***Original Research Article***

**Multivariate Analysis for Genetic Diversity in *Bt*-Cotton (*Gossypium hirsutum*) Genotypes for Yield and Fibre Quality Traits**

**ABSTRACT**

**Aim:** The primary objective of the present investigation was to evaluate the genetic diversity among twenty-four *Gossypium hirsutum* genotypes for yield and fibre quality traits using multivariate statistical approaches. This evaluation aims to identify genetically divergent parental lines suitable for breeding programs focused on yield improvement and fibre quality enhancement.

**Design**: A field experiment was conducted in *Kharif* 2022–23 using a randomized block design in three number of replications with the spacing of 67.5 x 30 cm.

**Methodology:** Twenty-four cotton genotypes were analyzed using multivariate techniques specifically, cluster analysis (based on Ward’s method) and principal component analysis (PCA using Kaiser’s criterion) to quantify genetic diversity and trait contribution to variability. Eight yield and fibre quality attributes were recorded. Data analysis was executed using R software (version R4.2.1).

**Results:** Cluster analysis grouped the genotypes into three distinct clusters, with Cluster 2 having the highest number of genotypes (12), followed by Cluster 1 (11) and Cluster 3 comprising a single genotype (HAU Bt-5). PCA revealed that the first two principal components accounted for 71.08% of the total phenotypic variance. PC1, contributing 57.70%, was predominantly influenced by boll number, seed cotton yield, ginning out turn and micronaire. The trait biplot and clustering patterns underscored significant diversity among genotypes, highlighting potential combinations for hybridization.

**Conclusion:** Significant genetic variability was observed among genotypes. Traits such as seed cotton yield, boll number, boll weight and ginning out turn were key contributors to diversity. The multivariate approach effectively identified superior parental lines for use in cotton breeding programs.

**Keywords:** *Cotton,**Cluster, Diversity, fibre quality, Seed Cotton Yield*

1. **INTRODUCTION**

Cotton (*Gossypium spp.*), often termed as "white gold," is an economically significant fibre crop that underpins both global and national economies (Nadeem *et al*., 2024). Among the cultivated species, upland cotton (*Gossypium hirsutum* L.), an allotetraploid, is distinguished by its exceptional adaptability and superior yield potential. This species dominates global cotton production, accounting for approximately 95% of the total fibre output (Zhang *et al*., 2008). Cotton is cultivated across an estimated 314.7 lakh hectares worldwide, with India leading in the acreage, encompassing approximately 113.60 lakh hectares. Notably, India contributes 36.09% of the global cotton-growing area and accounts for approx 20% (5.09 MT) of the total global cotton production, which stands at 25.55 MT (<https://cotcorp.org.in/>).

In recent decades, cotton breeding efforts have predominantly concentrated on enhancing yield. However, with evolving consumer preferences and advancements in textile technology, there is an increasing demand for superior fibre quality (Zafar *et al*., 2023). Consequently, fibre quality improvement has emerged as a key objective in modern cotton breeding programs. Critical fibre attributes, including fibre strength (FS), length (FL) micronaire (FM), and uniformity (FU), serve as primary determinants of fibre quality, significantly influencing the global cotton market. These traits are vital across various stages of textile processing, including spinning, weaving and dyeing. However, the genetic complexity and strong environmental dependence of these fibre quality parameters present substantial challenges in breeding programs aimed at fibre quality improvement (Rehman *et al*., 2015).

The increasing emphasis on fibre quality assessments and a deeper understanding of fibre properties' influence on textile performance have driven cotton breeders to prioritize fibre quality in genetic improvement programs. It is imperative to evaluate the genetic diversity within existing cultivars to develop cotton genotypes with superior fibre quality alongside high seed cotton yield. Biometrical techniques such as principal component analysis (PCA) and cluster analysis have proven invaluable in assessing genetic diversity among cotton genotypes (Saeed *et al*., 2014). PCA effectively quantifies phenotypic variability, providing a robust framework for evaluating genetic diversity based on fibre attributes (Sun *et al*., 2019). While PCA assigns germplasm lines to distinct groups, cluster analysis offers a more nuanced perspective on genetic relationships and kinship among cotton accessions (Munir *et al*., 2020). Recognizing the critical role of genetic diversity, this study aims to elucidate the genetic architecture of key fibre traits in *Gossypium hirsutum* which helps to form a foundation for the development of improved cultivars.

