, fiber fineness, uniformity ratio and fiber strength were recorded according to Ali *et al.* (2008).

***Statistical analysis*:** The data were subjected to analysis of variance using software Statistix version 8.1. Genotypes mean for each character were further compared by using the Tukey test at 5% probability level (Steel *et al.,* 1997). The Factor analysis was performed using XLSTAT. From cluster analysis, the genotypes were grouped based on Euclidean distance matrix. A relative dendrogram was constructed using Ward's method to find the similarity among the genotypes in each group (Neyman and Pearson, 1928).

Table 1. **List of genotypes used in experiment from various research institutes in Punjab, Pakistan.**

Sr.	Genotypes	Institutes
1	FH 171	Cotton Research Institute, AARI, Faisalabad
2	FH159	Cotton Research Institute, AARI, Faisalabad
3	VH324	Cotton Research Station, Vehari
4	VH325	Cotton Research Station, Vehari
5	VH333	Cotton Research Station, Vehari
6	VH339	Cotton Research Station, Vehari
7	VH228	Cotton Research Station, Vehari
8	FH182	Cotton research Institute, AARI, Faisalabad
9	MNH586	Cotton Research Station, Multan
10	MG6	Private seed company
11	CIM598	Central Cotton Research institute, Multan
12	CIM602	Central Cotton Research institute, Multan
13	CIM599	Central Cotton Research institute, Multan
14	FH177	Cotton Research Station, AARI, Faisalabad
15	FH187	Cotton Research Station, AARI, Faisalabad
16	VH283	Cotton Research Station, Vehari
17	VH295	Cotton Research Station, Vehari
18	AS01	Private seed company
19	NS131	Neelum Seed Corporation, Multan
20	SB149	Private seed company
21	FH113	Cotton Research Station, AARI, Faisalabad
22	CRS2007	Cotton Research Station, Multan
23	IR3701	NIBGE, Faisalabad
24	C26	Private seed company
25	VH282	Cotton Research Station, Vehari
26	VH329	Cotton Research Station, Vehari
27	FH175	Cotton Research Station, AARI, Faisalabad
28	IR901	NIBGE, Faisalabad
29	IR3	NIBGE, Faisalabad
30	NIAB820	NIAB, Faisalabad
31	FH114	Cotton Research Station, AARI, Faisalabad
32	NS121	Neelum Seed Corporation, Multan
33	FH118	Cotton research Station, AARI, Faisalabad
34	CRS456	Cotton Research Station, Multan
35	KZ191	Private seed company
36	FH169	Cotton Research Station, AARI, Faisalabad
37	FH172	Cotton Research Station, AARI, Faisalabad
38	FH154	Cotton Research Station, AARI, Faisalabad
39	FH142	Cotton Research Station, AARI, Faisalabad
40	FH170	Cotton Research Station, AARI, Faisalabad
41	VH148	Cotton Research Station, Vehari
42	KZ189	Private seed company
43	KZ181	Private seed company
44	VH259	Cotton Research Station, Vehari
45	AA703	Private seed company
46	AA802	Private seed company
47	MNH888	Cotton Research Station, Multan
48	FH4243	Cotton Research Station, AARI, Faisalabad
49	FH941	Cotton Research Station, AARI, Faisalabad
50	MNH886	Cotton Research Station, Multan

RESULTS AND DISCUSSION

The statistical parameters were determined in 50 *Bt*-cotton accessions for yield and quality related traits. The results demonstrated that seed cotton yield has maximum variance followed by plant height and bolls per plant, suggesting maximum variation for these traits among cotton genotypes (Table 2).

Table 2. **Estimation of basic statistics for seed cotton yield and fiber quality characters in 50 genotypes of *G. hirsutum*.**

Variables	Min	Max	Mean	SD	CV%	Variance
PH	108.50	173.18	140.72	15.16	10.77	229.97
NM	0.90	6.30	3.02	1.18	39.16	1.40
NS	11.70	27.00	18.65	3.32	17.83	11.07
NB	7.90	48.40	20.30	7.43	36.57	55.31
BW	1.96	4.69	2.63	0.51	19.47	0.26
SD/BL	3.97	25.50	9.89	3.83	38.72	14.69
SCY	17.60	162.87	51.95	22.87	44.02	523.27
SI	4.39	6.62	5.59	0.56	10.18	0.32
LI	2.39	7.22	4.12	0.90	21.88	0.81
GOT%	30.23	56.61	42.08	5.11	12.15	26.15
FL	24.95	31.10	27.37	1.30	4.75	1.69
FF	3.70	5.40	4.55	0.39	8.71	0.15
UR	44.30	51.75	47.80	1.90	3.97	3.62
FS	22.50	30.65	26.62	1.64	6.19	2.72

