Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 3, 2026

GENETIC VARIABILITY, TRAIT ASSOCIATIONS, AND MULTIVARIATE ANALYSIS FOR YIELD AND FIBER QUALITY IMPROVEMENT IN UPLAND COTTON (GOSSYPIUM HIRSUTUM L.) GENOTYPES

Syed Muhammad Salman[1], Inshara Eman[2], Qasim Ram[3], Alishba Saeed[4], Eman Ashfaq[5], Muhammad Noman Manzoor[*6], Hafiz Usama Manzoor[7], Muhammad Abdullah[8]

*[1,2,3,4,5,*6,7]Plant Breeding and genetics The Islamia University of Bahawalpur*
[8]Department of Plant Breeding and Genetic The Islamia University of Bahawalpur,Pakistan

[1]syedmuhammadsalman131@gmail.com, [2]insharae@gmail.com, [3]qasimrampbg.iub@gmail.com
[4]saeedalishba19@gmail.com, [5]emanashfaq3610@gmail.com, [*6]muhammadnomanmanzoor8@gmail.com
[7]sahabusama712@gmail.com, [8]abdullah73db@gmail.com,

DOI: https://doi.org/10.5281/zenodo.23432140

Keywords
upland cotton, genetic variability, yield, nodes, genotypic correlation, phenotypic correlation, path coefficient, PCA, RCBD, LSD

Article History
Received: 07 January 2026
Accepted: 17 March 2026
Published: 31 March 2026

Copyright @Author
Corresponding Author: *
Muhammad Noman

Abstract

A multi-trait evaluation framework was applied to 25 coded upland cotton (Gossypium hirsutum L.) varieties using observations recorded for three replications. Eleven traits were analyzed: plant height (PH), nodes, monopodial branches (MNP), sympodial branches (SYMP), green bolls (GB), open bolls (OB), yield, Pn, ginning outturn (GOT), fiber length, and fiber uniformity. A randomized complete block design (RCBD) model was used to demonstrate analysis of variance (ANOVA), Fisher's least significant difference (LSD), descriptive statistics and coefficients of variation (CV), phenotypic and genotypic correlation, path analysis, and principal component analysis (PCA). In this study, variety effects were significant at $P < 0.001$ for all 11 traits. Mean yield was 3,272.13 (yield/ha), ranging from 1,936.82 to 5,555.50; V15 had the highest mean yield (5,317.53). Yield showed a weak positive association with nodes (phenotypic $r = 0.191$; genotypic $r = 0.188$), but stronger associations with GOT ($rP = 0.283$; $rG = 0.304$) and PH ($rP = 0.280$; $rG = 0.285$). A five-predictor genotypic path model explained 17.7% of variation among genotype means. PCA indicated that PC1 and PC2 explained 36.63% and 19.56% of standardized variation, respectively. These outputs illustrate an integrated breeding-analysis workflow for genetic potential or recommending varieties.

INTRODUCTION

Cotton (Gossypium hirsutum L.) is an important commercial crop cultivated for its natural fiber and oil-rich seeds. It contributes to the agricultural economy and supports textile industries, employment, and rural livelihoods in many cotton-producing countries. Improving cotton productivity while maintaining desirable fiber quality remains a major objective of cotton breeding programs (Shahzad et al., 2022).

Seed cotton yield is a complex quantitative trait influenced by genetic constitution, environmental conditions, and interactions among yield-contributing characters. Plant height, number of nodes, monopodial and sympodial branches, boll production, ginning outturn, and fiber characteristics contribute to the overall performance of cotton genotypes. Understanding the variability and relationships among these traits is essential for identifying suitable selection

criteria (SAHAR et al., 2021).Genetic variability provides the fundamental basis for crop improvement because selection is more effective when useful differences exist among genotypes. Previous studies in cotton have reported associations between seed cotton yield and important yield components, particularly boll number and sympodial branche(Chapara et al., 2025; REHMAN et al., 2020; SAHAR et al., 2021) also reported positive associations of yield with plant height, sympodial branches, and boll number. However, the strength and direction of trait relationships may vary across genetic material and growing environments.