1. **MATERIALS AND METHODS**

Twenty-four lines of *Gossypium hirsutum* were screened for genetic diversity based on eight yield and quality ascribing traits. The field experiment was carried out in *Kharif* 2022-23 with three replicated trials in a randomized block design (RBD) with the spacing of 67.5 x 30 cm at the research area of the Cotton section, Department of Genetics and Plant Breeding, CCS Haryana Agricultural University Hisar. Hisar is located at 2905’5’’ north latitude and 75045’55’’ east longitudes. Climate of Hisar owes to its continental location on the outer margin of south west monsoon region. It has semi-arid and sub-tropical climate with warm and dry winds during summer season, hot humid in monsoon season and cold dry weather in winter. Data were recorded from each entry of three replications for boll weight, boll number, seed cotton yield, ginning out turn (GOT), upper half mean length (UHML), uniformity index, tenacity and micronaire. The mean data were put through cluster analysis and principal component analysis using R software version R4.2.1 (R core team, 2022) based on Ward’s minimum variance method (1963) and Kaiser’s (1958) approach of using principal components, respectively.

1. **RESULTS AND DISCUSSION**
	1. **Cluster analysis:** D2 cluster analysis is the most commonly used clustering method that groups the plant material into different clusters based on D2 values and construct a line diagram known as cluster diagram. In the present study, the total 24 lines are grouped into three major clusters on the basis of D2 values. The cluster pattern revealed that highest number of genotypes was present in Cluster 2 (12) followed by Cluster 1 (11) and only one genotype (HAU Bt-5) was observed in cluster 3 (Table 1). Similar findings were also observed by Reddy *et al*. (2025) and Arif *et al*. (2024). A dendrogram was constructed using Ward’s minimum variance method that displays division of all 24 genotypes into different clusters (Fig.1). A heatmap was also constructed depicting the similarity index among various 24 genotypes. As the colour of intersecting cells changes from dark blue to light blue to light orange to dark orange, the dissimilarity between these genotypes increases. Therefore, the blue colour diagonal represents the exactly similar lines whereas the orange intersections represent highly dissimilar lines (Fig. 2). These dissimilar genotypes can be crossed in different fashions to obtain segregating population.

**Table 1**. Clustering of 24 genotypes into three cluster based on D2 statistics

|  |  |  |
| --- | --- | --- |
| **CLUSTER** | **NO. OF GENOTYPES** | **GENOTYPES** |
| Cluster. 1 | 11 | HAU Bt-12, HAU Bt-13, HAU Bt-14, HAU Bt-8, HAU Bt-9, HAU Bt-6, HAU Bt-7, HAU Bt-10, HAU Bt-11, HAU Bt-15, HAU Bt-19 |
| Cluster. 2 | 12 | HAU Bt-4, PAU Bt-3, H1098i, HAU Bt-1, HAU Bt-16, HAU Bt-17, HAU Bt-18, HAU Bt-20, HAU Bt-21, HAU Bt-22, HAU Bt-23, PAU BT-3, |
| Cluster. 3 | 1 | HAU Bt-5 |

Inter-cluster distances indicate the genetic dissimilarity between two clusters, while intra-cluster distances reflect the genetic variation within individual clusters. Table 2 presents the magnitude of the intra- and inter-cluster distances obtained from the statistical analysis of 24 cotton genotypes. The highest intra-cluster distance was observed in cluster 1 (1296.32) followed by cluster 2 (1214.04). The maximum inter-cluster distance was recorded between clusters 1 and 2 (2016.58) followed by the distance between clusters 2 and 3 (1707.94). The smallest inter-cluster distance was observed between clusters 1 and cluster 3 (1387.68) indicating minimal divergence between these clusters. The results are in agreement with the findings of Luqman *et al*. (2025) and Kaleri *et al*. (2025).

The cluster means for different eight morphological and quality traits in 24 cotton genotypes is presented in Table 3. Cluster 3 exhibits the highest mean values for SCY (3621.40 g), BW (3.73 g), BN (45.78), GOT (36.90) and mic (4.90). Cluster 2 outperforms in terms of UHML (27.93), UI (82.00) and Tenacity (28.46). Genotypes from clusters with complementary performance across different traits can serve as potential parents for transgressive breeding. Similar conclusions were depicted by Zafar *et al*. (2024) and Bilal *et al*. (2024).



Fig 1. Circular dendrogram of all 24 genotypes obtained by Ward’s minimum variance method



Fig 2. Heatmap depicting similarity index among all 24 genotypes of cotton

**Table 2.** Inter and intra (diagonal) cluster distance among all three different clusters

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cluster. 1** | **Cluster. 2** | **Cluster. 3** |
| **Cluster. 1** | **1296.32** |  |  |
| **Cluster. 2** | 2016.58 | **1214.04** |  |
| **Cluster. 3** | 1387.68 | 1707.94 | **0.00** |

