The *Journal of Cotton Science 29:52–58* (*2025*)
http://journal.cotton.org, © The Cotton Foundation 2025

WCRC-8 SPECIAL ISSUE
BREEDING, GENETICS, & GENOMICS

Unleashing Bt Cotton Hybrids Potential Through Deciphering Yield Traits Using Principal Components and Correlation Studies

Sunayana Punia* and Manpreet Singh

ABSTRACT

India, the largest cotton producer globally, is experiencing stagnation in productivity. Enhancing productivity necessitates selecting high-yielding hybrids based on genetic relationships among traits and principal component analysis (PCA). PCA is a statistical method that simplifies complex data by identifying key contributors to total variability, aiding hybrid selection. However, cotton yield is a complex trait influenced by multiple genes. Understanding the interaction and contribution of each trait is crucial for effective selection and yield improvement. This study examined the relationship between yield components in 45 Bt cotton hybrids for three years (2021-2023) in Abohar, Punjab. Pearson correlation coefficient and PCA were employed to analyze data. Most yield components exhibited positive correlations except for days to 50% flowering. Plant height, node ratio, number of nodes per plant, monopodial and sympodial branches per plant, bolls per plant, and boll weight showed significant positive correlations with seedcotton yield per plant. Among 10 principal components identified, three had eigenvalues exceeding one, accounting for 81.93% of total variability. The principal component with the highest variability was primarily associated with plant height, nodes per plant, seedcotton yield, number of sympodial branches, number of bolls per plant, and boll weight. Findings highlight the importance of considering multiple yield components and their interrelationships to enhance cotton productivity. Specifically, plant height, node ratio, number of nodes, monopodial and sympodial branches, bolls per plant, and boll weight were identified as key traits influencing seedcotton yield, providing valuable insights for targeted breeding efforts.

Cotton (*Gossypium hirsutum* L.), a globally significant crop accounting for 90% of worldwide production, holds a unique position in India's agricultural landscape. India occupies the first position in cotton area (37% of the world's acreage) and as the largest cotton producer (22% of world cotton production) (Cotton Corporation of India, 2023). Despite this dominance, productivity has stagnated at approximately 447 kg/ha in recent decades. This stagnation poses a significant challenge to meet the growing demand for cotton fiber and ensure the economic viability of cotton cultivation in the country. Various factors contribute to this yield plateau, including biotic stresses such as pests and diseases, abiotic stresses such as drought and heat, and the inherent genetic potential of existing cultivars. To address this productivity plateau, a multifaceted approach is required, encompassing advancements in agronomic practices, pest and disease management, and genetic improvement. Among these, genetic enhancement through the development of high-yielding cultivars holds immense potential to revolutionize cotton production in India. In the quest to develop cotton cultivars capable of flourishing in diverse environments and withstanding various stresses, pinpointing superior genotypes possessing desirable yield-related traits is paramount. Principal component analysis (PCA) is an invaluable statistical tool in this pursuit. Recognizing the need to leverage PCA for showcasing the outcomes of cotton breeding experiments, numerous researchers have employed this technique to evaluate the relationships and diversity within various cotton germplasms Thus, many researchers (Abasianyanga et al., 2017; Abdel-Monaem et al., 2020; Nandhini et al., 2018; Shah et al., 2018; Vinodhana and Gunasekaran, 2019; Yehia and El-Hashash, 2021) have used PCA to assess the relationship and diversity between several cotton germplasms. Moreover, PCA has been instrumental in unraveling the intricate connections between yield and its constituent components.

S. Punia* and M. Singh, Punjab Agricultural University Regional Research Station, Abohar, Punjab, India, 152116.
*Corresponding author: sunayanapunia19@pau.edu

By simplifying complex datasets and identifying key contributors to yield variability, PCA facilitates the selection of superior parental lines for breeding programs (Isong et al., 2017). Understanding the genetic relationships among yield-related traits is paramount, as cotton yield is a complex trait influenced by multiple genes. These traits encompass various aspects of plant growth and development, including flowering time, plant architecture, boll formation, and fiber quality. This study aimed to unravel the character association in seedcotton yield (SCY) and its attributing traits in cotton hybrids, employing PCA to dissect the intricate interplay of factors contributing to yield. By identifying the most influential traits and their interrelationships, breeders can make informed decisions in selecting the hybrids for consistent performance. The insights gained from this research will pave the way for targeted breeding strategies, ultimately aiming to break the yield plateau and bolster cotton productivity in India. This, in turn, will have far-reaching implications for the livelihoods of millions of farmers who depend on cotton cultivation and contribute significantly to the country's economy. Additionally, the identification of key yield-related traits and their genetic associations will provide valuable information for developing molecular markers and genomic selection tools, further accelerating the breeding process and enhancing the efficiency of hybrid development.

