



Trait association and principal component analysis of morphological, biochemical and yield traits in upland cotton (*Gossypium hirsutum* L.)

Om Prakash Yadav^{1*}, Minakshi Jattan¹ and Harpreet Singh²

<https://doi.org/10.37992/2026.1703.037>

Abstract

The present investigation evaluated 127 upland cotton genotypes during *Kharif* 2021 and 2022 at CCS Haryana Agricultural University, Hisar, to characterize morphological, biochemical, and yield traits using principal component, correlation, and cluster analyses. PCA analysis showed four components with eigenvalues greater than one, explaining 71.57 % of the total variation. PC1 and PC2 together contributed 53.67 % of the variability and were mainly influenced by biochemical and yield-contributing traits, respectively. K-means clustering of the PCA score plot grouped the genotypes into four distinct clusters, indicating substantial genetic divergence. Correlation analysis showed that seed cotton yield was significantly and positively associated with lint yield, number of bolls per plant, seed index, boll weight, and leaf trichome density. Cluster II exhibited distinct divergence from Clusters I and III, while Cluster IV displayed intermediate variability among groups. Genotypes HS 45, SA 320 F, and Atlas 59-182 showed unique trait combinations, making them valuable for future cotton improvement programs.

Keywords: Clusters, Correlation, Cotton, Principal component analysis.

INTRODUCTION

Cotton, globally recognized as the most important natural fibre sustains the textile industry and also contributes substantially to India's rural economy. Cotton is one of the oldest cultivated crops and is grown predominantly in tropical and subtropical regions for its fibre and seed oil. India has a major role in global cotton production, accounting for approximately 20.67 % of worldwide output (ICAR-AICRP Annual Report 2024–25). Because of its substantial contribution to the agricultural economy, cotton is widely referred to as "white gold" (Komala *et al.*, 2018).

The genus *Gossypium* includes 46 diploid ($2n = 26$) and 7 allotetraploid ($2n = 52$) species representing the AA, BB, CC, DD, EE, FF, GG, KK and AADD genomes (Shim *et al.*, 2018). Eight diploid genomes (A to G and K) have been assigned to these species. The genomes A, B, E and F are prevalent in Africa and Asia. The D genome is native to America, while C, G, and K genomes have appeared in Australia (Malik *et al.*, 2014). Out of 53 cotton species, only four are cultivated. Among cultivated species, two diploid species ($2n = 26$) i.e., *G. arboreum* L. (A_2 genome) and *G. herbaceum* L. (A_1 genome) are commonly known as 'Asiatic' or 'old world cotton' whereas, two tetraploid species ($2n = 52$) i.e., *G. barbadense* L. (AD_2 genome) and *G. hirsutum* L. (AD_1 genome) are known as 'new world cotton' (Sheidai, 2008 and Kusugal *et al.*, 2014). *G. hirsutum* is known as upland cotton or Mexican cotton. The word *hirsutum* is derived from the Latin word "hirsutus" means 'hairy' and more than 90 % of the cotton growing area is covered by *G. hirsutum* (Ding *et al.*, 2015).

¹Department of Genetics and Plant Breeding, Chaudhary Charan Singh Haryana Agriculture University, Hisar, Haryana-125004, India.

²Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, Punjab -141004, India.

Corresponding author: Om Prakash Yadav, Department of Genetics and Plant Breeding, Chaudhary Charan Singh Haryana Agriculture University, Hisar, Haryana-125004, India.
*E-Mail: opyadavpbg@gmail.com

How to cite this article: Om Prakash Yadav, Minakshi Jattan and Harpreet Singh. 2026. Trait association and principal component analysis of morphological, biochemical and yield traits in upland cotton (*Gossypium hirsutum* L.). *Electronic Journal of Plant Breeding*, 17(3): 296-306
doi:10.37992/2026.1703.037

Received: 27.05.2026 **Revised:** 06.08.2026 **Accepted:** 18.08.2026

Cotton productivity continues to be limited by several production constraints, among which the scarcity of high-yielding cultivars remains a key factor. In addition, drought and heat are common abiotic stresses during the cotton-growing season and can severely impair crop growth, yield, and fibre quality (Sen *et al.*, 2022). Therefore, the breeders are focusing to develop high yielding superior genotypes. A narrow genetic base and limited diversity among cotton genotypes are considered major factors for yield reduction (Rathinavel, 2017 and Sagar *et al.*, 2023). Therefore, broadening the genetic base of cultivated

varieties requires the use of genetically divergent parents in breeding programmes aimed at developing high-yielding cotton genotypes (Shakeel *et al.*, 2015 and Sun *et al.*, 2019). Effective utilization of the available germplasm through hybridization, along with the introduction of novel genetic resources, is essential for generating substantial genetic variation and identifying superior genotypes. Principal component analysis (PCA) and cluster analysis are widely used to evaluate genetic similarities and differences among genotypes and to classify them into distinct groups (Asha *et al.*, 2013; Naik *et al.*, 2016 and Sarwar *et al.*, 2021). PCA is particularly useful because it explains the total variation in the data and helps identify elite parental lines by pinpointing the traits that contribute most to divergence. It also facilitates the assessment of phenotypic variability and genetic diversity among genotypes, thereby supporting informed selection in breeding trials. By revealing the traits that contribute most strongly to variation, PCA provides valuable guidance for selecting genotypes of genetic importance (Isong *et al.*, 2017). The present study was undertaken to evaluate genetic diversity in morphological, biochemical, yield, and yield-related traits among 127 upland cotton genotypes and to determine the associations among the traits.

