Vol. 05, No. 02
e-ISSN. 2830-7097

EXPLORING GENETIC DIVERSITY AND TRAIT RELATIONSHIPS FOR COTTON IMPROVEMENT THROUGH MULTIVARIATE ANALYSIS

Muhammad Noman Manzoor[1], Muhammad Saad Rafique[1], Arslan Arshad[1], Ali Haider[1], Inshara Eman[1], Hafiz Usama Manzoor[1], Muhammad Abdullah[1], Jahanzaib[1], Qasim Ram[1], Umar Bashir[1] *, Hafiz Rafi U deen[2]

[1] Department of Plant Breeding and Genetics, The Islamia University of Bahawalpur, 63100 Pakistan.
[2]. Department of agronomy Bahauddin Zakariya University, Multan

Email: Umarbashir.phd@gmail.com

Received: 31-08-2026 Accepted: 30-09-2026 Published: 25-10-2026

ABSTRACT

Genetic diversity and relationships among agronomic, physiological, fiber, and yield traits were described in 16 cotton (*Gossypium hirsutum L.*) genotypes evaluated in a randomized complete block design (RCBD) with three replications at the Research Farm of The Islamia University of Bahawalpur, Pakistan. Traits included plant height, number of nodes, monopodial and sympodial branches, green and open bolls, net photosynthetic rate, fiber length, ginning out-turn (GOT), and seed cotton yield. The supplied correlation matrix showed positive associations of yield with GOT (r = 0.492), open bolls (r = 0.457), sympodial branches (r = 0.354), and nodes (r = 0.330), while yield was nearly uncorrelated with net photosynthetic rate (r = 0.008) and negatively correlated with fiber length (r = −0.261) and green bolls (r = −0.163). The reported first three principal components explained 81.968% of total variation. Cluster summaries indicated higher mean yield in Cluster 1 (4.936) than in Cluster 2 (4.606), whereas Cluster 2 had slightly higher mean photosynthetic rate and fiber length. These results suggest useful diversity for parent selection, but statistical outputs and some reported significance tests require confirmation from the original plot-level data before definitive inference.

Keywords: *cotton; genetic diversity; correlation; principal component analysis; cluster analysis; seed cotton yield*

INTRODUCTION

Upland cotton (*Gossypium hirsutum L.*) is a major source of natural textile fiber and an economically important crop in Pakistan (Shahzad *et al*., 2022). Cotton productivity is affected by crop management, pests, heat and water stress, and the genetic potential of cultivated varieties. Breeding progress depends on identifying useful variation in yield components and fiber traits while maintaining acceptable quality. Because agronomic characters are interrelated, evaluating them together can help breeders summarize variation and identify contrasting parents(Topuz *et al*., 2026)(Shamkuwar *et al*., 2025). Correlation analysis describes the direction and strength of pairwise trait associations. Principal component analysis (PCA) reduces a multi-trait dataset to a smaller set of components, while cluster analysis groups genotypes according to overall similarity. These methods have been used to characterize cotton germplasm and support selection of genetically divergent parents (Kumar *et al*., 2025). This study summarizes variation and trait associations among 16 cotton genotypes using the supplied ANOVA, correlation, PCA, and cluster summaries.

MATERIALS AND METHODS

Experimental material and design

The experiment was conducted at the Research Farm of The Islamia University of Bahawalpur, Punjab, Pakistan. Sixteen (16) genotypes were evaluated in a randomized complete block design with three replications. Genotypes were Amp-1, SI-21, Emp-3, SI-3, FMP-4, EMP-5, Emp-6, SI_7, HB-1, HB-2, HB-HSP-1, SI-9, SS-10, Emp-11, SI-12, and Emp-12. Seeds were dibbled at three seeds per hill and thinned at the seedling stage to retain one healthy plant. Standard agronomic and plant-protection practices were applied.

Traits and statistical analyses

The recorded traits were plant height (PH), number of nodes (NODES), monopodial branches (MNB), sympodial branches (SYMB), number of green bolls (NGB), number of open bolls (NOOB), net photosynthetic rate (Pn), fiber length (FL), ginning out-turn (GOT), and seed cotton yield. The supplied analyses included ANOVA, Pearson correlation, PCA, and hierarchical clustering. The manuscript reports Euclidean distance and Ward's method for clustering. Exact measurement protocols, season, plot dimensions, units for all traits, and software versions were not supplied and should be added from field records.