Correlation analysis helps breeders identify traits associated with yield and fiber quality. Phenotypic correlation describes the overall association expressed among observed traits, whereas genotypic correlation provides an estimate of the underlying genetic relationship after accounting for environmental variation. Nevertheless, correlation alone cannot distinguish direct contributions from indirect associations among interrelated traits (Chapepa et al., 2020; SAHAR et al., 2021). Path coefficient analysis can therefore provide additional insight by partitioning trait correlations into direct and indirect effects on yield (Chapepa et al., 2020).Multivariate techniques offer complementary tools for evaluating several traits simultaneously. Principal component analysis (PCA) reduces a large set of correlated variables into fewer components that explain major patterns of variation among genotypes. A PCA biplot further illustrates relationships among traits and the relative positions of genotypes, helping breeders examine phenotypic similarities and identify entries with potentially desirable combinations of characteristics (Zafar et al., 2022)refer.

Considering the importance of these approaches, the present study evaluated 25 upland cotton varieties for morphological, yield-related, and fiber-quality traits. The objectives were to quantify trait variability, compare varietal performance through analysis of variance and least significant difference testing, estimate phenotypic and genotypic correlations, investigate direct and indirect trait associations through path analysis, and characterize multivariate variation using PCA and biplot analysis. The overall aim was to identify useful selection criteria and promising cotton genetic material for future breeding efforts directed toward improved yield and fiber quality.

Materials and Methods

The field experiment was conducted at the experimental field of the Department of Plant Breeding and Genetics, The Islamia University of Bahawalpur (IUB), Bahawalpur, Punjab, Pakistan. This study involved 25 upland cotton (Gossypium hirsutum L.) varieties (V1–V25) evaluated using a randomized complete block design (RCBD) with three replications. Eleven traits were considered: plant height (PH), number of nodes per plant (Nodes), monopodial branches (MNP), sympodial branches (SYMP), green bolls (GB), open bolls (OB), yield, Pn, ginning outturn (GOT), fiber length (F Length), and fiber uniformity (Fiber U). Data were analyzed using analysis of variance (ANOVA), and variety means were compared using Fisher's least significant difference (LSD) test at the 5% probability level(SAHAR et al., 2021). Descriptive statistics, including mean, standard deviation, range, and coefficient of variation (CV), were calculated for each trait. Phenotypic and genotypic correlation coefficients were estimated to examine associations among traits, while path coefficient analysis was performed to assess the direct and indirect relationships of selected traits with yield. Principal component analysis (PCA) was conducted using standardized variety means for all 11 traits. Eigenvalues, explained variance, principal component scores, and trait loadings were calculated, and a biplot was constructed to visualize relationships among varieties and traits.

Table 1:

Sr. No	Variety Name	Mean Yield	Sr. No	Variety Name	Mean Yield
1	IUB-13	2677.66	14	Line-19	3140.00
2	Line-07	2822.36	15	COT-25	5317.53
3	COT-18	3670.65	16	IUB-06	2149.97
4	IUB-09	3125.70	17	Cotton-17	3960.46
5	Cotton-22	2599.72	18	Line-23	4330.68
6	FH-Line-06	3392.88	19	IUB-31	3880.38
7	IUB-21	3173.45	20	COT-10	3247.30
8	Line-04	2769.17	21	Line-16	3304.14
9	COT-15	3159.15	22	IUB-24	3335.31
10	IUB-27	2152.59	23	Cotton-03	3015.54
11	Line-11	3333.57	24	FH-Line-12	3293.60
12	Cotton-08	3863.52	25	IUB-28	2800.45
13	IUB-BT-14	3287.49			

Results

Descriptive statistics and coefficients of variation

The overall mean yield was 3,272.13, with a range of 1,936.82 to 5,555.50. The descriptive CV for yield across plot records was 20.81%, while the residual CV from the RCBD model was 6.03%. Nodes averaged 32.64 and had a descriptive CV of 19.37%; their residual CV was 5.29% (Frey et al., 2024). The highest residual CV values occurred for MNP (77.32%) and GB (38.47%), partly because these count traits included many low or zero values. Fiber length and fiber uniformity had comparatively low residual CV values.