MATERIALS AND METHODS

The experimental material consisted of 127 genotypes of upland cotton (*G. hirsutum* L.) including two checks (RS 2013 and Ganganagar Ageti) obtained from Cotton Section, Genetics and Plant Breeding Department, Chaudhary Charan Singh Haryana Agricultural University, Hisar. Each genotype was sown in Randomized Block Design (RBD) with three replications during *Kharif* 2021 and 2022. The field experiment was conducted at Research Area of Cotton Section, Department of Genetics and Plant Breeding, CCS Haryana Agricultural University, Hisar. It is located in semi-tropical region of Western zone of India at 29°10'N latitude, 75°46'E longitude with an altitude of 215.2 m above mean sea level. The experimental studies were carried out in sandy loam with homogeneous nutrient levels. All the agronomic practices were followed as per recommended package of practices for cotton crop. Each genotype was sown in a single row of 6 m length with a row spacing of 67.5 cm and a plant spacing of 30 cm. For each genotype, five plants per replication were tagged for recording observations on morphological and biochemical traits. Data on field-based traits were recorded, except for leaf trichome density, trichome length, and biochemical parameters (total sugar, total phenol, tannin content, and gossypol content), which were estimated in the Entomology Laboratory and Biochemistry Laboratory of the Cotton Section, respectively. Observations on seed cotton yield, ginning outturn, plant height, seed index, boll weight, lint yield, lint index, and number of bolls per plant were recorded in the Field Laboratory of the Cotton Section.

Analysis of variance was performed to determine the genetic variation among the genotypes. Correlation

analysis between various yield attributing traits and biochemical components were derived as per the Pearson's correlation coefficient. Multivariate analysis such as principal component analysis (PCA) and biplot analysis was done. Data were analyzed using the software R Studio (R Studio, 2020).

RESULTS AND DISCUSSION

ANOVA

The analysis of variance components in the present study revealed substantial genetic variability among 127 cotton genotypes for morphological, biochemical and yield-related traits. Analysis of variance indicated highly significant ($P < 0.01$) differences among genotypes for all fourteen traits studied (**Table 1**), including seed cotton yield, plant height, ginning outturn, seed index, boll weight, lint yield, lint index, number of bolls per plant, leaf trichome density, trichome length, total sugar, total phenol, tannin content and gossypol content. The significant variation among genotypes reflects the existence of substantial genetic variability for these agronomic, morphological and biochemical traits. Replication effects were significant for seed cotton yield, plant height, boll weight, lint yield, lint index, number of bolls per plant, trichome length, total phenol and tannin content, whereas they were non-significant for ginning outturn, seed index, leaf trichome density, total sugar and gossypol content. The comparatively low error mean squares indicate good experimental precision, suggesting that the observed variation was primarily attributable to genetic differences among the evaluated cotton genotypes. Anjani *et al.* (2020) conducted ANOVA on 40 Upland cotton genotypes and observed significant genotype effects for all traits, supporting the need for variance analysis in our study.

Variability parameters

The mean performance along with standard error of mean for the fourteen morphological and biochemical traits studied is presented in **Table 2**. Among the yield and yield-contributing traits, seed cotton yield averaged 63.82 ± 3.61 g, plant height 148.20 ± 2.94 cm, ginning outturn 38.58 ± 0.74 %, seed index 5.90 ± 0.14 g, boll weight 2.91 ± 0.10 g, lint yield per plant 23.41 ± 1.90 g, lint index 4.02 ± 0.14 g, and number of bolls per plant 20.83 ± 0.87 . Among the leaf-surface and biochemical traits associated with sucking pest resistance, leaf trichome density recorded the highest mean value (126.94 ± 5.88), followed by trichome length (32.71 ± 1.59 μ m) and total sugars (27.43 ± 0.95 mg/g), while phenol, tannin, and gossypol contents averaged 11.51 ± 0.31 , 5.91 ± 0.19 , and 3.59 ± 0.12 mg/g, respectively. The relatively low standard error associated with most trait means indicate a reasonable degree of precision in the estimates and adequate replication in the experimental layout, lending confidence to the mean values used for subsequent genetic analysis.

Genetic variability parameters estimated for the fourteen morphological and biochemical traits (**Table 2**) revealed that phenotypic coefficient of variation (PCV) was only

Table 1. Analysis of variance (mean sum of squares) for yield, morphological, fibre quality and biochemical traits

S.No.	Traits	Replications MS (df = 2)	Genotypes MS (df = 126)	Error MS (df = 252)
1.	Seed cotton yield per plant	244.945	850.648**	39.2041
2.	Plant height	198.652	402.766**	25.9082
3.	Ginning outturn	0.219	16.654**	1.6501
4.	Seed index	0.177	0.632**	0.0618
5.	Boll weight	0.217	0.470**	0.0308
6.	Lint yield	156.076	49.103**	10.7767
7.	Lint index	0.299	0.476**	0.0573
8.	Bolls per plant	50.662	12.993**	2.2557
9.	Leaf trichome density	298.107	25060.592**	103.6718
10.	Trichome length	114.602	404.688**	7.6085
11.	Total sugar	5.046	223.936**	2.7197
12.	Total phenol	1.436	12.074**	0.2855
13.	Tannin content	0.873	6.075**	0.1077
14.	Gossypol content	0.064	3.922**	0.0419

**Significant at 1% level of probability

Table 2. Estimation of genetic components for yield and yield attributing traits

S.No.	Traits	Mean $\pm$ SE	GCV (%)	PCV (%)	h^2 (bs) (%)	GA (% of mean)
1.	Seed cotton yield per plant (g)	63.82 $\pm$ 3.61	25.77	27.58	87.34	49.69
2.	Plant height (cm)	148.20 $\pm$ 2.94	7.56	8.31	82.90	14.21
3.	Ginning outturn (%)	38.58 $\pm$ 0.74	5.80	6.68	75.28	10.38
4.	Seed index	5.90 $\pm$ 0.14	7.39	8.51	75.45	13.25
5.	Boll weight (g)	2.91 $\pm$ 0.10	13.15	14.47	82.58	24.65
6.	Lint yield (g)	23.41 $\pm$ 1.90	15.27	20.73	54.25	23.20
7.	Lint index	4.02 $\pm$ 0.14	9.29	11.03	70.93	16.14
8.	Number bolls per plant	20.83 $\pm$ 0.87	9.08	11.60	61.36	14.68
9.	Leaf trichome density (cm ²)	126.94 $\pm$ 5.88	71.85	72.30	98.77	47.32
10.	Trichome length (μ m)	32.71 $\pm$ 1.59	35.17	36.16	94.56	70.55
11.	Total sugar (mg/g)	27.43 $\pm$ 0.95	31.30	31.88	96.44	63.42
12.	Total phenol (mg/g)	11.51 $\pm$ 0.31	17.23	17.84	93.23	34.32
13.	Tannin content (mg/g)	5.91 $\pm$ 0.19	23.87	24.50	94.86	47.95
14.	Gossypol content (mg/g)	3.59 $\pm$ 0.12	31.70	32.21	96.90	64.38

marginally higher than genotypic coefficient of variation (GCV) for all traits, indicating limited environmental influence and a close correspondence between phenotypic and genotypic variability. High GCV and PCV (>20 %) were recorded for leaf trichome density, trichome length, gossypol content, total sugars, tannin content, and seed cotton yield, reflecting substantial exploitable genetic variability, while plant height, ginning outturn, seed index, lint index, and number of bolls per plant showed low GCV and PCV, indicating a narrow genetic base for these characters. Broad-sense heritability was high (>60 %) across all traits, ranged from 54.25 %