Statistical reporting note

The manuscript reports 16 genotypes and three replications, for which an RCBD ANOVA normally has 15 genotype degrees of freedom, 2 replication degrees of freedom, and 30 error degrees of freedom (total 47). The source ANOVA table instead lists genotype df = 14 and error df = 28These analysis statistically analysed using statistical tools (Software) XLSTAT, RStudio, and Statistix 8.1 ("Genetic Variability for Yield Related Traits in Some F1 Crosses of Upland Cotton," 2025).

RESULTS AND DISCUSSION

Analysis of variance (ANOVA)

The source manuscript reported genotype mean squares of 635.77 for plant height, 110.71 for nodes, 0.60 for monopodial branches, 126.35 for sympodial branches, 11.40 for green bolls, 224.15 for open bolls, 3.92 for Pn, 2.64 for GOT, 0.65 for fiber length, and 4.75 for yield (Ali *et al.*, 2026). The source marked plant height, monopodial branches, sympodial branches, open bolls, and yield as significant, but several printed significance labels are inconsistent with the printed mean squares and error mean squares. In addition, the reported yield error mean square (0.00088) is unusually small relative to the genotype mean square (4.75) (Nascimento *et al.*, 2024). These values should not be treated as validated inferential results until recalculated from the original data.

Table 1. Mean squares and significance labels as supplied in the manuscript; these require verification. Source degrees of freedom were inconsistent with the stated number of genotypes.

Trait	Replication MS	Genotype MS (source)	Error MS (source)	F-Value
Plant height	918.81	635.77*	302.76	2.10
Nodes per plant	120.33	110.71NS	42.75	2.59
Monopodial branches	1.27	0.60*	0.27	2.22
Sympodial branches	263.25	126.35**	43.58	2.90
Green bolls	15.27	11.40NS	10.00	1.14
Open bolls	301.18	224.15*	83.45	2.69
Net photosynthetic rate	0.11	3.92NS	2.96	1.32
GOT	0.81	2.64NS	3.79	0.70
Fiber length	1.57	0.65NS	0.38	1.71
Seed cotton yield	0.00013	4.75**	0.00088	5397.73

Note: *=for significant values and double, **= for highly significant values, Ns = shows for non-significant values

Correlation among traits

Yield had its strongest positive pairwise association with GOT (r = 0.492), followed by open bolls (r = 0.457), sympodial branches (r = 0.354), and nodes (r = 0.330) (Subhashini *et al.*, 2023). The relationship with monopodial branches was weak (r = 0.093), and the correlation with Pn was essentially zero (r = 0.008). Yield was negatively associated with green bolls (r = −0.163) and fiber length (r = −0.261) (Polat *et al.*, 2026). These are correlations, not evidence of

causal effects; significance tests and confidence intervals were not supplied. Plant height was positively correlated with nodes (r = 0.870), sympodial

branches (r = 0.886), and open bolls (r = 0.781). Nodes were strongly associated with sympodial branches (r = 0.986) and open bolls (r = 0.943), while sympodial branches and open bolls were also strongly correlated (r = 0.953). Such high correlations suggest substantial redundancy among architectural and boll traits, which should be considered when selecting traits for multivariate models. Fiber length showed negative correlations with green bolls (r = −0.539), monopodial branches (r = −0.496), and GOT (r = −0.386) (Hussain *et al*., 2026). Any interpretation of a yield–fiber trade-off should remain cautious and be validated in additional environments (D & V, 2026) (Kanta *et al*., 2025).

Table 2**.** Pearson correlation matrix for the ten cotton traits (coefficients reproduced from the supplied.

Trait	PH	NODES	MNB	SYMB	NGB	NOOB	Pn	FL	GOT	Yield
PH										
NODES	0.870									
MNB	0.600	0.587								
SYMB	0.886	0.986	0.601							
NGB	0.565	0.493	0.612	0.529						
NOOB	0.781	0.943	0.617	0.953	0.519					
Pn	-0.319	-0.257	0.220	-0.264	0.076	-0.193				
FL	-0.385	-0.268	-0.496	-0.276	-0.539	-0.244	-0.292			
GOT	0.303	0.320	0.353	0.307	0.227	0.351	0.188	-0.386		
Yield	0.183	0.330	0.093	0.354	-0.163	0.457	0.008	-0.261	0.492	

Cluster analysis

The reported hierarchical clustering separated the genotypes into two groups. Cluster 1 comprised Amp-1, SI-21, Emp-3, SI-3, Emp-5, SI-7, HB-2, SI-9, and SS-10; Cluster 2 comprised FMP-4, Emp-6, HB-1, HB-HSP-1, Emp-11, SI-12, and Emp-12. Cluster 1 had higher reported means for plant height, nodes, monopodial and sympodial branches, green and open bolls, GOT, and seed cotton yield (Prakash Reddy *et al*., 2024)(Tamboli *et al*., 2025). Cluster 2 had slightly higher mean Pn and fiber length. The reported mean yield was 4.936 for Cluster 1 and 4.606 for Cluster 2; the yield unit should be confirmed from field records (Muhammad Younas *et al*., 2025) (Navingautham *et al*., 2026)

Table 3. Mean trait values for the two clusters. No standard errors or within-cluster variance estimates were supplied.