Table 2. Descriptive statistics, coefficients of variation, and Fisher's LSD thresholds.

Trait	Mean	SD	Min	Max	Overall CV (%)	Residual CV (%)	LSD 5%
PH	133.39	18.38	94.00	170.00	13.78	2.71	5.94
Nodes	32.64	6.32	17.00	43.00	19.37	5.29	2.83
MNP	0.52	0.94	0.00	4.00	179.85	77.32	0.66
SYMP	28.76	6.74	12.00	42.00	23.45	6.90	3.26
GB	3.24	2.48	0.00	8.00	76.42	38.47	2.05
OB	25.99	8.71	8.00	46.00	33.52	13.07	5.58
Yield	3272.13	680.97	1936.82	5555.50	20.81	6.03	324.17
Pn	28.09	1.68	24.00	31.70	5.99	2.58	1.19
GOT	38.33	1.68	33.70	42.40	4.39	1.66	1.04
F Length	28.49	0.60	27.30	29.70	2.09	0.94	0.44
Fiber U	85.32	1.08	83.00	88.20	1.26	0.60	0.85

Analysis of variance and LSD comparisons

In this experiment, variety effects were significant at $P < 0.001$ for all 11 traits. The strongest F-ratios were recorded for PH ($F = 77.48$), Nodes ($F = 39.10$), Yield ($F = 34.64$), and SYMP ($F = 33.31$), indicating that variety means differed relative to the simulated residual error for these traits (REHMAN et al., 2020)

Table 3. RCBD ANOVA summary, error mean squares, residual CV, and LSD values.

Trait	Mean	MSE	F variety	P-value	LSD 5%	Residual CV (%)
PH	133.39	13.099	77.48	<0.001	5.942	2.71
Nodes	32.64	2.979	39.10	<0.001	2.834	5.29
MNP	0.52	0.162	14.62	<0.001	0.660	77.32
SYMP	28.76	3.941	33.31	<0.001	3.259	6.90
GB	3.24	1.553	10.08	<0.001	2.046	38.47
OB	25.99	11.537	18.20	<0.001	5.576	13.07
Yield	3272.13	38990.474	34.64	<0.001	324.166	6.03
Pn	28.09	0.525	14.62	<0.001	1.189	2.58
GOT	38.33	0.403	19.66	<0.001	1.042	1.66
F Length	28.49	0.072	13.32	<0.001	0.439	0.94
Fiber U	85.32	0.266	11.41	<0.001	0.846	0.60

The LSD yield was 324.17 units. V15 had the highest mean yield (5,317.53), which exceeded the next-highest variety V18 (4,330.68) by 986.85 units, larger than the LSD threshold. V18 exceeded V17 (3,960.46) by 370.22 units, also above the nominal LSD. V19 (3,880.38) and V12 (3,863.52) differed by only 16.86 units, below the LSD threshold, and were not separated by this pairwise comparison.