PH= Plant height (cm), NM= Number of monopodial branches, NS= Number of sympodial branches, NB= Bolls per plant, BW=Boll weight (g), SD/BL= Seed per boll, SCY=Seed cotton yield (g), LI=Lint index, SI= Seed index, GOT% = Ginning out turn, FL = Fiber length (mm), FF = Fiber fineness (µg/inch), UR=Uniformity ratio, FS = Fiber strength (g/tex)

Several researchers reported similar findings for all these attributes (Baloch *et al.*, 2014). Similarly, the number of bolls contributed affirmatively to the yield along with plant height. These results revealed that a great amount of genetic variability was present between genotypes for studied characters and selection pressure could be performed on these traits to separate desirable genotypes. The presence of variation is important for selection and improvement of seed cotton yield for cotton breeders (Abdullah *et al.*, 2016). Boll weight, seed weight, lint index, fiber fineness, fiber strength and uniformity ratio showed minimum variance which indicates that these traits have less amount of variation among genotypes. These results suggested that there was a need to create a wider spectrum of variability for efficient selection and improvement in these attributes. Breeders are aided in their selection processes by heritability estimation. Though Johnson *et al.* (1955) noted that heritability estimates alone do not provide a clear indication of predicted growth in the following generation. Rather, they only do so when combined with genetic advancement. Genetic progress provides the magnitude of projected genetic gain achieved by one cycle of selection (Idahosa *et al.*, 2010). Coefficient of variation provided a good estimate for comparing extent of variation in available germplasm. Seed cotton yield showed a maximum coefficient of variation (44%) followed by monopodial branches, seed per boll and number of bolls. The lowest values of CV% were found for uniformity ratio, fiber length and fiber strength (Table 2).

***Analysis of variance*:** Analysis of variance for each parameter indicated that Mean squares for the genotypes under study were all significant in each parameter using analysis of variance, indicating differences across all 50 genotypes for each character (Table 3). The cotton genotypes exhibited significant differences in various characteristics. Plant height, number of sympodial branches, number of bolls, seed per boll, and seed cotton yield had highly significant differences at (P<0.01). Additionally, the number of monopodial branches, boll weight, seed index, lint index, GOT%, fiber length, fiber fineness, fiber strength, and uniformity ratio showed significant differences at (P<0.05) (Table 3). Mean performance of all genotypes showed that the highest yield was recorded for CRS-456 (162.3 g) followed by 82.46 g for FH-187. The genotype VH-282 produced lowest yield (17.0 g) hence showing a decrease of about 80.41 from top yielding variety CRS-456.

Table 3. Mean squares for seed cotton yield and fiber quality related traits in 50 genotypes of *G. hirsutum.*

SOV	Replication	Genotype	Error
Degree of Freedom	1.00	49.00	49.00
Plant height (cm)	26.48	459.90**	83.77
Monopodial branches	1.29	2.80*	1.29
Sympodial branches	2.13	22.14**	4.80
Bolls per plant	113.97	110.62*	47.72
Boll weight (g)	0.45	0.52*	0.19
Seeds per boll	7.12	29.38**	2.98
Seed cotton yield (g)	367.80	1044.87**	276.39
Seed index	0.12	0.64*	0.23
Lint index	2.51	1.62*	0.72
Ginning out turn (%)	41.10	48.57*	26.16
Fiber length (mm)	5.06	3.38*	1.47
Fiber fineness (µg/inch)	9.06	0.31*	0.19
Fiber strength (g/tex)	2.19	7.24*	4.16
Uniformity ratio%	5.90	5.44*	2.99