(lint yield per plant) to 98.77 % (leaf trichome density), and genetic advance as percent of mean ranged from 10.38% (ginning outturn) to 70.55 % (leaf trichome length). High heritability accompanied by high genetic advance as percent of mean was observed for leaf trichome density, trichome length, total sugars, tannin, gossypol and phenol content, boll weight, lint yield per plant, and seed cotton yield, suggesting these traits are governed predominantly by additive gene action and would therefore respond well to direct phenotypic selection, since heritability alone, without genetic advance, can be inflated by non-additive effects and is a less reliable predictor of selection

response (Johnson *et al.*, 1955). Conversely, plant height, ginning outturn, seed index, lint index, and number of bolls per plant combined high heritability with low-to-moderate genetic advance as percent of mean, indicating that non-additive gene action (dominance and epistasis) likely contributes to their expression, for such traits, breeding approaches that exploit non-additive variance, such as heterosis breeding, would be more effective than simple mass selection (Panse, 1957).

Principal component analysis and PCA-Biplot

Principal component analysis (PCA) is a multivariate statistical approach used to examine complex and large datasets. Based on the association among the studied characters and extracted clusters, the variation pattern in

cotton genotypes was also examined by PCA to determine the relationship among different variables that contributed to variability in the studied data (Isong *et al.*, 2017). PCA revealed that four of the fourteen principal components (PCs) had eigenvalues greater than one (>1), collectively accounting for 71.57 % of the total trait variability (Table 3 & Fig. 1). This indicates that a substantial proportion of the variation present in the evaluated genotypes could be summarized by a limited number of components, reflecting the structured nature of trait association in the material studied.

The first principal component (PC1) explained the highest variation among the studied traits, with major contributions from gossypol content (19.47 %), leaf trichome density

Table 3. Eigen vector and variance of principal components of cotton genotypes

Principal components	Eigenvalue	Variance (%)	Cumulative variance (%)
PC1	4.18	29.83	29.83
PC2	3.34	23.84	53.67
PC3	1.40	10.00	63.67
PC4	1.11	7.91	71.57
PC5	0.84	6.00	77.58
PC6	0.83	5.94	83.51
PC7	0.61	4.37	87.89
PC8	0.58	4.11	92.00
PC9	0.34	2.46	94.46
PC10	0.30	2.11	96.56
PC11	0.16	1.12	97.68
PC12	0.14	1.03	98.71
PC13	0.11	0.80	99.51
PC14	0.07	0.49	100.00

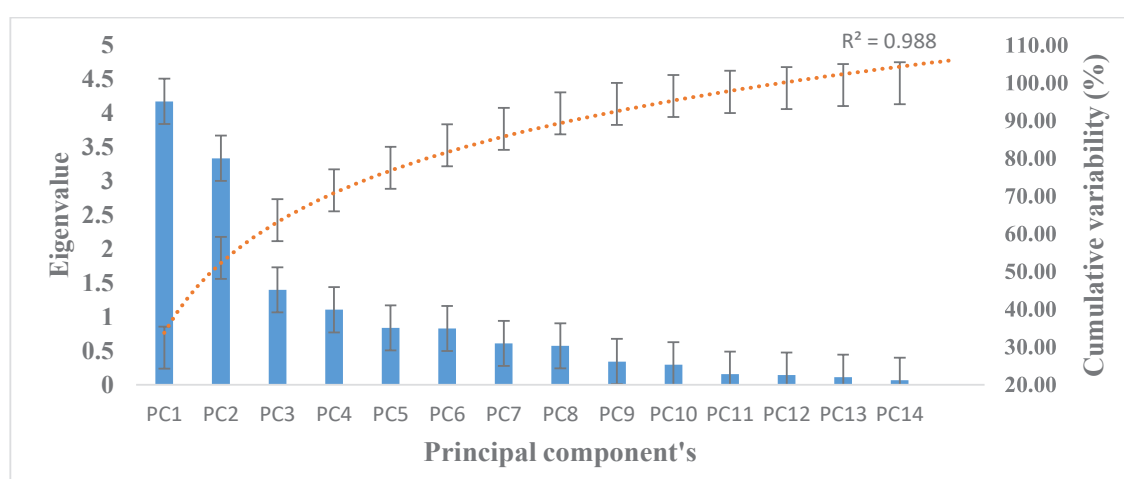


Fig. 1. Scree plot depicting the eigenvalues and cumulative percentage of variance explained by the principal components

SCY - Seed cotton yield per plant, GOT - Ginning outturn, PH - Plant height, SI - Seed index, BW - Boll Weight, LY - Lint yield, LI - Lint index, BPP - Number of boll per plant, LTD - Leaf trichome density on abaxial surface, TL - Trichome length, TS - Total sugar, TPh - Total phenol, TC - Tannin content, GC - Gossypol content

(17.24 %), total sugar content (15.85 %), total phenol content (15.82 %), and tannin content (14.51 %). The predominance of these biochemical and morphological defense-related traits in PC1 suggests that the first axis of variation was mainly associated with plant defense mechanisms. PC2 showed major contribution from seed index (20.70), lint index (19.66), lint yield (18.39), and boll weight (13.26), indicating that the second major axis of variation was related to economically important yield and fibre traits. PC3 showed major contribution from seed cotton yield (17.51), number of bolls per plant (13.70), boll weight (12.43), tannin content (12.30), total phenol content (10.10), and ginning outturn (9.73). PC4 was largely represented by plant height (38.77) and number of bolls per plant (32.36) (**Table 4**). Together, these results showed that the total variation in the material was distributed across both defense-related and yield-related attributes, with some overlap between these two trait groups.