MATERIALS AND METHODS

The research was conducted across three Kharif seasons (2021-2023) at the Punjab Agricultural University's Regional Research Station in Abohar. The experiment was executed with 45 cotton hybrids (Table 1) in an alpha lattice along with two replications. Each genotype was sown in four rows in each replication with a row length of 6 m and row width of 67.5 cm. Recommended agronomic practices were maintained throughout the crop season. For data collection, five plants were randomly chosen from each genotype in every replication. The following yield and yield-related traits were recorded: days to 50% flowering (DT50; average number of days from sowing to the appearance of the first flower on the selected plants), plant height (measured in cm from cotyledonary node to the tip of the main stem), number of nodes per plant (total number of nodes counted), plant height to node ratio (PHNR; calculated), number of monopodial branches per plant (lateral and axillary branches on the main stem with vertical, upward growth in acropetal succession), number of sympodial branches per plant (fruiting branches), number of bolls per plant (fully opened bolls were counted and averaged), boll weight (average weight of five well opened bolls in g), SCY per plant (open bolls of five competitive plants were picked in separate cloth bags and average weight was recorded for SCY per plant and converted in kg/ha), and ginning outturn (percentage ratio of lint weight to seed cotton weight). Pearson correlation analysis and PCA using Pearson correlation were performed on the collected data using the metan package in R software.

RESULTS AND DISCUSSION

Pearson correlation analysis (Table 2) revealed a complex network of relationships among yield components and SCY in Bt cotton hybrids. As expected, several traits exhibited strong positive correlations with SCY. Plant height (r = 0.791), number of nodes per plant (r = 0.774), number of sympodial branches per plant (r = 0.761), and boll weight (r = 0.697) showed significant positive associations with SCY, confirming their importance as yield-determining factors in cotton (Amein et al., 2020; Chaudhari et al., 2017; Manonmani et al., 2019; Memon et al., 2022; Satish et al., 2020).

The strongest positive correlation with SCY was observed for the number of bolls per plant (r = 0.863). This aligns with Kumbhar et al. (2020), who emphasized the significant role of boll number in determining upland cotton yield. Interestingly, Table 2 reveals that boll number is also positively correlated with several other traits, including plant height, number of nodes, and number of sympodial branches. This suggests that these traits might indirectly contribute to increased yield by promoting higher boll production, as indicated by Jarwar et al. (2019). Although monopodial branches showed a positive correlation with SCY (r = 0.489), their contribution appears less pronounced compared to sympodial branches. This might indicate that the architecture of sympodial branches is more conducive to boll development and yield in these cotton hybrids. Further investigation into the specific roles of monopodial and sympodial branches in yield determination could be valuable.

As anticipated, DT50 had a negative correlation (r = -0.562) with SCY. This confirms that earlier flowering genotypes tend to have higher yields, supporting the findings of Jangid et al. (2022) in desi cotton. The

Table 1. List of Hybrids

No.	Hybrid Name	No.	Hybrid Name	No.	Hybrid Name
1	ABCH 252	16	KSCH 207	31	RCH 653
2	ABCH 254	17	MH 5302	32	RCH 773
3	ABCH 4899	18	MRC 7041	33	RCH 776
4	ACH 133-2	19	MRC 7361	34	SHAKTI 9
5	ACH 155-2	20	MRC 7365	35	SOLAR 56
6	ACH 177-2	21	NCS 495	36	SOLAR 65
7	ACH 33-2	22	NCS 855 BG II	37	SOLAR 75
8	ANKUR 3224	23	NCS 9013	38	SOLAR 77
9	ANKUR 3228	24	NCS 9024	39	SUPER 5
10	ANKUR 3244	25	PCH 225	40	SUPER 544
11	KCH 172	26	PCH 877	41	SUPER 721
12	KCH 999 BG II	27	PCH 9611	42	SWCH 4711
13	KDCH 441	28	PRCH 333 BG II	43	SWCH 4748
14	KDCH 621	29	RCH 314 BG II	44	VICH 308
15	KDCH 641	30	RCH 650	45	VICH 310 BG II

Table 2. Pearson Correlation Coefficients among yield and its contributing traits cotton hybrids.