** Highly Significant = (P<0.001); *Significant = (P<0.005)

***Factor analysis (FA)*:** To classify important components that contribute to total variation, factor analysis was conducted. Factor 1 and Factor 2 were important as they contributed maximum towards the total variance. Hence, the contribution of each character as compared to other characters was achieved (Soomro *et al.*, 2017). Our results revealed that mean data of 14 traits in 50 Bt varieties was arranged and subjected in factor analyses to study genetic diversity for various yield and quality related attributes. The factor loadings and contribution of each character to total variance was estimated (Table 4). Out of 14 factors, the first five had Eigen values greater than 1 and these five factors contributed maximum to total diversity. F1 and F2 exhibited 24.72% and 16.56% variance respectively and were extracted for a two-dimensional representation of cotton genotypes. While F3 to F5 exhibited 12.797%, 9.682% and 8.36% variability respectively. These 5 FAs contributed 72.135 % of total variability but remaining nine FA showed >1 eigen values therefore could not be considered for further explanation (Table 4). Factor 1 contributing maximum (24.72%) towards total variability (72.13%) with positive factor loadings of yield attributing traits include number of bolls, seed cotton yield and fiber length. So, this factor could be designated as the yield factor. Factor2 contributed 16.56 % to variability with positive values for number of sympodial branches, lint index and GOT %. Shakeel *et al*. (2015) concluded similar results that F1 and F2 showed higher positive effects for seed cotton yield per plant, number of bolls, lint index and GOT % respectively. However, F3 and F4 had higher positive effects for monopodial branches, boll weight and negative effect for fiber length, respectively.

Table 4. Eigen values and factor analysis for yield and fiber traits in 50 genotypes of *G. hirsutum.*

Variables	Eigenvalue	Variability (%)	Cumulative %
F1	3.46	24.7	24.7
F2	2.31	16.5	41.2
F3	1.79	12.7	54.0
F4	1.35	9.68	63.7
F5	1.17	8.36	72.1
F6	0.94	6.72	78.8
F7	0.9	6.42	85.2
F8	0.69	4.93	90.2
F9	0.62	4.48	94.7
F10	0.37	2.69	97.4
F11	0.28	2.01	99.4
F12	0.04	0.34	99.7
F13	0.02	0.16	99.9
F14	0.009	0.06	100

Eigen vectors indicated the direction of the relationship between factors and variables. Factor 1 had high diversity of genotypes with positive values for seed cotton yield per plant, bolls number, seed /boll but plant height and GOT% while fiber length showed negative effects. Factor 2 with higher positive values related to GOT%, lint index, plant height and sympodial branches whereas monopodial branches, boll weight and seed cotton yield showed negative values The highest positive value in Factor 3 were observed for fiber strength and fiber uniformity and plant height, number of sympodial branches, bolls per plant showed negative values. Similarly, Factor 4 showed the highest positive values for monopodial branches, plant height, whereas seed index, lint index, fiber length and fineness exhibited negative effects. Whereas Factor 5 indicated maximum positive values for plant height, fiber length and fiber strength and negative values for monopodial branches, lint index and GOT%. These traits provide strong evidence of differences as characteristics of variability among genotypes suggesting that selection for the traits with higher positive values are effective (Table 5).

Table 5. Eigen vectors for fourteen characters in 50 genotypes of *G. hirsutum*.

Variables	F1	F2	F3	F4	F5
PH	-0.081	0.352	-0.154	0.320	0.363
NM	0.018	-0.108	0.239	0.582	-0.189
NS	0.174	0.283	-0.420	-0.080	0.335
NB	0.507	0.099	-0.049	0.123	0.050
BW	0.109	-0.357	0.134	-0.230	0.184
SD/BL	0.501	0.112	0.006	0.154	0.029
SCY	0.513	-0.039	0.042	0.091	0.118
SI	0.292	0.052	0.051	-0.370	-0.396
LI	-0.021	0.548	0.276	-0.143	-0.231
GOT%	-0.149	0.554	0.251	0.025	-0.070
FL	-0.036	0.049	0.276	-0.474	0.492
FF	0.205	0.105	-0.179	-0.253	-0.308
UR%	0.091	-0.094	0.418	0.022	-0.119
FS	0.143	0.007	0.544	0.068	0.322

PH= Plant height (cm), NM= Number of monopodial branches, NS= Number of sympodial branches, NB= Bolls per plant, BW=Boll weight (g), SD/BL= Seed per boll, SCY=Seed cotton yield (g), LI=Lint index, SI= Seed index, GOT% = Ginning out turn, FL = Fiber length (mm), FF = Fiber fineness (µg/inch), UR=Uniformity ratio, FS = Fiber strength (g/tex)

***Scatter plot*:** The scatter plot showed that genotypes CIM-599, IR3, FH-177, MNH-888, CRS-2007, NS-131, SB-149, AS01, MG-6, MNH-886, FH-175, VH-259, VH-295, FH-169 and VH-228 were not much diverse from each other due to narrow genetic base as they were crowded to the same area (Figure 1). While genotypes CRS-456, NIAB-820, NS-121, VH-329, VH-148, FH-114, CIM-598, VH-282 covered maximum distance from origin with respect to F1 and F2 and existed at peak of graph and were found more diverse as compared to other genotypes (Figure 1). Scatter plot analysis based on the first two factors confirmed diverse genotypes as the pattern of scattering demonstrates. The distribution of all genotypes in scatter plot discovered presence of broad genetic variation. Genotypes which were presented at the edge of plot due to greater distance from the origin are considered more diverse as compared to other genotypes. Likewise, the great distance between the base of graph and traits discussed diverse nature of studied genotypes and provided significant information about the genotypic performance (Saeed *et al.*, 2013).