PCA of the traits showed that the greatest amount of variation among genotypes was observed in leaf trichome density on the abaxial surface (33.90 %), followed by gossypol content (13.90 %), total sugar content (12.50 %), and tannin content (9.80 %). Moderate variation was observed for trichome length (7.80 %) and total phenol content (6.50 %). Traits related to yield and morphology showed lower variance such as, seed cotton yield (4.40 %), boll weight (2.50 %), plant height (2.40 %), seed index (1.90 %), ginning outturn (1.70 %), lint index (1.20 %), number of bolls per plant (1.10 %), and lint yield (0.50 %) (**Table 4**). The comparatively higher contribution of trichome density, gossypol, sugars, phenols, and tannins indicates that these traits played a major role in discriminating among the genotypes and may be important indicators of defense response in cotton.

Ullah *et al.* (2022) performed PCA on 40 cotton genotypes under water stress, the first two PCs captured 97.5 % of variance. This supports the present study that the use of PCA to summarize trait variation.

The PCA further substantiated these relationships by partitioning total variability into four significant components, collectively explaining 71.57 % of the variation. PC1 was dominated by biochemical defense-related traits (gossypol, phenols, tannins, and trichome density), while PC 2 captured major yield-related traits such as seed index, lint yield, and boll weight. This separation reflects the genetic independence of yield and defense pathways, offering opportunities for concurrent selection. Similar PCA-based clustering of yield and resistance traits has been observed in cotton germplasm diversity studies (Liang *et al.*, 2023, Joshi *et al.*, 2023 and Navingautham *et al.*, 2026). The PCA biplot explained 53.67 % of the total variation and clearly grouped traits into two functional clusters: first yield-associated traits (seed cotton yield, boll number, lint yield, boll weight) and second defense-related traits (trichome density, gossypol, tannins, phenols, total sugar). The negative association between total sugar and defense-related traits indicates possible physiological trade-offs, as previously suggested by Yue *et al.* (2024). The K-means clustering further highlighted the substantial genetic diversity, separating the 127 genotypes into four distinct groups. Genotypes such as HS 45, SA 320 F and Atlas 59-182 formed unique clusters, suggesting their potential as parental lines for hybridization to maximize heterotic expression. Their selection was based not only on genetic divergence but also on their desirable agronomic and biochemical attributes. HS 45 exhibited high seed cotton yield (109.29 g plant⁻¹), high leaf trichome density (639.10), and low total sugar content (6.57 mg g⁻¹). SA 320 F possessed

Table 4. Variability and PCA by different traits of cotton genotypes

S.No.	Characters	Variability contribution (%)	Contribution of the variables			
			PC1	PC2	PC3	PC4
1.	Seed cotton yield per plant	4.40	6.03	5.04	17.51	1.15
2.	Plant height	2.40	0.25	2.53	2.51	38.77
3.	Ginning outturn	1.70	1.06	7.88	9.73	2.89
4.	Seed index	1.90	0.64	20.70	1.25	1.13
5.	Boll weight	2.50	0.64	13.26	12.43	5.73
6.	Lint yield	0.50	1.42	18.39	1.01	0.89
7.	Lint index	1.20	1.73	19.66	1.21	1.88
8.	Bolls per plant	1.10	1.32	2.11	13.70	32.36
9.	Leaf trichome density on abaxial surface	33.90	17.24	1.66	5.50	2.66
10.	Trichome length	7.80	4.02	3.14	4.17	1.82
11.	Total sugar	12.50	15.85	1.09	5.45	2.17
12.	Total phenol	6.50	15.82	0.86	10.10	2.99
13.	Tannin content	9.80	14.51	1.12	12.30	4.24
14.	Gossypol content	13.90	19.47	2.56	3.13	1.32

high phenol (13.00 mg g^{-1}) and gossypol (5.15 mg g^{-1}) contents, together with lower seed cotton yield and fewer bolls per plant. Atlas 59-182 was characterized by a high ginning outturn (40.56%) and a lower number of bolls

per plant. These complementary traits, in combination with their genetic divergence, indicate their potential as promising parental genotypes for developing high-yielding and insect-tolerant cotton cultivars through breeding.

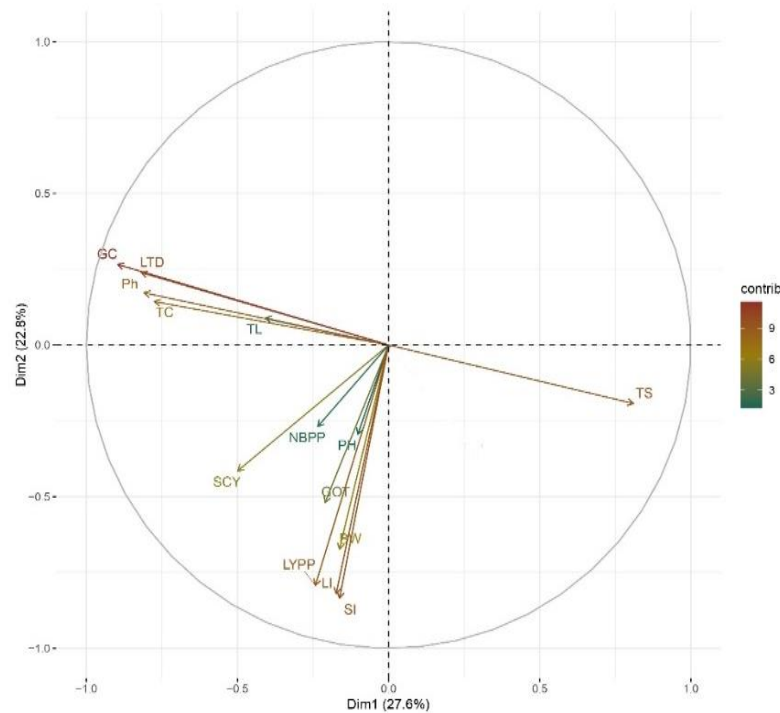


Fig. 2. PCA biplot representing the distribution of fourteen variables traits of cotton genotypes