Trait	Cluster 1	Cluster 2
PH	133.481	109.095
NODES	38.815	28.143
MNB	0.630	0.143
SYMB	34.259	23.095
NGB	8.259	6.333
NOOB	32.444	18.429
Pn	26.811	27.643
FL	28.181	28.405
GOT	37.370	37.095
Yield	4.936	4.606

The cluster means indicate a potential breeding contrast: Cluster 1 is more promising for several reported yield components, whereas Cluster 2 may contain useful variation for fiber length and photosynthetic rate. These conclusions are descriptive because cluster membership and means have not been independently regenerated from plot-level date.

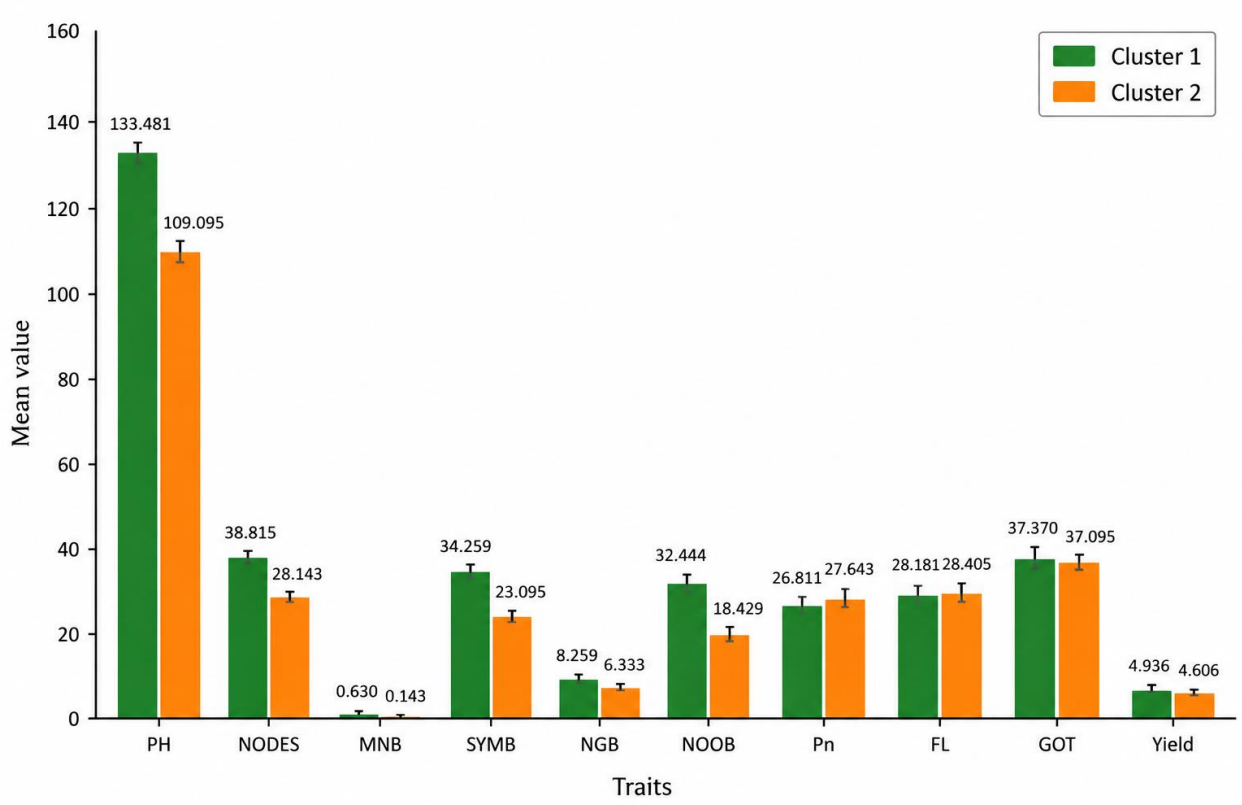


Figure 1. Comparison of mean values for traits, and Green and orange bars represent Cluster 1 and Cluster 2, respectively. Error bars are illustrative because standard errors were not provided.