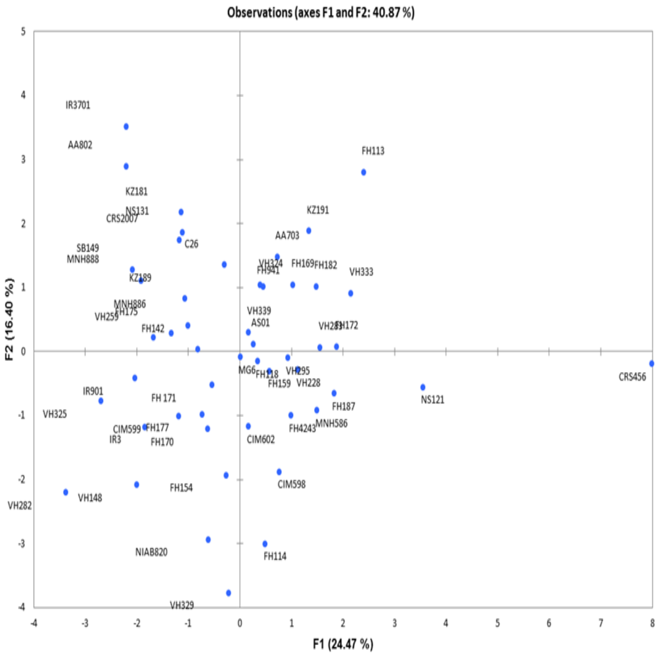

Figure 1. Two-dimensional orientation of 50 cotton genotypes of Faxis I and II.

***Biplot analysis*:** In a biplot presentation, the arrangement of morphological traits around the vectors and the angles of the vectors allows for the selection of many variables that are related to yield. In breeding efforts to create high yielding cotton cultivars, genotype evaluation is essential for further improving cotton seed yield and its related components (McCarty *et al.*, 2005). A factor biplot analysis (Figure 2) showed that genotypes and variables were loaded on plot as vectors. The range of each variable with respect to F1 and F2 presented with contribution of these variables in the diversity of genotypes. Based on F1 and F2 together, plant height, sympodial branches, seed cotton yield, seed per bolls, bolls per plant, boll weight, ginning out turn, lint index highly showed differences in the plot, but monopodial branches, staple length, micronaire value, fiber strength and UR%, and seed index exhibited minimum differences on the basis F1 and F 2 (Figure 2, 3). Correlation through biplot graph showed positive correlation between sympodial branches, bolls, boll number, boll weight and yield because these attributes showed less than 90º angle with cotton yield. In the literature, researchers have also shown the genetic variations between cotton genotypes for sympodial and monopodial branches, bolls per plant, seed cotton output per plant, boll weight, and lint percentage (Khan *et al.*, 2017).

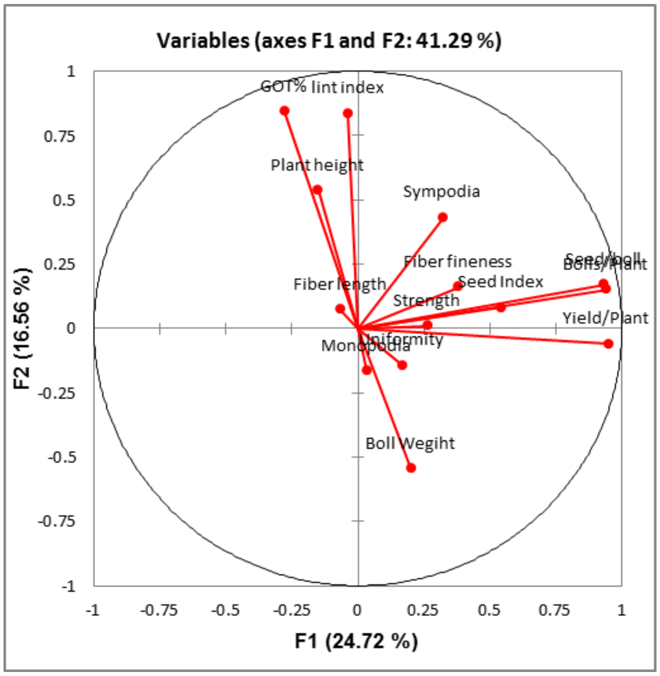

Figure 2. Factor biplot for contribution of trait.