	DT50	PH	PHNR	NPP	MPP	SPP	BW	BPP	SCY	GOT
DT50^{z}	1**	-0.463**	-0.184*	-0.441**	-0.093	-0.468**	-0.778**	-0.416**	-0.562**	0.316**
PH		1**	0.513**	0.874**	0.549**	0.859**	0.679**	0.686**	0.791**	0.104
PHNR			1**	0.041	0.049	0.035	0.324**	0.155	0.275**	-0.027
NPP				1**	0.598**	0.989**	0.611**	0.722**	0.774**	0.132
MPP					1**	0.471**	0.299**	0.488**	0.489**	0.321**
SPP						1**	0.616**	0.703**	0.761**	0.085
BW							1**	0.479**	0.697**	-0.163
BPP								1**	0.863**	0.104
SCY									1**	0.042
GOT										1**

**Significant at 0.01 level. *Significant at 0.05 level.

^{z}DT50: days to 50% flowering, PH: plant height, PHNR: plant height to node ratio, NPP: number of nodes per plant, MPP: number of monopodial branches per plant, SPP: number of sympodial branches per plant, BW: boll weight, BPP: number of bolls per plant, SCY: seedcotton yield and GOT: ginning outturn.

negative correlation between DT50 and boll weight (r =-0.778) further suggests that early flowering might be associated with a longer boll development period, potentially leading to heavier bolls. Interestingly, PHNR showed a relatively weak correlation with SCY (r = 0.275) despite the strong positive correlation between plant height and SCY. This implies that simply increasing plant height might not necessarily translate to higher yields unless accompanied by a proportional increase in node number. This detailed analysis of the correlation matrix provides valuable insights for cotton breeders. By understanding the complex interplay between yield components, breeders can make informed decisions to optimize selection strategies. Prioritizing traits with strong positive correlations to SCY, such as boll number, plant height, nodes per plant, and sympodial branches, while also considering earliness (DT50), can lead to the development of high-yielding cotton varieties.

Principal Component Analysis. Principal component (PC) analysis was conducted to explore the pattern of variability and potential for grouping within the cotton hybrid population. This analysis yielded 10 PCs, with the first three having eigenvalues greater than one, indicating their substantial contribution to the overall variation. These three PCs collectively accounted for

81.93% (Table 3) of the total variability observed in yield and other measured traits. This aligns with previous research in cotton, where the first few PCs have been shown to capture a majority of the total variance (Abasianyanga et al., 2017; Abdel-Monaem et al., 2020; Shah et al., 2018). For example, Abasianyanga et al. (2017) reported that the first five PCs accounted for more than 70% of the cumulative variance in their study of cotton yield components.

For further analysis, we focused on the first two PCs, which together explained 71% of the total variability, exceeding the explanatory power of any single trait. This approach allows for a more comprehensive understanding of the data's variability and aids in selecting traits with positive loading factors. PC1, with the largest contribution (54.95%), was primarily associated with plant architecture and boll production traits, including plant height, nodes per plant, SCY, number of sympodial branches, number of bolls per plant, and boll weight. This is visually confirmed by the Circos diagram (Fig. 1), which shows strong connections between PC1 and these traits. PC2 explained 16.17% of the variability and was mainly linked to ginning outturn, DT50, monopodial branches per plant, and boll weight. PC3, explained 10.8% of the variation, and was associated with the PHNR (Fig. 1).