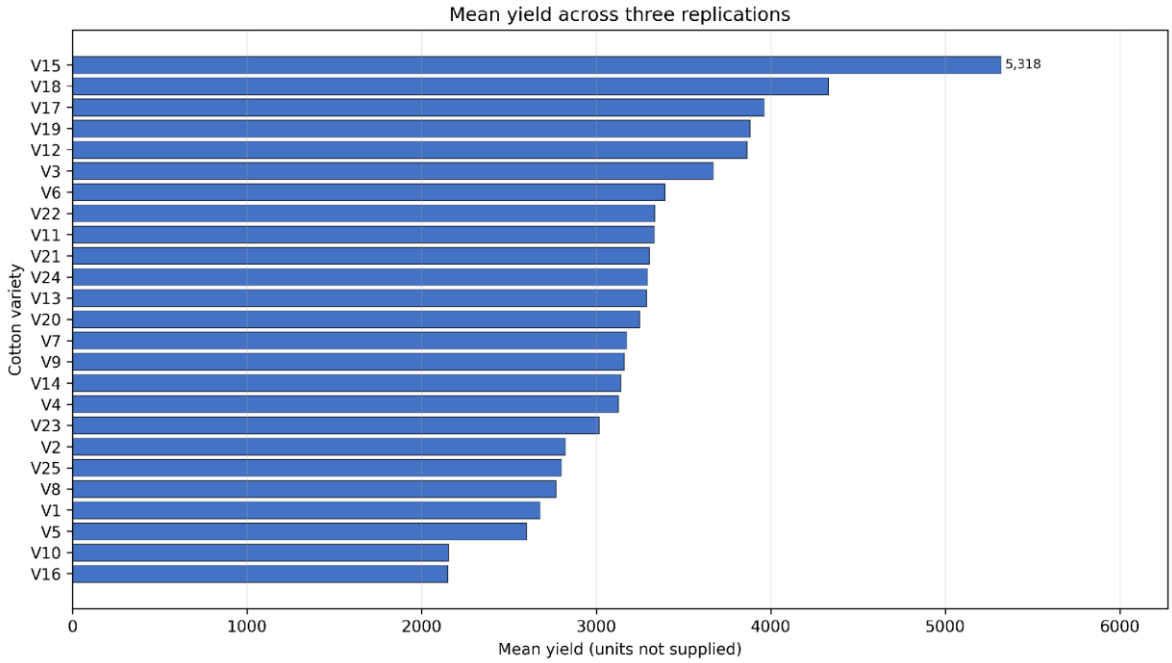

Figure 1. Variety mean yield across three replications; yield units were not supplied.

Phenotypic and genotypic correlations

Plant architecture and boll traits showed strong positive associations. Phenotypic correlations were high for Nodes–SYMP (r = 0.954), SYMP–OB (r = 0.952), and Nodes–OB (r = 0.940). Similar relationships occurred in the genotypic correlation matrix (Nodes–SYMP rG = 0.984; SYMP–OB rG = 0.993; Nodes–OB rG = 0.984)(REHMAN et al., 2020; SAHAR et al., 2021). These associations indicate considerable redundancy among node, sympodial branch, and open-boll measures in this study.

The node–yield association was weakly positive: phenotypic r = 0.191 and genotypic r = 0.188. Thus, the observations do not support describing nodes as a strong stand-alone correlate of yield. Yield had positive associations with GOT (rP = 0.283; rG = 0.304) and PH (rP = 0.280; rG = 0.285), and negative associations with F Length (rP = -0.225; rG = -0.231) and Fiber U (rP = -0.250; rG = -0.264) (REHMAN et al., 2020). Correlation alone does not distinguish direct from indirect effects.

Table 4. Selected phenotypic and genotypic correlations among important breeding traits.

Trait pair	Phenotypic r	Genotypic r
Nodes – SYMP	0.954	0.984
SYMP – OB	0.952	0.993
Nodes – OB	0.940	0.984
Yield – Nodes	0.191	0.188
Yield – PH	0.280	0.285
Yield – GOT	0.283	0.304
Yield – F Length	-0.225	-0.231
Yield – Fiber U	-0.250	-0.264