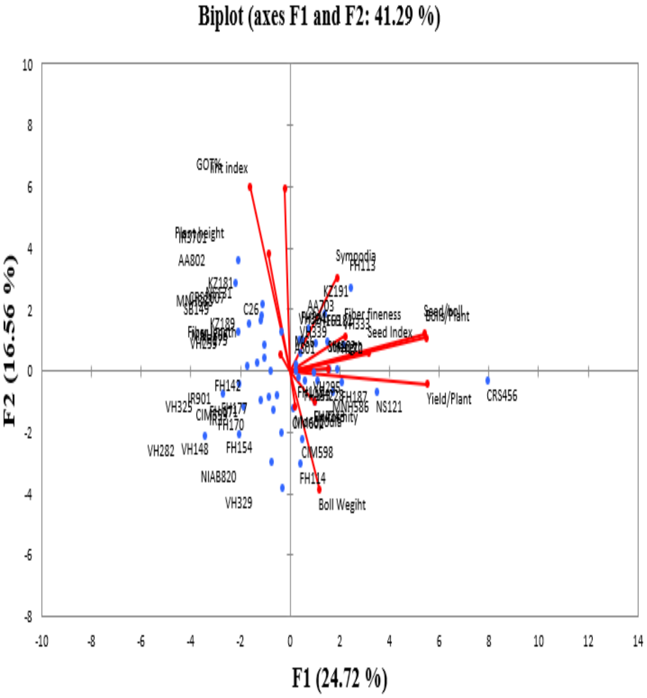

Figure 3. Two-dimensional ordination for different traits on factor biplot axis I and II.

***Cluster analysis*:** Cluster analysis is used to classify similar or dissimilar objects into several different groups based on measured variables in such a way that similar genotypes are placed in the same group. It is the simplest method to differentiate the entries using measurement that describe similarity or distance among the genotypes (Iqbal *et al.*, 2015). In the current study, the cluster analysis indicated that 50 genotypes of cotton were grouped into 3 clusters based on various morphological attributes (Table 6). Cluster 1 is composed of 29 genotypes while cluster 2 and 3 contain 21 and 1 genotypes respectively. The genotypes of cluster 1 performed best in some of characters like sympodial branches, seed cotton yield, ginning out turn and fiber length. Cluster 2 has 20 genotypes with highest mean values of lint index and ginning out turn. It showed the lowest value for these characters includes number of sympodial branches, bolls per plant, boll weight, seed per boll and seed index among all three clusters. The 3rd cluster with one genotype CRS-456 has higher mean values for plant height, seed cotton yield and number of bolls per plant. Genotype CRS-456 performed excellent for all characters under studied except for lint index and GOT%, suggesting that cluster 3 contained genotype that was more diverse for yield and quality related traits (Table 6 and 7). Similarly, the study of genetic data collections and the use of statistical tools was the main emphasis of Mohammadi and Prasannas (2003) work. They discovered that PCA and cluster analysis are most frequently used and appear to be mostly beneficial. The outcomes of cluster analysis based on factor analysis were showed in scatter plot which indicated relationship between different clusters because genotypes showing similarity or dissimilarity between them. The genotypes falling in the same cluster were presented close to each other in a scattered plot. It can be viewed through dendrogram which classified genotypes into different clusters based on their genetic variability (Chandrashekar *et al.*, 2015). Moreover, low D^2 values reflected genetic similarity while high values of D^2 indicated genetic dissimilarity among cotton genotypes. The existence of this wide variation between clusters is of great genetic value in providing genotypes aimed at cotton selection for seed cotton yield and fiber quality traits (Udaya and Patil, 2019).

Dendrogram was obtained from cluster analysis which classified 50 genotypes into three clusters (Fig. 4). Each cluster in dendrogram represented morphological characteristics. The member of cluster 3 showed the maximum intra cluster distance while cluster 2 indicating larger variability. The inter cluster distances between clusters 2 and cluster 3 was 136.077 indicated that genotypes in these cluster were highly different for the characters under studied. However, cluster 1 and cluster 3 exhibited 108.213 Euclidean distance and could be considered for heterosis breeding (Table 8). As the study conducted by Liaqat *et al.*, (2015) indicated that the proportion of contribution was largest for plant height, number of bolls, seed cotton yield and uniformity ratio toward genetic variance.

