

Research Article

Genetic diversity for yield and fiber traits in upland cotton (*Gossypium hirsutum* L.) genotypes

Rajab Ali Kaleri^{1*}, Younis Malik¹, Tariq Majidano², Mohsin Zulqarnain Hussain Abbasi³, Saleem Ejaz³, Ghulam Hassan Bahalkani¹, Nabila Maryam³, Muhammad Ubaidullah Aleem², Zuhaib Saleem⁴, Asma Kaleri¹, Benazir Chutto¹, Farman Ali Kaleri¹, Kamran Khan Kaleri¹ and Irfan Ali Kaleri¹

1. Department of Plant Breeding & Genetics, Sindh Agriculture University, Tandojam, Pakistan

2. Pakistan Tobacco Board, Peshawar, Ministry of National Food Security & Research, Pakistan

3. Agriculture Research Center, Tandojam, Sindh, Pakistan

4. Department of Soil Science, Sindh Agriculture University, Tandojam, Pakistan

*Corresponding author's email: rajabkaleri@gmail.com

Citation

Rajab Ali Kaleri, Younis Malik, Tariq Majidano, Mohsin Zulqarnain Hussain Abbasi, Saleem Ejaz, Ghulam Hassan Bahalkani, Nabila Maryam, Muhammad Ubaidullah Aleem, Zuhaib Saleem, Asma Kaleri¹, Benazir Chutto, Farman Ali Kaleri, Kamran Ali Kaleri and Irfan Ali Kaleri. Genetic diversity for yield and fiber traits in upland cotton (*Gossypium hirsutum* L.) genotypes. Pure and Applied Biology. Vol. 14, Issue 2, pp524-530.

<http://dx.doi.org/10.19045/bspab.2025.140051>

Received: 09/11/2024

Revised: 06/01/2025

Accepted: 07/01/2025

Online First: 10/01/2025

Abstract

The present study was conducted at the CRI Cotton Research Institute in Tandojam to evaluate the genetic diversity for yield and fiber characteristics in upland cotton. The crop was grown using a randomized complete block design with 3 replications and 8 genotypes, including a single check variety. The cluster analysis divided the genotypes into four groups based on their genotypic homogeneity. Cluster 1 contains two genotypes (TH-25/23, TH-23/23), and there is a close relationship between them. Cluster 2 included two genotypes (TH-22/23, TH-24/23), cluster 3 contains genotypes TH-27/23 and Sindh-1, and cluster 4 contains two genotypes TH-26/23 and TH-28/23 with a maximum distance between them. Similarly, the presence of high loading was significantly observed in the first principal component PCA-4, 5, 6, and 7 exhibited positive loadings for sympodial branches plant-1, height of plant, and monopodial branches plant-1, respectively. A greater percentage of variance was observed in PCA-7 and PCA-8 (99.90% and 100%). Among the genotypes, TH-26/23 and TH-22/23 should be preferred for further breeding programs, and they could be utilized in hybridization and selection programs for the development of a new variety.

Keywords: Cotton; Germplasm; Fiber; Yield traits; Cluster analysis

Introduction

The cotton crop is a highly profitable and globally valuable cash crop due to its natural

fiber content [1]. It is widely famous for its fiber used in textile manufacture in the old and new world [2]. Cotton is a flowering

crop belonging to genus *Gossypium* which comprises fifty species and is cultivated in tropical and subtropical regions around the world. Pakistan is ranked 4th largest cotton producing country for producing edible oil and feed for livestock animals [3]. However, despite extensive research on the genetic variability of wheat but variations still exist due to environmental factors and equipment used in breeding materials [4]. Genetic diversity is major tool used in selective breeding program to conserve the germplasm for future breeding policies. Genetic diversity should be utilized to produce favorable traits. Knowledge of advanced genetics and variations is needed for selecting superior genotypes to observe the association of fiber traits and production characteristics [5]. The cluster and PCA (Principal Component Analysis) have been utilized to explore different and similar genotypes and group them into different clusters. PCA is a highly important statistical procedure for breaking down the total variation and selecting elite parental breeding lines. Genotypic variability can be explored through PCA, which is helpful for observing the physical variability of different traits under study. The information gained from PCA regarding variability is beneficial for selecting superior genotypes based on agronomic and genetic factors [6]. Considering the importance of genetic variability and variations in production and fiber characteristics of upland cotton genotypes, their relationship should be taken into account.