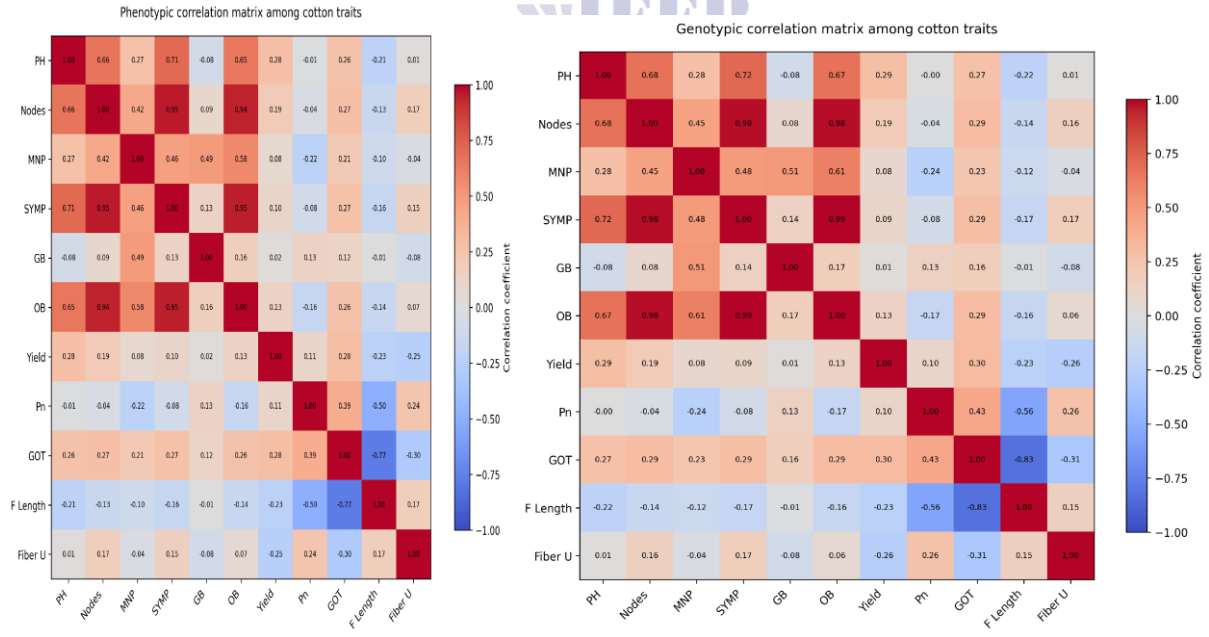

Figure 2 and 3. Phenotypic and genotypic correlation heatmap among the 11 traits.

Path coefficient analysis

The genotypic path model using PH, Nodes, GOT, F Length, and Fiber U explained $R^2 = 0.177$ (17.7%) of variation in yield among genotype means. The residual path coefficient was 0.907, indicating that most variation remained unexplained by the selected five predictors. The

largest positive direct effects were estimated for PH (0.226) and GOT (0.151); Nodes had a small direct effect (0.022) despite its weak positive total genotypic correlation with yield (0.188) (Chapepa et al., 2020). Fiber U had a negative direct effect (-0.218), while F Length had a near-zero negative direct effect (-0.020)(SAHAR et al., 2021). The model is exploratory and based on only 25 genotype means; coefficients can be sensitive to predictor selection and shared variation.

Table 5. Summary of genotypic path effects for yield using five selected predictors.

Predictor	Genotypic r with yield	Direct effect	Combined indirect effect	VIF
PH	0.285	0.226	0.059	1.94
Nodes	0.188	0.022	0.166	2.25
GOT	0.304	0.151	0.153	4.33
F Length	-0.231	-0.020	-0.210	3.64
Fiber U	-0.264	-0.218	-0.046	1.32