Cluster 1 and cluster 2 showed the minimum distance indicating genotypes in these cluster were less diverse for the characters. It is concluded that cluster 2 and cluster 3 genotypes can be combined to obtained high yielding genotypes with better fiber quality. Cluster 3 contained genotypes with higher seed cotton yield and number of bolls while cluster 2 with high GOT% and uniformity ratio. FA and cluster analysis gave useful information which can be utilized to obtain valuable results. Accordingly, Saeed *et al.* (2013) observed higher mean values for plant height in cluster-I and cluster-II and ginning out turn in cluster-I. Further, Farooq et al (2015) grouped 18 genotypes into three clusters which indicated that genotypes in cluster-I and cluster-III might be selected for higher yield potential.

Table 6. Cluster wise distribution of 50 upland cotton genotypes into three clusters.

Cluster	Frequency	Genotypes
1	29	FH 171, FH159, VH324, VH333, VH339, VH228, FH182, MNH586, MG6, CIM598, CIM602, FH177, FH187, VH283, VH295, AS01, VH329, NIAB820, FH114, NS121, FH118, KZ191, FH169, FH172, FH154, FH113, AA703, FH4243, FH941
2	20	FH-175, IR-901, IR-3, VH325, CIM599, FH175, IR901, IR3, FH142, FH170, VH148, KZ189, KZ181, NS131, SB149, CRS2007, IR3701, C26, VH282, VH259, AA802, MNH888, MNH886.
3	1	CRS-456

Table 7. Cluster means for fourteen traits in Bt cotton genotypes.

Clusters	1	2	3
Plant height	135.20	148.10	152.20
Monopodial branches	2.96	3.04	4.30
Sympodial branches	19.20	17.75	20.80
Bolls per plant	23.47	14.37	48.40
Boll weight	2.76	2.43	3.36
Seeds per boll	11.24	7.17	25.50
Seed cotton yield	60.68	33.76	162.87
Seed index	5.74	5.25	5.99
Lint index	3.95	4.31	3.60
Ginning out turn	40.45	44.66	37.57
Fiber length	27.64	27.05	26.05
Fiber fineness	4.54	4.52	5.30
Uniformity ratio%	47.53	48.05	51.75
Fiber strength	26.83	26.17	29.55

Table 8. Intra and inter cluster D^2 distances.

Class	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0	33.935	108.213
Cluster 2		0	136.077
Cluster 3			0

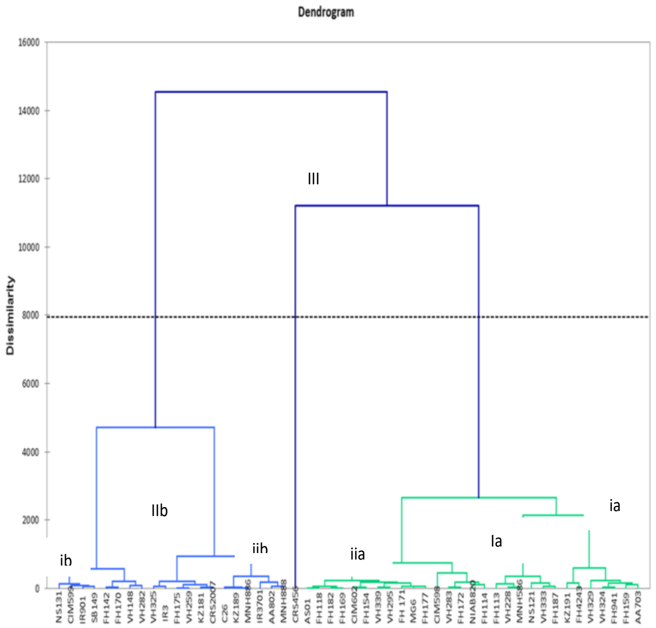

Figure 4. Classification of 50 upland cotton genotypes into three clusters.

***Conclusion*:** The variance components and factor analysis have highlighted the significant variations between cotton accessions for all yield-related parameters. The outcome of factor analysis suggested that CRS-456 (36.55%), VH-282 (6.8%), VH-235 (4.3%) and FH-113 were the greatest variability in germplasm. The members in Cluster-II and cluster-III genotypes can be crossed to obtained high yielding genotypes with better fiber quality. The germplasm used in the current study also confirmed strong broad-sense heritability and moderate to high hereditary progress. The results showed that germplasm has the potential to be used in breeding developments in the future to increase the production of seed cotton by choosing genotypes with favorable features.