**Table 3.** Mean performance of all clusters for various traits under study

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SCY** | **BW** | **BN** | **GOT** | **UHML** | **UI** | **Tenacity** | **mic** |
| **Cluster. 1** | 3331.46 | 3.62 | 41.83 | 36.56 | 26.42 | 81.18 | 27.57 | 4.59 |
| **Cluster.2** | 2451.99 | 3.41 | 33.83 | 34.23 | 27.93 | 82.00 | 28.46 | 4.36 |
| **Cluster.3** | 3621.40 | 3.73 | 45.78 | 36.90 | 23.20 | 81.00 | 23.20 | 4.90 |

*SCY: Seed cotton yield, BW: Boll weight, BN: Boll number, GOT: Ginning Out Turn, UHML: Upper Half Mean length, UI: Uniformity Index, Mic: Micronaire (FF)*



Fig. 3. Scree plot presenting contribution of each principal component



Fig. 4. Biplot of fibre quality traits along with the evaluated genotypes based on PCA analysis



Fig. 5. Biplot depicting four distinct groups of traits based on PCA analysis

**Table 4.** Eigen values, eigen vectors and proportion of variation accounted for by first two principal components.

|  |  |  |
| --- | --- | --- |
| **Trait** | **PCI** | **PC2** |
| **SCY** | 0.867 | 0.262 |
| **BW** | 0.435 | -0.647 |
| **BN** | 0.885 | 0.300 |
| **GOT** | 0.788 | 0.188 |
| **UHML** | -0.89 | 0.092 |
| **UI** | -0.827 | -0.297 |
| **Tenacity** | -0.621 | 0.505 |
| **mic** | 0.641 | -0.326 |
| **eigenvalue** | 4.616 | 1.071 |
| **Variance** | 57.697 | 13.389 |
| **cumulative** | 57.697 | 71.086 |

*SCY: Seed cotton yield, BW: Boll weight, BN: Boll number, GOT: Ginning Out Turn, UHML: Upper Half Mean length, UI: Uniformity Index, Mic: Micronaire (FF)*

**3.2 Principal component analysis (PCA):** Principal component analysis (PCA) was conducted to gain a deeper understanding of the sources of variance among cotton genotypes. PCA reduces the number of traits influencing the total variance. Out of the eight principal components (PCs) generated, only the first two are discussed, as their eigen values were greater than one and explained 71.08% of the total variation, as depicted in fig. 3. Similar results were observed by Reddy *et al*. (2025) and Liu & Kim (2025). PC1 accounted for 57.7% of the total variation and had highest positive factor loadings for boll number (0.885), seed cotton yield (0.867), ginning out turn (0.788), mic (0.641) and boll weight (0.435) and negatively influenced by UHML (-0.890), UI (-0.827) and tenacity (-0.621). Whereas, PC2 explained 13.39% of the variation, with tenacity, boll number, SCY and ginning out turn being the highest contributors, showing factor loadings of 0.505, 0.300, 0.262 and 0.188, respectively, while boll weight, mic and UHML were the negative contributor with a factor loading value of -0.647, -0.327 and -0.297, respectively (Table 4), as also observed by Chapara *et al*. (2024) and Agarwal *et al*. (2024).

Furthermore, four distinct groups of traits were identified based on the trait biplot shown in fig. 4. The first group included BN, SCY and GOT which exhibited a positive association with the first two PCs. The second group consisted of mic and BW were negatively correlated with PC2, while third group included only UI, which was negatively correlated with PC1 and positively correlated with PC2. The fourth group included tenacity and UHML which was positively correlated with PC3, while negatively correlated with PC1 and PC2. The vectors of traits *viz*., BN, SCY and GOT made an acute angle with the PC1 vector, indicating their positive correlation, while UHML and tenacity made an obtuse angle with it showing their negative correlation, as also concluded by Agarwal *et al.* (2024). The association is stronger in the genotypes, included in that group *i.e*., HAUBt-6, HAUBt-14, HAUBt-10, HAUBt-12, HAUBt-13, HAUBt-11 as shown in fig. 5.

1. **CONCLUSION**

The present study has revealed substantial degrees of variation among the twenty-four cotton lines. The principal component analysis identified traits *viz*., seed cotton yield, boll weight, boll number and ginning out turn based on which selection would be efficacious. Aforesaid characters have high positive factor loadings and statistically valid differences that contributed more to genetic diversity and were crucial in illustrating the clusters. The grouping of genotypes into three clusters and reduction of traits to two major principal components by PCA would be of practical value to cotton breeders to design experiments for further germplasm collection as well as for hybridization activities to produce heterotic progenies.

**REFERENCES**

Agarwal, N., Ray, S., & Tripathi, K. C. (2024). Enhancing cotton crop yield prediction through principal component analysis and regression modelling. *Springer Nature*, 2(3), 237-248.

Arif, T., Chaudhary, M. T., Majeed, S., Rana, I. A., Ali, Z., Elansary, H. O., & Azhar, M. T. (2023). Exploitation of various physio-morphological and biochemical traits for the identification of drought tolerant genotypes in cotton. *BMC Plant Biology*, *23*(1), 508-512.