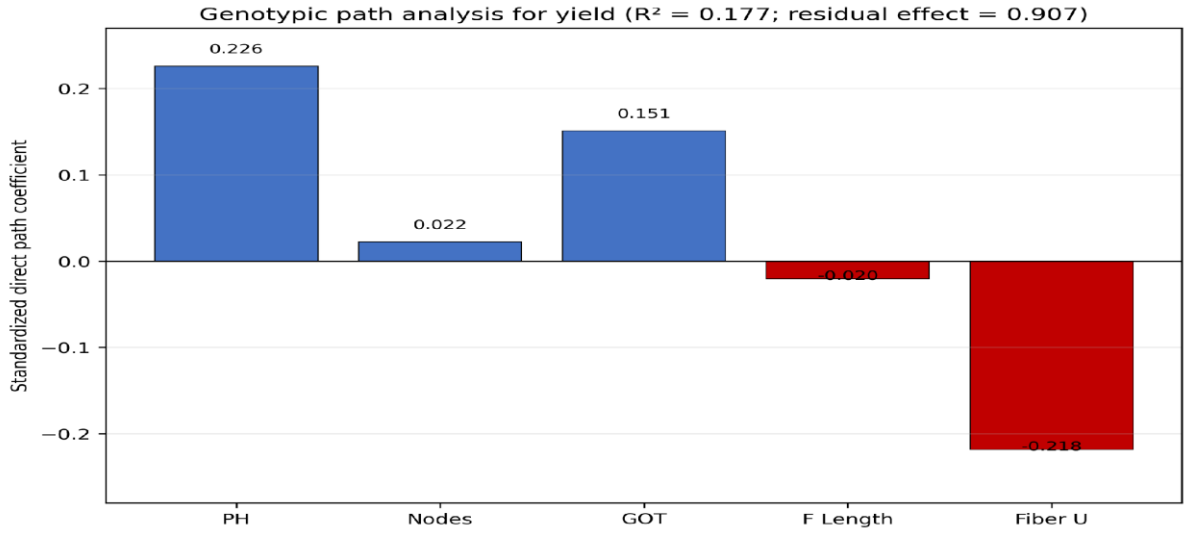

Figure 4. Standardized direct genotypic path coefficients for selected yield-associated traits.

Principal component analysis and biplot

PCA of standardized genotype means showed that PC1 explained 36.63% of the total variation and PC2 explained 19.56%, for a combined 56.18%. PC1 had high positive loadings for OB (0.940), SYMP (0.935), Nodes (0.919), PH (0.761), and MNP (0.603), representing an architecture/reproductive-trait axis. PC2 contrasted F Length (loading = 0.817) with GOT (-0.764) and Pn (-0.674), indicating that fiber length and these traits contributed in opposite directions to the second axis (Zafar et al., 2022).

Table 6. Trait loadings on the first two principal components.

Trait	PC1 loading	PC2 loading
PH	0.761	0.037
Nodes	0.919	0.206
MNP	0.603	0.168
SYMP	0.935	0.229
GB	0.216	-0.045
OB	0.940	0.253
Yield	0.286	-0.380