***Authors' contributions*:** Hafza Anam mainly conduction the study, Amir Shakeel*, Asif Saeed, and Azeem Iqbal Khan validated the data and supervised the study, Muhammad Jabran, Shahid Iqbal, Amjad Abbas helped in analysis and writeup, Muhammad Amjad Ali supervised the study

***Funding*:** Not Applicable

***Ethical statement*:** Not Applicable

***Availability of data and material*:** Could be made available upon request

***Acknowledgement*:** The authors are thankful to different institutions for provision of germplasm for this study

Code Availability: Not Applicable.

REFERENCES

Abbas, H., A. Mahmood, Q. Ali, M. Khan, W. Nazeer, T. Aslam and W. Zahid. 2013. Genetic variability, heritability, genetic advance and correlation studies in cotton (Gossypium hirsutum L.). International Research Journal of Microbiology 4:156-161.

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

Ahmad, M.Q., S.H. Khan and F.M. Azhar. 2012. Decreasing level of genetic diversity in germplasm and cultivars of upland cotton (*Gossypium hirsutum*) in Pakistan. Journal of Agriculture and Social Sciences 8: 92-96.

Ali, M. A., S. Niaz, A. Abbas, W. Sabir and K. Jabran, 2009. Genetic diversity and assessment of drought tolerant sorghum landraces based on morph-physiological traits at different growth stages. Plant Omics Journal 2:214-227

Ali, M.A. and I.A. Khan. 2007. Assessment of genetic variation and inheritance mode of some metric traits in cotton (*Gossypium hirsutum* L.). Journal of Agriculture and Social Sciences 4:112-116.

Ali, M.A., I.A. Khan, S.I. Awan, A. Shiraz and N. Shahid. 2008. Genetics of fibre quality traits in cotton (*Gossypium hirsutum* L.). Australian Journal of Crop Sciences 2:10-17.

Arifuzzaman, M., S. Barman, S. Hayder, M. Azad, M. Turin, M. Amzad and M. Masuda. 2020. Screening of bread wheat (*Triticum aestivum* L.) genotypes under drought stress conditions using multivariate analysis. Cereal Research Communications 48:301-308.

Arshad, M. U., Zhao, Y., Hanif, O., & Fatima, F. 2022. Evolution of Overall Cotton Production and Its Determinants: Implications for Developing Countries Using Pakistan Case. Sustainability 14:840.

Baloch, M.J., N.U. Khan, M.A. Rajput, W.A. Jatoi, S. Gul, I.H. Rind and N.F. Veesar 2014. Yield related morphological measures of short duration cotton genotypes. The Journal of Animal and Plant Sciences 24:1198-1211.

Chandrashekar, R., B. Prasad, N. Vankudothu, J. Chaitanya, L. Bhavani, R. Reddy and K. Rao. 2015. A comparative genetic diversity study among elite Bt and Non-Bt cotton varieties of Andhra Pradesh by RAPD markers. Discovery Biotechnology. 6:30-37.

Economic Survey of Pakistan. 2020-21. Govt. of Pakistan Finance division, Islamabad.

Farooq, J., M. Rizwan, M. Anwar, M. Riaz, K. Mahmood and I.V. Petrescu-Mag. 2015. Multivariate analysis for CLCuD and various morphological traits in some advanced lines of cotton *Gossypium hirsutum* L. Advances in Agriculture and Botany 7:241-247.

Idahosa, D. O., J. E. Alikaand A. U. Omoregie. 2010. Genetic variability, heritability and expected genetic advance as indices for yield and yield components selection in cowpea (*Vigna unguiculata* L.). Academic Arena 2:22-26.

Iqbal, A., U. Arshad, S. Ayub, , A. Jabeen, M. Saeed. 2023. Biopriming of seed with *Trichoderma* sp. and *Bacillus* sp. leads to improved morpho-physiological parameters related to drought tolerance in maize. Integrative Plant Biotechnology 1:43-51.

Iqbal, M.A., A. Ammad and Y. Zafar. 2015. Characterization of indigenous *Gossypium arboreum* L. Genotypes for various fiber quality traits.Pakistan Journal of Botany, 47:2347-2354.