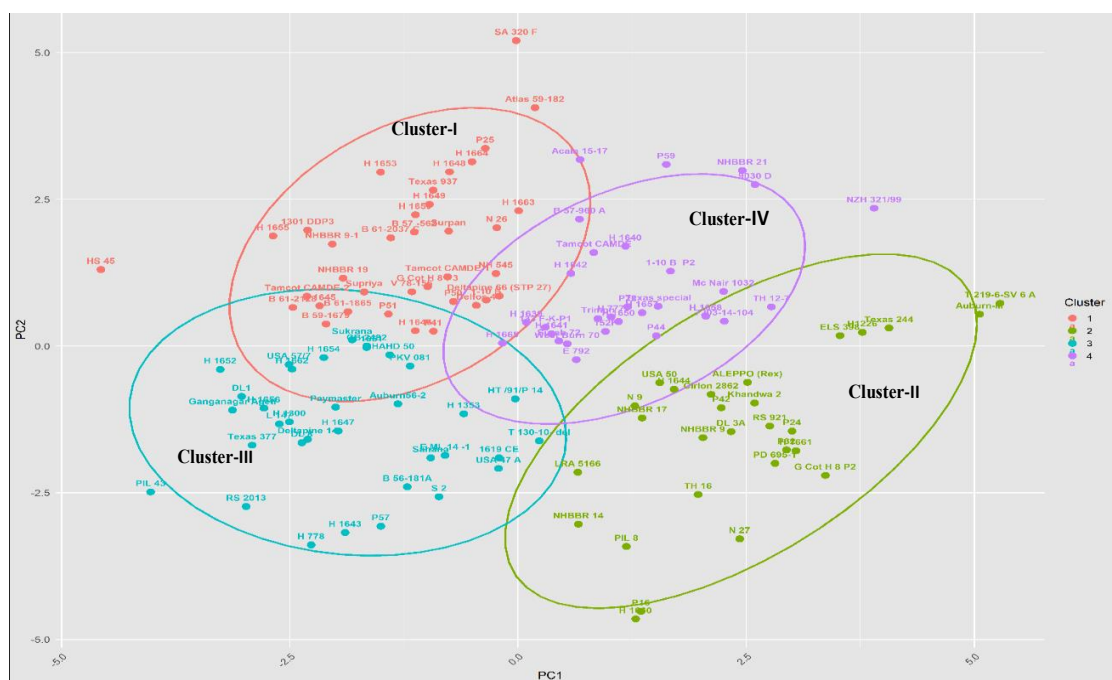


Fig. 3. PCA biplot illustrating the distribution of cotton genotypes grouped into four distinct clusters

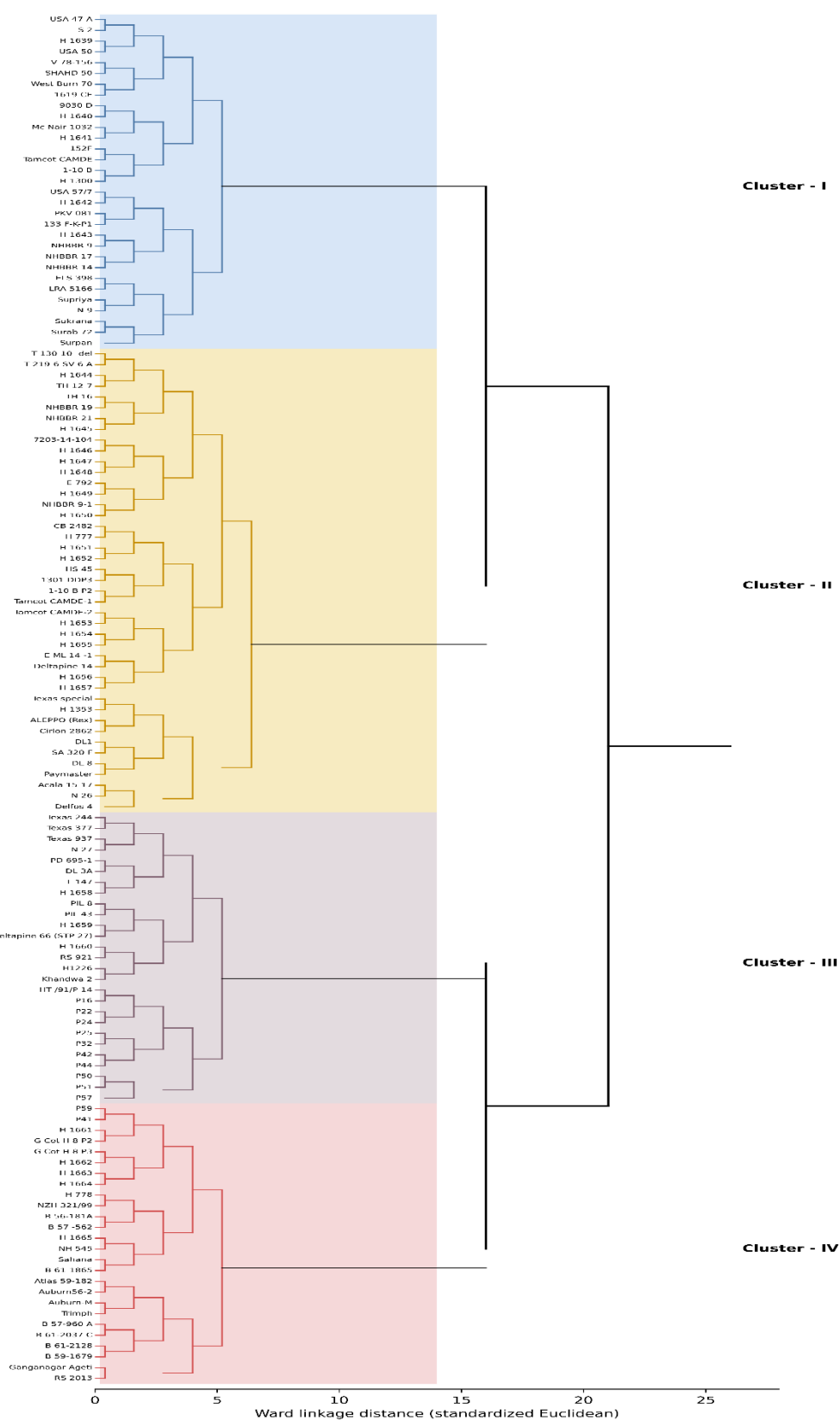


Fig. 4. Dendrogram illustrating the genetic relationships among 127 cotton genotypes as revealed by cluster analysis