Principal componentanalysis

The source reports that PC1, PC2, and PC3 had eigenvalues of 5.077, 1.729, and 1.391, explaining 50.773%, 17.289%, and 13.906% of total variation, respectively. Their cumulative contribution was 81.968%. PC1 was reported as the dominant component, followed by PC2 and PC3 (Yadav *et al*., 2025) (Celestin *et al*., 2025). The eigenvalues and variance percentages are reproduced below; PCA loadings and genotype scores were not available for independent verification and are not inferred here (Elhaik, 2022)(Beattie & Esmonde-White, 2021) .

Table 4. Eigenvalues and variance explained by principal components. Values should be checked against the original PCA output.

Component	Eigenvalue	Variance (%)	Cumulative (%)
PC1	5.077	50.773	50.773
PC2	1.729	17.289	68.062
PC3	1.391	13.906	81.968
PC4	0.624	6.241	88.209
PC5	0.506	5.057	93.265
PC6	0.315	3.154	96.420
PC7	0.238	2.377	98.797
PC8	0.092	0.923	99.720
PC9	0.021	0.205	99.925
PC10	0.008	0.075	100.000

The PCA results suggest that a small number of components summarize much of the variation. However, claims about which traits load most strongly on each component, or which individual genotypes occupy particular positions on a biplot, require the original loading and score matrices (Fuchedzieva & Georgiev, 2025)(Poonia *et al*.,

2025). Those claims are therefore not presented as confirmed findings in this concise version.

CONCLUSION

The supplied results indicate variation among the 16 cotton genotypes and identify positive yield associations with GOT, open bolls, sympodial branches, and nodes. Cluster 1 had higher reported mean seed cotton yield and several yield-related traits, while Cluster 2 had slightly higher reported fiber length and net photosynthetic rate. The reported first three principal components explained 81.968% of total variation. These patterns may help guide parent selection, but the manuscript is not statistically final: the ANOVA degrees of freedom conflict with the stated design, some mean-square/significance combinations are inconsistent, and the PCA outputs need verification. Reanalysis using the original plot-level data is essential before submission or making definitive breeding recommendations.

REFERENCES

Ali, B, Yousaf, MI, Jameel, S, Shaheen, M, Akram, M, Ammar, A, Bukhari, SJ, Sheikh, AA, Zubair, M, Hussain, S. 2026, 'Agronomic Performance and Yield Stability of Late Sown Cotton Genotypes under Semi-Arid Conditions of Bahawalpur', *Biological and Clinical Sciences Research Journal*, vol *7*, no 03, hh 103–107.

Beattie, JR, Esmonde-White, FWL. 2021, 'Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra'. In *Applied Spectroscopy*, vol. 75 no 04, hh 361–375

Celestin, M, Vasuki, M, Sujatha, S, Dinesh Kumar, A. 2025, 'Principal Component Analysis for Simplifying Multivariate Financial Data in Portfolio Risk Analysis', *In Brainae Journal of Business*, vol 9, no 02, hh 171–179.

D, R, V, K. 2026, 'Multi-environment correlation and trait association analysis for seed cotton yield and its component traits in cotton (Gossypium hirsutum L.)' *International Journal of Advanced Biochemistry Research*, vol 10, no 8, hh 131–136.

Elhaik, E. 2022, 'Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, vol 12, no 1.

Fuchedzieva, I, Georgiev, P. 2025, 'Principal Component Analysis of the environmental impact of ships visiting the Port of Varna', *IOP Conference Series: Earth and Environmental Science*, vol 1532, no 1.

Genetic Variability for Yield Related Traits in Some F1 Crosses of Upland Cotton. (2025). *Trends in Biotechnology and Plant Sciences*. https://doi.org/10.62460/tbps/2025.078

Hussain, Y, Abro, A, Soomro, MK, Abro, HH, Kaleri, A, Soomro, TA, Memon, MH, Rai, G, Dayanand, R, Unar, B, Kaleri, I, Khaskeli, AA, Syed, S, Kaleri, SA. 2026, 'Correlation and Regression Coefficient Studies on Yield and Fiber Quality Traits in Upland Cotton (Gossypium hirsutum L.) Genotypes' *Jammu Kashmir Journal of Agriculture*, vol 6, no 2, hh 149–154.

Kanta, CMM, Chamundeswari, N, Reddy, KVS, Sreekanth, B. 2025, 'Study of Correlation and Path Coefficient Analysis for Seed Cotton Yield, Yield Attributing and Fibre Quality Traits in Upland Cotton (Gossypium hirsutum L.)', *Journal of Advances in Biology & Biotechnology*, vol 28, no 10, hh 515–526.