Pn	0.006	-0.674
GOT	0.479	-0.764
F Length	-0.353	0.817
Fiber U	0.018	0.334

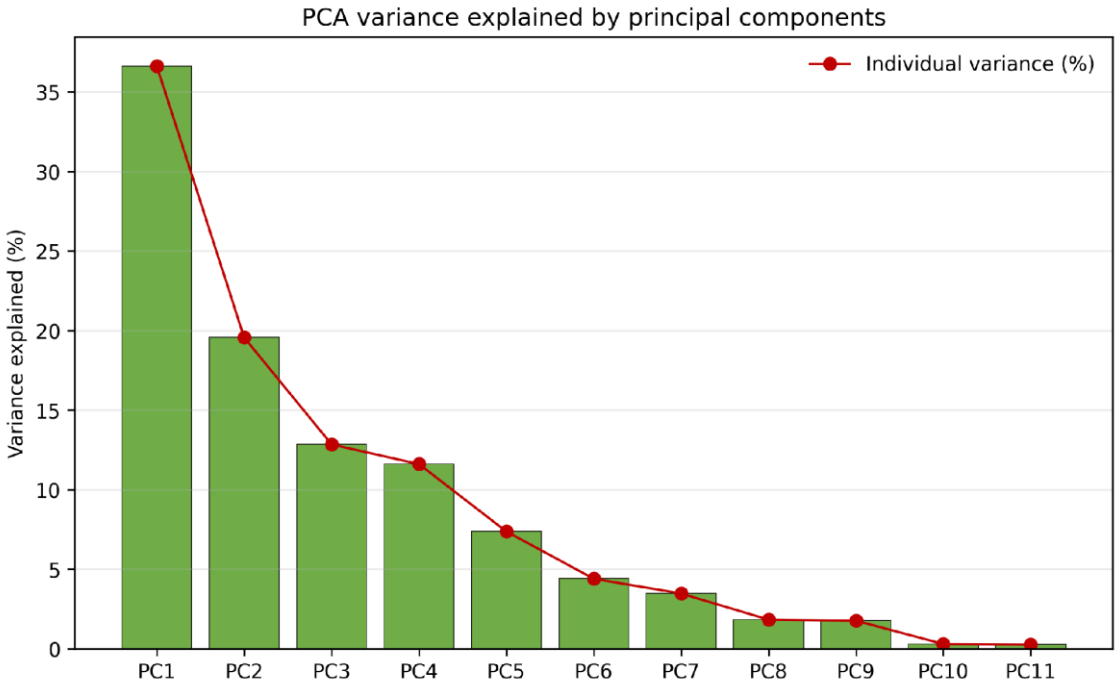

Figure 5. Percentage of standardized variance explained by principal components.

IEER
Institute for Excellence in Education & Research

Discussion

The experiment demonstrated how several statistical approaches can be combined to evaluate cotton varieties. The RCBD analysis detected differences among variety means for all traits, while the LSD threshold translated yield comparisons into a practical pairwise criterion. V15 was the highest-yielding entry numerically and was separated from V18 by more than the yield LSD. However, confidence in these comparisons depends on valid randomization and a residual mode(Frey et al., 2024).

The strong relationships among Nodes, SYMP, and OB suggest that these traits share substantial information. In a breeding program, selecting simultaneously on several highly correlated architectural traits can be redundant. Yet the observed node–yield relationship was weak (rP = 0.191; rG = 0.188), meaning that selection for higher node number alone is not supported as an efficient yield strategy by these records. This is a meaningful result: a hypothesized association should be tested, not imposed by adjusting data (Chapepa et al., 2020).

The path model further showed that Nodes had a small direct genotypic effect after PH, GOT, F Length, and Fiber U were included. PH and GOT had larger positive direct effects in this model, while Fiber U had a negative direct coefficient. The total explained fraction was low (R^2 = 0.177) (SAHAR et al., 2021) and the predictor VIFs indicated some shared variation, particularly for GOT and F Length. These estimates should be treated as exploratory guides for designing a more robust analysis rather than as definitive causal effects(Kalnins & Praitis Hill, 2025).

PCA summarized the dominant multi-trait structure: PC1 primarily captured variation in plant architecture and open-boll/sympodial traits, whereas PC2 reflected a contrast between fiber length and GOT/Pn. Repeating the analysis on verified field observations, and ideally across environments or years, is essential for robust variety selection(Elsamman et al., 2024).

Finally, large residual CVs for MNP and GB require contextual interpretation, as these traits have many zeros and low counts. A count-data or generalized mixed-model approach may be worth considering if residual assumptions are poor. The standard RCBD ANOVA used here is a practice baseline, and residual plots/model diagnostics should be reviewed before final inference (Li et al., 2025).

Conclusions

The supplied practice dataset illustrates substantial differences among coded varieties for yield and several architectural/fiber traits. V15 had the highest mean yield, followed by V18 and V17; however, the data provenance prevents a real breeding recommendation. Node number was positively but weakly correlated with yield and had a small direct path coefficient in the selected genotypic path model. Consequently, node number should not be used as a sole proxy for yield. PH and GOT showed relatively stronger positive relationships with yield in the selected model, while fiber-quality traits showed negative correlations with yield in this dataset; those relationships warrant confirmation rather than automatic selection decisions.