PCA projects multidimensional trait data onto a two-dimensional plane called a biplot, where genotypes appear as points and traits as vectors. In this study, the biplot was constructed using PC1 and PC2, showing genotype dispersion and trait contributions to variability. The PCA biplot based on the first two components (PC1: 29.83 % & PC2: 23.84 %) explained 53.67 % of the total variability among genotypes. Yield-contributing traits such as seed cotton yield, number of bolls per plant, boll weight, and lint yield were closely associated, indicating their positive interdependence. In contrast, traits like trichome density, trichome length, gossypol, tannin, phenol, and total sugar formed a separate group, highlighting their distinct contribution to variability. Subsequently, it was observed that total sugar was negative correlated with traits like trichome density, trichome length, gossypol, tannin, and phenol (Fig. 2). Reddy *et al.* (2015) also used PCA and hierarchical clustering on 63 cotton genotypes and found 8 clusters, with the first 7 PCs explaining 87.89 % of variability. This demonstrates similar diversity patterns in cotton and validates the choice of multivariate analysis in the present study.

The PCA scatterplot combined with K-means clustering (Fig. 3) and dendrogram (Fig. 4) revealed clear differentiation of the cotton genotypes into four distinct clusters, reflecting substantial genetic diversity. Genotypes HS 45, SA 320 F, and Atlas 59-182 were positioned on the positive side of PC2, indicating unique trait combinations that distinguish them from other groups. Cluster I comprised 31 genotypes, while Cluster II included 43 genotypes mainly aligned along positive PC1, showing a distinct grouping with greater genetic distance from Clusters I and III. Cluster III contained 27 genotypes spread toward the negative side of PC2, reflecting moderate divergence with some overlap, whereas Cluster IV, comprising 26 genotypes, exhibited intermediate variability and shared traits with Clusters I, II, and III.

Correlation analysis

Correlation analysis (Fig. 5) revealed that seed cotton yield exhibited a significant positive correlation with leaf trichome density on the abaxial surface (0.46), lint yield (0.43), seed index (0.38), number of bolls per plant (0.33),

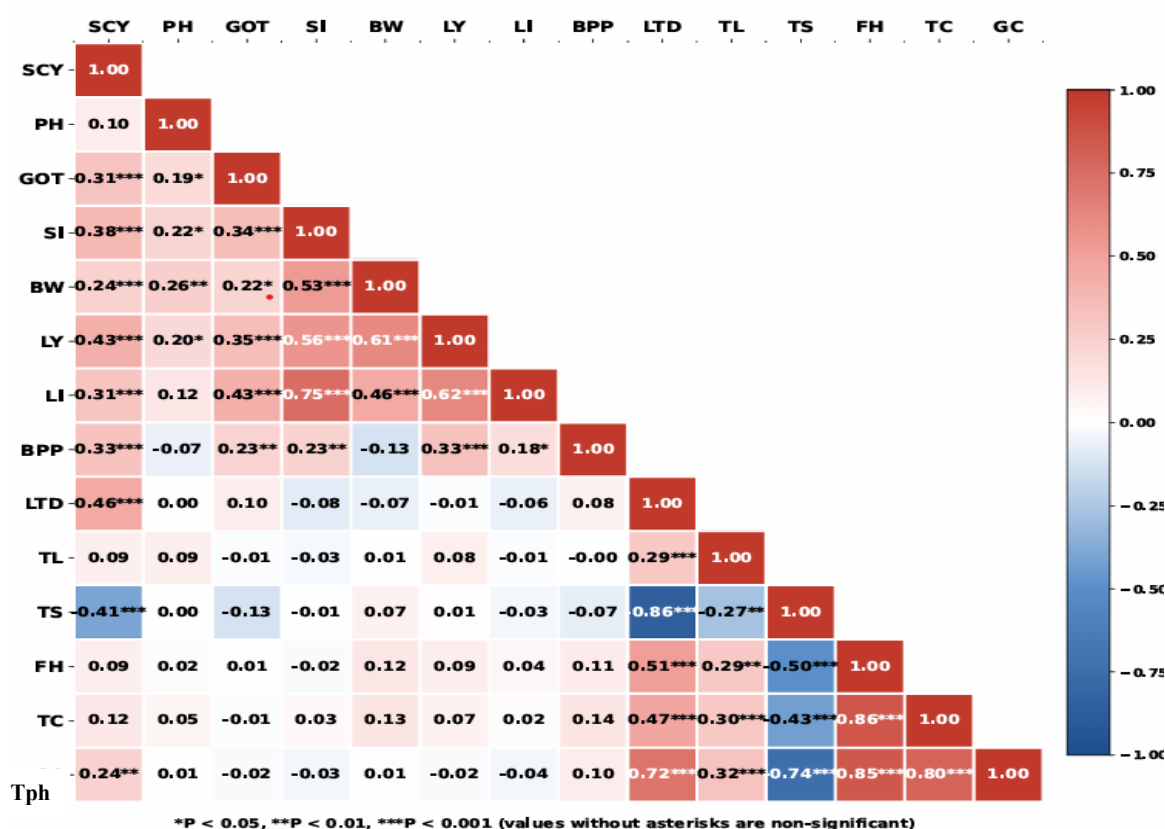


Fig. 5. Correlogram showing correlations and scatterplots matrix of 14 traits

SCY - Seed cotton yield per plant, GOT - Ginning outturn, PH - Plant height, SI - Seed index, BW - Boll Weight, LY - Lint yield, LI - Lint index, BPP - Number of bolls per plant, LTD - Leaf trichome density on abaxial surface, TL - Trichome length, TS - Total sugar, TPh - Total phenol, TC - Tannin content, GC - Gossypol content