Kumar, S, Jattan, M, Kumar, D, Sharma, A, Malik, KS, Saini, AK, Mandhania, S. 2025, 'Genetic Diversity in Bt-Cotton (Gossypium hirsutum L) Genotypes for Yield and Fibre Quality Traits using Multivariate Analysis', *Journal of Advances in Biology & Biotechnology*, vol 28, no 6, hh 1081–1088.

Muhammad Y, Muhammad Z, Babar HB, Muhammad J, Muhammad KM, Muhammad MI, Ali AA, Sobia I, Muhammad T, Abdul GK, Musarrat S, Syed WHS, Yousaf, MI, Sabir H. 2025, 'Multivariate Analysis of Elite *Bt* Cotton Genotypes for Seed Cotton Yield and Fiber Quality Traits Under Semi-Arid Conditions' *Integrative Plant Biotechnology*, vol 3, no 2, hh 101–111.

Nascimento, JS, Batista, LG, Barbosa, WF, Belo, LP, Silva, JF, Costa, GHS, Emiliano, PC. 2024, 'Data transformations for ANOVA assumption violations in agrarian sciences subfields: a systematic review, simulation study, and practical guidelines'. *REVISTA DELOS*, vol 17 no 61, e2782.

Navingautham, NS, Vishnuvardhan, KM, Sabitha, N, Naidu, GM, Reddisekhar, M, 2026, 'Multivariate analysis for earliness, yield and fibre quality traits in cotton using correlation, PCA and cluster analysis', *Electronic Journal of Plant Breeding*, vol 17, no 2, hh 188–197.

Polat, F, Çınar, VM, Ünay, A. 2026, 'Correlation, Stepwise and Path Analysis in Cotton (Gossypium hirsutum L.)', *Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi*, vol 29, no 1, hh 71–78.

Poonia, MK, Kumar, A, Rana, M, Kumar, R. 2025, 'Genetic Diversity Assessment of Wheat Genotypes in Bundelkhand Region Using Principal Component Analysis', *Asian Journal of Soil Science and Plant Nutrition*, vol 11, no 3, hh 241–246.

Prakash RBVR, Vishnuvardhan, KM, Sagar, UN, Kalyani, DL, Ramakrishna, MS, Amarnath K, Venkateswarlu, NC. 2024, 'Unravelling the Complex Interactions among Yield Attributes in Upland Cotton (Gossypium hirsutum L.) through Principal Component Analysis', *Ecology, Environment and Conservation*, vol 30, hh S1–S5.

Shahzad, K, Mubeen, I, Zhang, M, Zhang, X, Wu, J, Xing, C. 2022, Progress and perspective on cotton breeding in Pakistan. In *Journal of Cotton Research',* vol. 5, no 1.

Shamkuwar, SG, Kayande, NV, Uike, AR, Ganvir, MM, Shamkuwar, GR. 2025, 'Comparative Analysis of Yield and Fibre Traits in Newly Developed Bt and Non-Bt Cotton Hybrids', *Journal of Advances in Biology & Biotechnology*, vol 28, no 1, hh 346–353.

Subhashini, S, Rajeswari, S, Premalatha, N, Kalaimagal, T, Muthuswami, M, Jeyakumar, P, Keerthivarman, K. 2023, 'Association and path analysis studies for yield contributing and fibre quality traits in the F1 population of Gossypium hirsutum L', *Environment Conservation Journal*, vol 24, no 2,hh 70–76.

Tamboli, N, Patel, DH, Chandramaniya, CP, Patel, VB, Savaliya, D, Suthar, MC, Mukeshkumar, P, Kale, B, Modha, K. 2025, 'Multivariate analysis of genetic variability and divergence in upland cotton (Gossypium hirsutum L.) for yield and fiber quality traits', *Plant Science Today*, vol 12, no 4, hh 1–9.

Topuz, T, Çinar, VM, Ünay, A. 2026, 'Inheritance of Yield and Fiber Quality Traits and Inbreeding Depression in F3, F4 and F5 Generations of Cotton (Gossypium hirsutum L.)', *Turkish Journal of Field Crops*, vol 31, no 1, hh 1–9.

Yadav, D, Singh, L, Jafri, SKF, Singh, A, Singh, V. 2025, 'Study on genetic variability and principal component analysis in Indian mustard [Brassica juncea (L.) Czern and Coss]', *International Journal of Advanced Biochemistry Research*, vol 9, no 5S, hh 07–10.