Jabran, M., U. Arshad, H. M. U. Aslam, A. Abbas, A. Haseeb, A. Hussain and M. A. Ali. 2021. Multivariate analysis of morpho-physiological and grain yield traits in advance lines of bread wheat under different leaf rust disease regimes. Pakistan Journal of Agricultural Sciences 58:

Johnson, H.W, H.F Robinson and R.E Comstock. 1955. Estimation of genetic and environmental variability in soybeans. Agronomy Journal 47:314-318.

Khan, A., Tan, D. K. Y., Afridi, M. Z., Luo, H., Tung, S. A., Ajab, M., & Fahad, S. 2017. Nitrogen fertility and abiotic stresses management in cotton crop: a review. Environmental Science and Pollution Research 24:14551-14566.

Khan, F., S. Rehman, M. Abid, W. Malik, C. Hanif, M. Bilal, Q. Qanmber, A. Latif, J. Ashraf and U. Farhan. 2015. Exploitation of germplasm for plant yield improvement in cotton (*Gossypium hirsutum* L.). Journal of Green Physiology, Genetics and Genomics 1:1-10.

Khan, M.I., H.A. Haq, K. Ullah, M. Arshad and A. Majid. 2017. Genetic Diversity and Correlation Studies for Cotton Leaf Curl Disease (CLCuD), Fiber & Yield Related Attributes in Exotic Lines of *Gossypium arboreum* L. American Journal of Plant Science 8:615-624.

Malik, T.H. and M.Z. Ahsan. 2016. Review of the cotton market in Pakistan and its future prospects. Oilseeds and fats, Crops and Lipids, 23:D606.

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

Mohammadi, S.A. and B.M. Prasanna. 2003. Analysis of genetic diversity in crop plants, salient statistical tools, and considerations. Crop Science 43:1235-1248.

Nawaz, B., M. Naeem, Malik, T. A., G. Muhae-Ud-Din, Q. Ahmad, & Sattar, S. 2019. A review about cotton leaf curl viral disease and its control strategies in Pakistan.

International Journa of Innovative and Applied Agricultural Research 3:132-147.

Neyman, J. and E.S. Pearson. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. Biometrika.pp.175-240.

Rajani VV, Davara DK, Dhaduk LK. 2019 .Management of fungal foliar diseases of cotton. Journal of Cotton Research and Development 33:110-5.

Rashid, B., Yousaf, I., Rasheed, Z., Ali, Q., Javed, F., and Husnain, T. 2016. Roadmap to sustainable cotton production. Life Science Journal 13:41-48.

Riaz, B., Saeed, A., Fiaz, S., & Riaz, A. 2019. Genetic diversity among Bt Cotton (*Gossypium hirsutum* L.) germplasm assessed through morphological and within-boll yield attributes. Journal of Animal and Plant Sciences 29:226-231.

Saeed, F., J. Farooq, A. Mahmood, M. Riaz, T. Hussain and A. Majid. 2013. Estimation of genetic variability for Cotton leaf curl virus disease, fiber quality and some morphological traits using multivariate analysis in exotic Gossypium (Diploid and tetraploid) species. Journal of. Agrobiology 30:107-115.

Sanahuja, G., Banakar, R., Twyman, R. M., Capell, T., & Christou, P. 2011. Bacillus thuringiensis: a century of research, development and commercial applications. Plant Biotechnology Journal 9:283-300.

Shahzad, K., Li, X., Qi, T., Guo, L., Tang, H., X. Zhang and WU, J. 2019. Genetic analysis of yield and fiber quality traits in upland cotton (*Gossypium hirsutum* L.) cultivated in different ecological regions of China. Journal of Cotton Research 2:1-11.

Shakeel, A., I. Talib, M. Rashid, A. Saeed, K. Ziaf and M.F. Saleem. 2015. Genetic diversity among upland cotton genotypes for quality and yield related traits. Pakistan Journal of Agricultural Sciences 52:73-77.

Soomro, A.W., R. Anjum and K. Soomro. 2017. Evaluation of Genetic Diversity for Cotton Leaf Curl Virus (Clcuv), Yield, Got% and Fiber Traits In Exotic Germplasm Of *Gossypium hirsutum* L. FUUAST Journal of Biology 7:199-205.

Steel, R.G. 1997. Pinciples and procedures of statistics a biometrical approach. 0070610282.

Udaya, V. and R.S. Patil. 2019. Genetic Studies on Parental Diversity in Relation to Yield of Promising Hybrids in Upland Cotton (Gossypium hirsutum L.). International Journal of Current Microbiology and Applied Sciences 8:513-520.