ginning outturn (0.31), lint index (0.31), gossypol content (0.24) and boll weight (0.24). Conversely, a significant negative correlation was observed with total sugar content (-0.41). Non-significant positive associations were found with plant height (0.10), total phenol content (0.09), tannin content (0.12) and trichome length (0.09). These associations suggest that simultaneous selection for boll number, boll weight, and ginning outturn, lint index, lint yield and seed index boll number, boll weight, ginning outturn, lint index, lint yield and seed index traits could enhance yield potential. Similar positive associations of yield with boll weight, lint yield, and boll number have been reported in recent studies on upland cotton (Sahar *et al.*, 2021). The significant negative association between seed cotton yield and total sugar content may be due to reason that enhanced sugar accumulation may redirect carbon partitioning away from lint biomass and Luo *et al.* (2019) also reported similar results. Rao and Gopinath (2013) found strong positive correlations between seed cotton yield and yield components (bolls per plant, boll weight, ginning out-turn, lint yield). This aligns with the present results and underscores the importance of these traits in selection.

The leaf trichome density on the abaxial surface showed a significant positive correlation with gossypol content (0.72), phenol content ($r = 0.51$), tannin content (0.47) and trichome length (0.29) confirming its role as a key morphological trait linked to biochemical defense pathways. Boll weight showed a significant positive correlation with plant height (0.24), ginning outturn (0.22), seed index (0.53), lint index (0.46) and lint yield (0.61). The strong positive correlation of trichomes with gossypol content indicates a coordinated defense mechanism against insect pests, consistent with findings by Conzemius *et al.* (2023) and Subhashini *et al.* (2025). Similarly, boll weight showed positive associations with seed index, lint index, and lint yield, suggesting its utility as a reliable yield component trait (Li *et al.*, 2022).

Collectively, the results demonstrate that integrating correlation and PCA provides valuable insight into the complex interaction of morphological, biochemical and yield attributing traits. Yield improvement in cotton can be achieved by simultaneously targeting lint yield, boll weight, and seed index, while resistance can be strengthened by selecting for trichome density and associated biochemical traits. The identified divergent clusters highlight promising genetic resources for developing high-yielding, pest-resilient cultivars. The study highlights that seed cotton yield is primarily influenced by lint yield, boll weight, and seed index, while trichome density and associated biochemical traits strengthen pest resistance. PCA confirmed the independence of yield and defense pathways, offering opportunities for simultaneous improvement. Divergent genotype clusters provide valuable parental resources for breeding high-yielding, pest-resilient cotton cultivars.

ACKNOWLEDGEMENT

The authors are grateful to the Department of Genetics and Plant Breeding, College of Agriculture, CCS Haryana Agricultural University, Hisar, Haryana, for providing the essential facilities during the experiment.

REFERENCES

- Anjani, A., Padma, V., Ramana, J.V. and Satish, Y. 2020. Evaluation of genetic parameters of agro-morpho-quality traits in American cotton (*Gossypium hirsutum* L.). *Electronic Journal of Plant Breeding*, **11**:279-282. [\[Cross Ref\]](#)
- Annual Report (2024-25). ICAR-All India Coordinated Research Project on Cotton. ICAR, Nagpur.
- Asha, R., Ahamed, M.L., Babu, D.R. and Kumar, P.A. 2013. Multivariate analysis in upland cotton (*Gossypium hirsutum* L.). *Madras Agricultural Journal*, **100** (4-6): 333-335.
- Conzemius, S.R., Reay-Jones, F.P., Greene, J.K., Campbell, B.T., Reisig, D.D., Wang, H. and Bridges, W.C. 2023. Field screening of wild cotton, *Gossypium hirsutum*, landraces for resistance to thrips (Thysanoptera: Thripidae). *Crop Protection*, **163**. [\[Cross Ref\]](#)
- Ding, M., Ye, W., Lin, L., He, S., Du, X., Chen, A., Cao, Y., Qin, Y., Yang, F., Jiang, Y., Zhang, H., Wang, X., Paterson, A.H. and Rong, J. 2015. The hairless stem phenotype of cotton (*Gossypium barbadense*) is linked to a copia-like retrotransposon insertion in a homeodomain-leucine zipper gene (HD1). *Genetics*, **201** (1):143-154. [\[Cross Ref\]](#)
- Isong, A., Balu, P.A. and Ramakrishnan, P. 2017. Association and principal component analysis of yield and its components in cultivated cotton. *Electronic Journal of Plant Breeding*, **8** (3):857-864. [\[Cross Ref\]](#)
- Johnson, H.W., Robinson, H.F. and Comstock, R.E. 1955. Estimates of genetic and environmental variability in soybeans. *Agronomy Journal*, **47** (7):314-318. [\[Cross Ref\]](#)
- Joshi, B., Singh, S., Tiwari, G.J., Kumar, H., Boopathi, N.M., Jaiswal, S., Adhikari, D., Kumar, D., Sawant, S.V., Iquebal, M.A. and Jena, S.N. 2023. Genome-wide association study of fiber yield-related traits uncovers the novel genomic regions and candidate genes in Indian upland cotton (*Gossypium hirsutum* L.). *Frontiers in Plant Science*, **14**:1-22. [\[Cross Ref\]](#)
- Komala, M., Ganesan, N.M. and Kumar, M. 2018. Genetic variability, heritability and correlation analysis in F2 populations of ratoon upland cotton hybrids. *International Journal of Agriculture, Environment and Biotechnology*, **11** (6):815-827. [\[Cross Ref\]](#)