Materials and Methods

The present research was carried out during the Kharif season of 2022 at the Cotton Research Institute, Agriculture Research Center (ARC), Tandojam. Eight varieties of cotton were used in the experiment was laid out in a Randomized Complete Block Design (RCBD) with three replications.

3.1 Genotypes

1. TH-22/23
2. TH-23/23
3. TH-24/23
4. TH-25/23
5. TH-26/23
6. TH-27/23
7. TH-28/23
8. Sindh-1

Observation recorded

- Plant height (cm)
- Number of monopodial branches plant⁻¹
- Number of sympodial branches plant⁻¹
- Boll weight (g)
- Number of bolls plant⁻¹
- Seed cotton yield plant⁻¹
- GOT (%)
- Staple length

Statistical analyses

The analysis of variance was determined according to the method given by [7]. The means of genotypes for all characters were compared by using the least significant difference (LSD) test at 5% level of probability. The cluster analysis was determined using the squared Euclidean distance, and PCA, was carried out using multivariate statistical package (MVSP-3) [8].

Results

Cluster analysis

The genotypes were classified into four categories in cluster analysis, utilizing the mean value of the study's parameters. The total computed distance, ranging from 1 to 30, was examined using the Jaccard similarity index. The UPGMA cluster analysis algorithm was employed to construct a dendrogram based on these distances. Meanwhile, the similarity indices were calculated using the Euclidean distance method. The genotypes were grouped into four clusters based on genetic similarities and homology, with distances ranging from 1 to 30. Cluster 1 included genotypes TH-23/23 and TH-23/23, cluster 2 comprised TH-24/23 and TH-22/23, cluster 3 featured

TH-27/23 and Sindh-1, while cluster 4 exhibited the highest homogeneity, with genotypes TH-28/23 and TH-26/23 having the least significant distance. Setting the mean distance threshold at 10, the cluster groups can be consolidated into two major groups. Specifically, cluster 1, 2, and 3 form Group A, while the genotypes in cluster 4 constitute Group B (Fig. 1).

Principle component analysis

The principal component analysis method was used to examine the average data. According to principal component analysis, each differentiation axis's biggest contributor to overall variance is significant. The total eight results were rapidly utilized to select how much element is present. The sum of the total 8 values was similar to the values of variables. Different morpho-physiological characteristics were studied through PCA to find the relationship within the parameters of these. In this study, the positive and negative loadings indicate that the variables and components have both positive and negative correlation patterns. The findings regarding the PCI principle component showed positive and high loading by boll weight, height of plant, cotton seed production, monopodial branches plant-1, GOT %, and boll per plant, while the length of staple and sympodial branches plant-1 was recorded negative loadings in this study. The 2nd PCII revealed more positive for cotton seed production and bolls plant-1, while it was observed more negative for boll weight, height of plant, cotton seed production, monopodial branches plant-1, GOT %, length of staple, and boll plant-1 in (Table 1). In this study, important contributing traits for variation among 3rd PCIII were recorded positive normal for cotton seed production, boll weight, height of plant, and sympodial branches plant-1, while it was recorded negative loading for GOT %, monopodial branches plant-1, bolls plant-1,

and length of staple traits. Similarly, the results for 4th PCIV revealed high and positive loading by boll weight, height of plant, cotton seed production, monopodial branches plant-1, while the length of staple and sympodial branches plant-1, while it was observed negative for GOT percentage, height of plant, and bolls plant-1. The results regarding 5th PCA loading were observed maximum for height of plant and monopodial branches plant-1, while it was recorded negative for bolls plant-1, cotton seed production, and sympodial branches plant-1. The results for 7th PCA exhibit highest positive loading values for cotton seed production, monopodial branches plant-1, GOT percentage, and boll weight, while the negative loading was recorded for length of staple, height of plant, bolls plant-1, and sympodial branches plant-1. The findings for the 8th component exhibit the highest positive loading in length of staple, GOT percentage, and cotton seed production, while it was observed negative in bolls plant-1, height of plant, monopodial branches, boll weight, and sympodial branches per plant. The first principal component showed the highest Eigen value having 3.4678. While others showed lesser Eigen value symmetrically as shown in (Table 2). The highest variability was accumulated in PC-VIII and PC-VII, which showed 100% and 99.90% variability, and the least variability, was observed in PC-I with 43.30% among different traits of cotton accessions.