For a practical breeding program, the next step is to replace synthetic replication records with plot observations, confirm all trait definitions and units, validate the RCBD randomization, and rerun the analysis. A selection index should combine yield with fiber quality and agronomic suitability, and promising entries should be evaluated in independent environments before advancement.

REFERENCES:

Chapara, R., Sudha Rani, M., Siva Reddy, K. V., Chamundeswari, N., Rajamani, S., Sree Lakshmi, B., Sreekanth, B., Diana Grace, G. A., Venkata Lakshmi, N., Rajesh Chowdary, L., Jyothi, Ch. N., Manoj Kumar, V., Prasada Rao, G. M. V., Durga Prasad, N. V. V. S., & James, M. (2025). Evaluation of Genetic Diversity and Marker Trait Association for Yield and Fiber Quality-Related Traits in Upland Cotton. *Journal of Natural Fibers*, *22*(1), 2508290. https://doi.org/10.1080/15440478.2025.2508290

Chapepa, B., Mubvekeri, W., Mare, M., & Kutywayo, D. (2020). Correlation and path coefficient analysis of polygenic traits of upland cotton genotypes grown in Zimbabwe. *Cogent Food & Agriculture*, *6*(1), 1823594. https://doi.org/10.1080/23311932.2020.1823594

Elsamman, E. Y., Ge, Q., Wang, X., Lamlom, S. F., Gong, J., Li, J., Yan, H., Zhong, Y., Bai, B., Qiao, D., Gong, W., Yuan, Y., & Abdelghany, A. M. (2024). Elucidating the phenotypic basis of multi-environment stability for fiber yield and quality traits of cotton (*Gossypium hirsutum* L.) using 498 recombinant inbred lines. *Industrial Crops and Products*, *215*, 118593. https://doi.org/10.1016/j.indcrop.2024.118593

Frey, J., Hartung, J., Ogutu, J., & Piepho, H.-P. (2024). Analyze as randomized—Why dropping block effects in designed experiments is a bad idea. *Agronomy Journal*, *116*(3), 1371–1381. https://doi.org/10.1002/agj2.21570

Kalnins, A., & Praitis Hill, K. (2025). The VIF Score. What is it Good For? Absolutely Nothing. *Organizational Research Methods*, *28*(1), 58–75. https://doi.org/10.1177/10944281231216381

Li, H., Baek, E., Luo, W., Du, W., & Lam, K. H. (2025). Using Generalized Linear Mixed Models in the Analysis of Count and Rate Data in Single-case Eperimental Designs: A Step-by-step Tutorial. *Evaluation & the Health Professions*, *48*(1), 143–155. https://doi.org/10.1177/01632787241259500

REHMAN, A., MUSTAFA, N., DU, X., & AZHAR, M. T. (2020). Heritability and correlation analysis of morphological and yield traits in genetically modified cotton. *Journal of Cotton Research*, *3*(1), 23. https://doi.org/10.1186/s42397-020-00067-z

SAHAR, A., ZAFAR, M. M., RAZZAQ, A., MANAN, A., HAROON, M., SAJID, S., REHMAN, A., MO, H., ASHRAF, M., REN, M., SHAKEEL, A., & YUAN, Y. (2021). Genetic variability for yield and fiber related traits in genetically modified cotton. *Journal of Cotton Research*, *4*(1), 19. https://doi.org/10.1186/s42397-021-00094-4

Shahzad, K., Mubeen, I., Zhang, M., Zhang, X., Wu, J., & Xing, C. (2022). Progress and perspective on cotton breeding in Pakistan. *Journal of Cotton Research*, *5*(1), 29. https://doi.org/10.1186/s42397-022-00137-4

Zafar, M. M., Jia, X., Shakeel, A., Sarfraz, Z., Manan, A., Imran, A., Mo, H., Ali, A., Youlu, Y., Razzaq, A., Iqbal, M. S., & Ren, M. (2022). Unraveling Heat Tolerance in Upland Cotton (Gossypium hirsutum L.) Using Univariate and Multivariate Analysis. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.727835