Bilal, N., Tipu, A. L., Liaqat, S., Sarwar, M., & Basheer, M. (2024). Diversity and correlation studies of intraspecific hybrids for their morphological traits in cotton (Gossypium hirsutum L.). *BMC Plant Biology*, *13*(5), 516-30.

Chapara, R., Reddy, K. S., Rani, M. S., Sudhamani, K., Grace, A. D., Lakshmi, N. V., & Rao, G. P. (2024). Studies on genetic divergence through principal component analysis and mahalanobis D2 analysis for seed cotton yield and important fibre traits in (*G. hirsutum* L.) *Pure and Applied Biology, 12* (3), 624-636.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187 200.

Kaleri, R. A., Malik, Y., Majidano, T., Abbasi, M. Z. H., Ejaz, S., Bahalkani, G. H., & Kaleri, I. A. (2025). Genetic diversity for yield and fibre traits in upland cotton (*Gossypium hirsutum* L.) genotypes*. Pure and Applied Biology, 14* (2), 524-530.

Liu, Y. & Kim, H. J. (2024). Attenuated total reflection fourier transform infrared spectroscopy with soft independent modelling of class analogy–principal component analysis for classifying cotton fibre maturity phenotypes of cotton population composed of various genotypes. *Applied Spectroscopy*, *78*(1), 99-110.

Luqman, T., Hussain, M., & Khan, M.K.R. (2025). Harnessing multivariate insights coupled with susceptibility indices to reveal morpho-physiological and biochemical traits in heat tolerance of cotton. *BMC Plant Biology*, *25*(1), 126-137. doi: 10.1186/s12870-025-06141-5.

Munir, S., Qureshi, M.K., Shahzad, A.N., Nawaz, I., Anjam, M.S., Rasul. S., & Zulfiqar, M.A. (2020). Genetic dissection of interspecific and intraspecific hybrids of cotton for morpho yield and fiber traits using multivariate analysis. *Pakistan Journal of Agriculture Research*, *33*(1), 9-16.

Nadeem, S., Riaz Ahmed, S., Luqman, T., Tan, D. K., Maryum, Z., Akhtar, K. P., & Liu, Y. (2024). A comprehensive review on Gossypium hirsutum resistance against cotton leaf curl virus. *Frontiers in Genetics*, *15*(2), 130-145.

R Core Team. (2022). R: a language and environment for sta tistical computing. R foundation for statistical computing, Vienna.

Reddy, P., Vishnuvardhan, K.M., & Amarnath K. (2025). Insights into the genetic divergence in asiatic cotton (*Gossypium arboreum* L.) germplasm for fibre-quality traits. *Plant Genetic Resources: Characterization and Utilization*, *23*(1), 18-26.

Rehman, S., Abid, M. A., Bilal, M., Asharaf, J., Liaqat, S., Ahmad, R.I., & Qambar, G. (2015). Genotype by trait analysis and estimates of heritability of wheat (Triticum aestivum L.) under drought and control conditions. *Basic Research Journal of Agricultural Science Review*, *4*(4), 127–134.

Saeed, F., Farooq, J., Mahmood, A., Hussain, T., Riaz, M., & Ahmad, S. (2014). Genetic diversity in upland cotton for cotton leaf curl virus disease, earliness and fibre quality. *Pakistan Journal of Agriculture Research*, *27*(3), 226 236.

Sun, Z.W., Xingfen, L.I.U., Zhengwen, G.U., Qishen, Z., Yan, L.I., Zhikun, K.E., Huifeng, Y., Jun, W.U., Jinhua, W.U., Liqiang, Z., Guiyin, M.A., & Zhiying. (2019). Evaluation of the genetic diversity of fibre quality traits in upland cotton (Gossypium hirsutum L.) inferred from phenotypic variations. *Journal of Cotton Research*, *2*(22), 1-8.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244.

Zafar, M. M., Ijaz, A., Anwar, Z., Iqbal, M. S., Zafar, S., Subhan, M., & Jiang, X. (2024). Harnessing genetic diversity in cotton for enhanced resilience against salt stress by using agro-physiological characters. *Plant Production Science*, *28*(1), 51–68.

Zafar, S., Afzal, H., Ijaz, A., Mahmood, A., Ayub, A., Nayab, A., & Moosa, A. (2023). Cotton and drought stress: An updated overview for improving stress tolerance. *South African Journal of Botany*, *161*(3), 258-268.

Zhang, H.B., Li, Y., & Wang, B., & Chee, P.W. (2008). Recent advances in cotton genomics. *International Journal of Plant Genomics*, *7*(3), 234-343.