Discussion

The cotton is major source of fiber which is utilized for various purposes including making curtains, socks, cloth and towels, etc. The fiber of cotton widely used for making cloth in textile industry of the world [9]. Cotton is valuable economic crop that provides clothes and shelter all around the world [10]. It is among the natural fiber crop used in the textile industry. The experiment

was conducted at Agriculture Research Cotton (ARC) [11]. Tandojam. The experiment was conducted on genetic

diversity on yield and fiber trait in upland cotton genotypes.

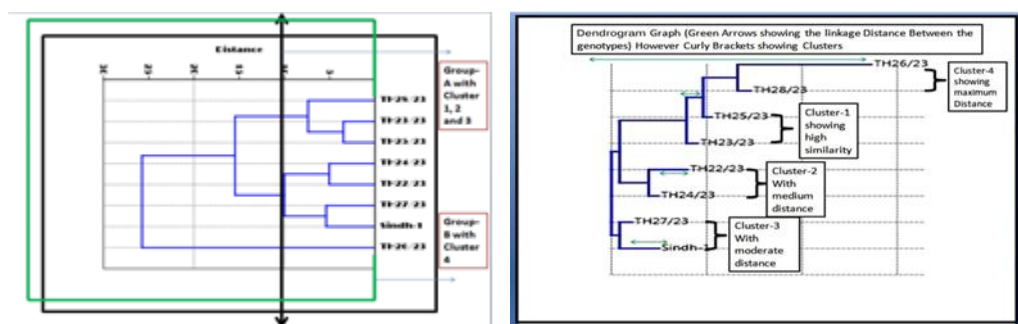


Figure 1. Dendrogram gram showing relationship among different cotton genotype

Table 1. Factor loadings for different traits in cotton genotypes

Factor	PCA-1	PCA-2	PCA-3	PCA-4	PCA-5	PCA-6	PCA-7	PCA-8
Plant height	0.465	-0.273	0.165	-0.107	0.137	0.145	-0.793	-0.036
Monopodial branches plant ⁻¹	0.320	-0.201	-0.692	0.046	0.430	-0.418	0.107	-0.065
Sympodial branches plant ⁻¹	-0.408	-0.306	0.294	0.006	-0.064	-0.774	-0.225	-0.005
Bolls plant ⁻¹	0.107	0.651	-0.228	-0.088	-0.395	-0.279	-0.296	-0.429
Boll weight	0.421	-0.160	0.401	0.447	-0.046	-0.091	0.326	-0.565
Seed cotton yield plant ⁻¹	0.460	0.255	0.150	0.289	-0.202	-0.288	0.055	0.700
GOT%	0.186	-0.506	-0.239	-0.295	-0.736	0.026	0.151	0.022
Staple length	-0.282	-0.144	-0.340	0.780	-0.228	0.188	-0.298	0.024

Table 2. Principal component (PCs) analysis for various morpho-physiological traits in cotton genotypes

	PCA-1	PCA-2	PCA-3	PCA-4	PCA-5	PCA-6	PCA-7	PCA-8
Eigen value	3.4678	1.8880	0.8910	0.8870	0.4909	0.2713	0.0935	0.0104
Percent of variance	43.30	23.60	11.10	11.10	6.10	3.40	1.20	1.00
Cum. Percentage	43.30	66.90	78.10	89.20	95.30	98.70	99.90	100