- Kusugal, A., Patil, S.S., Pranesh, K.J., Rajeev, S. and Reddy, H.K. 2014. Improving genetic base of *Gossypium barbadense* for developing potential inter specific hybrids. *Electronic Journal of Plant Breeding*, **5**:594- 601.
- Li, Z., Liu, S., Conaty, W., Zhu, Q.H., Moncuquet, P., Stiller, W. and Wilson, I. 2022. Genomic prediction of cotton fibre quality and yield traits using Bayesian regression methods. *Heredity*, **129**:103-112. [\[Cross Ref\]](#)
- Liang, Y., Zheng, J., Wang, J., Gong, Z., Li, Z., Ling, M., Zhang, Z., Sang, Z., Qu, Y., Li, X. and Chen, Q. 2023. Genetic diversity, population structure, and genome-wide association study of seven agronomic traits in 273 diverse upland cotton genotypes. *Phyton*, **92**(12):33-45. [\[Cross Ref\]](#)
- Luo, H., Chen, G., Shao, D., Gao, X., Yuan, C., Peng, J. and Tang, F. 2019. Genotypic differences in sucrose metabolism within cotton bolls in relation to lint percentage. *Field Crops Research*, **236**:33-41. [\[Cross Ref\]](#)
- Malik, W., Ashraf, J., Iqbal, M.Z., Khan, A.A., Qayyum, A., Abid, M.A., Noor, E., Ahmad, M.Q. and Abbasi, G.H. 2014. Molecular markers and cotton genetic improvement: current status and future prospects. *The Scientific World Journal*, 1-15. [\[Cross Ref\]](#)
- Naik, B.M., Satish, Y. and Babu, D.R. 2016. Genetic diversity analysis in American cotton (*Gossypium hirsutum* L.). *Electronic Journal of Plant Breeding*, **7** (4):1002-1006. [\[Cross Ref\]](#)
- Navingautham, N.S., Vishnuvardhan, K.M., Sabitha, N., Naidu, G.M. and Sekhar, M.R. 2026. Multivariate analysis for earliness, yield and fibre quality traits in cotton using correlation, PCA and cluster analysis. *Electronic Journal of Plant Breeding*, **17** (2):188-197. [\[Cross Ref\]](#)
- Panase, V.G. 1957. Genetics of quantitative characters in relation to plant breeding. *Indian Journal of Genetics and Plant Breeding*, **17**: 318–328.
- Rao, P.J.M. and Gopinath, M. 2013. Variability and association studies for yield and yield components in upland cotton (*Gossypium hirsutum* L.) under red chalka soils. *Electronic Journal of Plant Breeding*, 1093-1096.
- Rathinavel, K. 2017. Exploration of genetic diversity for qualitative traits among the extant upland cotton (*Gossypium hirsutum* L.) varieties and parental lines. *International Journal of Current Microbiology and Applied Sciences*, **6** (8):2407-2421. [\[Cross Ref\]](#)
- Reddy, K.B., Reddy, V.C., Naidu, T.C. and Srinivasarao, V. 2015. Multivariate analysis in upland cotton (*Gossypium hirsutum* L.). *Electronic Journal of Plant Breeding*, **6** (4):1019-1026.
- R Studio. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.
- Sagar, U.N., Reddy, B.V., Vishnuvardhan, K.M., Prasanna, R.A., Rao, E.S.V., Sabitha N., Kalyani D.L., Reddy Y.R. and Suresh G. 2023. Genetic analysis of fibre quality and yield related traits in desi cotton (*Gossypium arboreum* L.). *Electronic Journal of Plant Breeding*, **14** (3): 833-840. [\[Cross Ref\]](#)
- Sahar, A., Zafar, M.M., Razzaq, A., Manan, A., Haroon, M., Sajid, S., Abdul, R., Huijuan, M., Muhammad, A., Maozhi, R., Amir, S. and Yuan, Y. 2021. Genetic variability for yield and fiber-related traits in genetically modified cotton. *Journal of Cotton Research*, **4**(1):19. [\[Cross Ref\]](#)
- Sarwar, G., Nazir, A., Rizwan, M., Shahzadi, E. and Mahmood, A. 2021. Genetic diversity among cotton genotypes for earliness, yield and fibre quality traits using correlation, principal component and cluster analyses. *Sarhad Journal of Agriculture*, **37** (1):307-314. [\[Cross Ref\]](#)
- Sen, E.Z.A., El-Dahan, M.A., Badawy, S.A., Katta, Y.S., Aljuaid, B.S., El-Shehawi, A.M., El-Saadony, M.T. and El-Tahan, A.M. 2022. Evaluation of genetic behavior of some Egyptian cotton genotypes for tolerance to water stress conditions. *Saudi Journal of Biological Sciences*, **29** (3):1611-1617. [\[Cross Ref\]](#)
- Shakeel, A., Talib, I., Rahid, M., Saeed, A., Zia, K. and Saleem, M.F. 2015. Genetic diversity among upland cotton genotypes for quality and yield-related traits. *Pakistan Journal of Agricultural Sciences*, **52** (1):73-77.
- Sheidai, M. 2008. Cytogenetic distinctiveness of sixty-six tetraploid cotton (*Gossypium hirsutum* L.) cultivars based on meiotic data. *Acta Botanica Croatica*, **67** (2):209-220.
- Shim, J., Mangat, P.K. and Angeles-Shim, R.B., 2018. Natural variation in wild *Gossypium* species as a tool to broaden the genetic base of cultivated cotton. *Journal of Plant Science Current Research*, **2**: 005.
- Subhashini, S., Keerthivarman, K., Rajeswari, S., Premalatha, N., Kalaimagal, T., Muthuswami, M. and Jeyakumar, P. 2025. Molecular, cytological and morphological studies on Jassid resistance in cotton (*Gossypium hirsutum* L.) based on hairiness trait. *Journal of Cotton Research*, **8** (1):15. [\[Cross Ref\]](#)

- Sun, Z., Xingfen, W., Liu, Z., Gu, Q., Yan, Z., Li, Z., Ke, H., Jun, Y., Wu, J., Wu, L., Zhang, G. and Ma, Z. 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. [\[Cross Ref\]](#)
- Ullah, A., Shakeel, A., Ahmed, H.G.M.D., Naeem, M., Ali, M., Shah, A.N. and Hasan, M.E. 2022. Genetic basis and principal component analysis in cotton (*Gossypium hirsutum* L.) grown under water deficit condition. *Frontiers in Plant Science*, **13**:981369. [\[Cross Ref\]](#)
- Yue, Z., Tseng, T.M., Reddy, K.R., Krishnan, N. and Lu, S. 2024. Utilization of secondary metabolites in cotton production. In: *Best Crop Management and Processing Practices for Sustainable Cotton Production*. Intech Open. [\[Cross Ref\]](#)