Cluster analysis

A large number of potentially associated variables can be reduced using the principle component analysis (PCA) to a smaller set of variables known as principal components. According to UPGMA these genotypes formed four separate clusters. According to their genotypic homogeneity Cluster analysis divided the cotton genotypes into 4

main groups [12]. Cluster 1 contains two genotypes (TH-25/23, TH-23/23) and there is a close relationship among them. Cluster 2 included two genotypes (TH-22/23, TH-24/23), Cluster 3 contains genotypes TH-27/23 and Sindh-1 and Cluster 4 contains two genotypes TH-26/23 and TH-28/23 and there is a maximum distance between these both genotypes same researcher like [7]. As

indicated above, further study in this area may be conducted by choosing superior genotypes, since promising genotypes have been found from this cotton experiment. According to a dendrogram recombinant inbred lines have large genetic distance between them, the genetic distance study have demonstrated that there is space for improving the genetics of cotton varieties, and the data produced may be utilized to design extensive crosses and to take advantage of genetic diversity. Increase heterosis expression. Scientists led an examination to ponder the phenotypic variety and hereditary decent variety, significant elements to choose germplasm for cotton rearing [10]. The phenotypic relationship and head part investigation demonstrated that a large portion of the qualities assessed are significant for choice of high yielding genotypes and contributing their offer for more extensive hereditary fluctuation of the genotype. In this way, the general outcome shows the nearness of enough inconstancy for advancement of improved cotton assortments and the contemplated attributes can be utilized for choice.

Factor loadings and PCA analysis

PCA was used to analyses several yield and fiber traits in order to determine the general connection between these factors. The loading plot (Fig. 1) showed how these factors were related to one another. The positive and negative associations among the components, as well as between variables, are indicated by negative and positive loadings in this study. These negative and positive genotypes shared a major part in the diversity and were considered the most vitiated clusters. The results of our study are supported by [13, 14], they reported more similar values recorded with the results for main component and major diversity about successful component analysis. The results

regarding the PC one showed positive and high loading for height of plant cotton seed production, monopodial branches per plant, bolls weight, GOT percentage and bolls per plant, while length of staple and sympodial branches per plant showed negative and high loading in this study. In our study the 3rd PC was also exhibit negative and positive loading results cotton seed production and bolls per plant, while it was recorded negative and high for sympodial branches per plant, height of plant, GOT percentage, monopodial branches per plant and boll weight. The results for present study was supported by findings of [11], who had also reported negative and high loading values for sympodial branches per plant, height of plant, GOT percentage, monopodial branches per plant and boll weight and Positive higher range loadings for cotton seed production and bolls per plant. The key role sharing traits for variation among 3rd PC was observed cotton seed production, sympodial branches per plant, height of plant and boll weight showed positive and higher values, while negative values was recorded for GOT percentage, monopodial branches per plant, length of staple and boll per plant. Another study was conducted by [10], whose results were less as compared with present findings for PCI, II and III, the variation between the loading values might to due to data size, environment factors as well as seed production and variety variation between these. In same way 4th PC was observed positive and high loading values for cotton seed production, monopodial branches per plant, boll weight, length of staple and sympodial branches per plant, whereas it was recorded negative and high loading values for GOT percentage, height of plant and bolls per plant as suggested by [8]. The findings regarding 5th PC exhibit maximum and positive loading values for monopodial branches per plant and height of plant, while it was observed negative cotton

seed production, GOT percentage, height of plant and boll weight. Such type of result were also revealed by [4, 5] and reported highest positive loading values for monopodial branches per plant and height of plant and negative values for cotton seed production, GOT percentage, height of plant and boll weight. In 7th PC showed positive and high loading for GOT percentage and height of plant and length of staple, while it was observed negative for sympodial branches per plant, boll weight monopodial branches per plant, cotton seed production and boll per plant, whereas negative results for length of staple, sympodial branches per plant and bolls per plant was recorded. The results about 8th PC was observed high and positive loading for length of staple, cotton seed production and GOT percentage, whereas it was recorded negative for monopodial branches per plant, boll weight, height of plant and bolls per plant. The findings of our investigations are supported by [2, 13], they reported similar values as observed in our study for the first principal component showed the highest Eigen value having 3.4678. while others showed lesser Eigen value symmetrically as shown in (Table 2). The highest variability was accumulated in PC-VIII and PC-VII which showed 100% and 99.90% variability and the least variability was observed in PC-I with 43.30% among different traits of cotton accessions. Each principal component's Eigen value and variance dropped gradually until they reached and stopped at 0.104.

Conclusion

The cluster analysis was divided into four groups based on their genotypic homogeneity. Cluster 1 consists of two genotypes (TH-25/23, TH-23/23) that show a close relationship. Cluster 2 includes two genotypes (TH-22/23, TH-24/23), Cluster 3 contains genotypes TH-27/23 and Sindh-1, and Cluster 4 consists of two genotypes TH-26/23 and TH-28/23 with a maximum

distance between them. Similarly, high loading was significantly observed in the first principal component (PC-4, 5, 6, and 7) with positive loadings for sympodial branches per plant, plant height, and monopodial branches per plant, respectively. A higher percentage of variance was observed in PCA-7 and PCA-8 (99.90% and 100%).

Authors' contributions

Conceived and designed the experiments: RA Kaleri, Y Malik & T Majidano, Performed the experiments: RA Kaleri, MZH Abbasi, S Ejaz GH Bahalkani, MU Aleem & N Maryam, Analyzed the data: Z Saleem, KK Kaleri, IF Kaleri & B Chutto, Contributed materials/ analysis/ tools: A Kaleri & Farman Kaleri, Wrote the paper: KK Kaleri & IF Kaleri.

References

1. Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ & Zhu Y (2020). Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Gene* 52(23): 516–524.
2. Akhtar T, Islam MA, Rasul MG, Kundu S, Khaleq UZ & Ahmed JU (2019). Evaluation of genetic diversity in short duration cotton (*Gossypium hirsutum* L.). *J Cot Res* 2(12): 11-12.
3. Shakeel A, Talib I, Rashid M, Saeed A, Ziaf K & Saleem F (2015). Genetic diversity among upland cotton genotypes for quality and yield related traits. *Pak J Agric Sci* 52(1): 73-77.
4. Morris TT, Davies NM, Dorling D, Richmond RC & Smith GD (2018). Testing the validity of value-added measures of educational progress with genetic data. *Bri Edu Res J* 44(5): 725-747.
5. Baloch AW, Sahito JA, Ali M, Baloch GA, Abro A, Channa SA, Baloch AM,

- Baloch GH & Baloch GM (2014). Association analysis of yield and fiber traits in advance Pakistani cotton cultivars (*Gossypium hirsutum* L). *Adv Appl Agric Sci* 2(12): 73-80.
6. Saragih RIK & Coworker D (2019). Studies Keraaman Galur F4 Hasil Perisilangan Padi Verites IPB 4S Dengan Situ Patenggang. *Buletin Agrohorti*, 7(1), 38-39.
 7. Gomez, M. I. (1984). Statistical extremes and applications. *Inter J for Theore and Appl Statis* 33(1): 100-120.
 8. Kovach WL (1999). Multivariate statistical package, Version 3.1 users' Munual. Kovach Computing Services, London.
 9. Stewart A & Rossi M (2010). Association and heritability analysis for yield and fiber traits in promising genotypes of cotton (*Gossypium hirsutum* L.). *Sindh Uni Res J* 47(2): 303-306.
 10. Yehia WMB & Essam FH (2021). Correlation and multivariate analysis across non segregation and segregation generations in two cotton crosses. *Egy J of Agric Res* 99(3): 354-364.
 11. Shah SAI, Khan SJ, Ullahand K & Sayal OU (2018). Genetic diversity in cotton germplasm using multivariate analysis. *Sarhad J of Agric* 34(1): 130-13.
 12. Timpson NJ, Greenwood CM, Soranzo N, Lawson DJ & Richards JB (2018). Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Gene* 19(2): 110-124.
 13. Ullah A, Shakeel A, Ahmed HGMD, Naeem M, Ali M, Shah AN & Hasan ME (2022). Genetic basis and principal component analysis in cotton (*Gossypium hirsutum* L.) grown under water deficit condition. *Fron in Plant Sci* 13: 981-994.
 14. Khan NU, Hassan G, Kumbhar MB, Kang S, Khan I, Parveen A, Aiman U & Saeed M (2013). Heterosis and inbreeding depression and mean performance in segregation generation in upland cotton. *Eur J Sci Res* 1(17): 531-546.