These results suggest that PC1 and PC2 can effectively summarize the original variables and provide insights into the overall variance and grouping within the dataset. As such, we used PC1 and PC2 to construct a biplot (Fig. 2) for visualization and further interpretation following the approach used in other cotton research studies (Abasianyanga et al., 2017; Nandhini et al., 2018; Yehia and El-Hashash, 2021). The biplot visually represents the variables as vectors, with the angles between them indicating correlations. Acute angles signify positive correlations, whereas obtuse angles represent negative correlations. The biplot confirms the strong positive correlations observed among SCY, sympodial branches per plant, nodes per plant, bolls per plant, boll weight, plant height, and PHNR. Conversely, a negative correlation is evident between DT50 and SCY. Importantly, the biplot reveals a distinct and non-overlapping clustering of genotypes based on the environments (E1, E2, and E3), which represent the years 2021, 2022, and 2023, respectively. The traits demonstrated varying levels of contribution to genetic variability across different years, indicating that environmental conditions influence trait expression and genotype performance. The year 2021 (E1) provided optimum growing conditions resulting in the increased productivity across all Bt cotton hybrids. The traits associated with PC1 had a greater influence on variability in this environment. In contrast, 2022 (E2) was marked by the incidence of whitefly infestation and cotton leaf curl disease (CLCuD), which reduced yield levels and created considerable variation in the data, likely due to the differential responses of the hybrids to these biotic stresses. The wider spread of genotypes in E2 (2022) likely reflects the differential responses of the hybrids to whitefly and CLCuD infestation, with some genotypes potentially exhibiting greater resistance or tolerance. Both PC1 and PC2 contributed to genetic variability in the year 2023 (E3). Although E3 experienced favorable climatic conditions, the incidence of pink bollworm [*Pectinophora gossypiella* (Saunders)] infestations led to reduced yield levels in some hybrids, further contributing to the

Table 3. Estimates of Principal Component Analysis (PCA) in cotton hybrids.

S. No.	Principal Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
			%	%
1	PC1	5.4	54.9	54.9
2	PC2	1.6	16.2	71.1
3	PC3	1.1	10.8	81.9
4	PC4	0.5	5.4	87.4
5	PC5	0.5	5.1	92.5
6	PC6	0.4	4.8	97.3
7	PC7	0.1	1.8	99.1
8	PC8	0.0	0.9	99.9
9	PC9	0.0	0.0	100.0
10	PC10	0.0	0.0	100.0

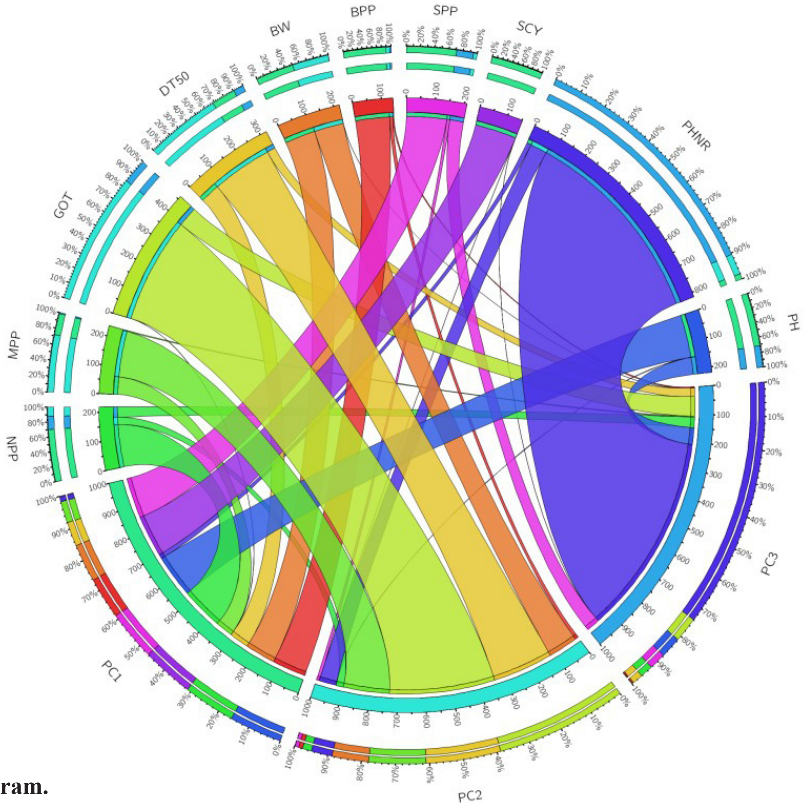

Figure 1. Circos Diagram.

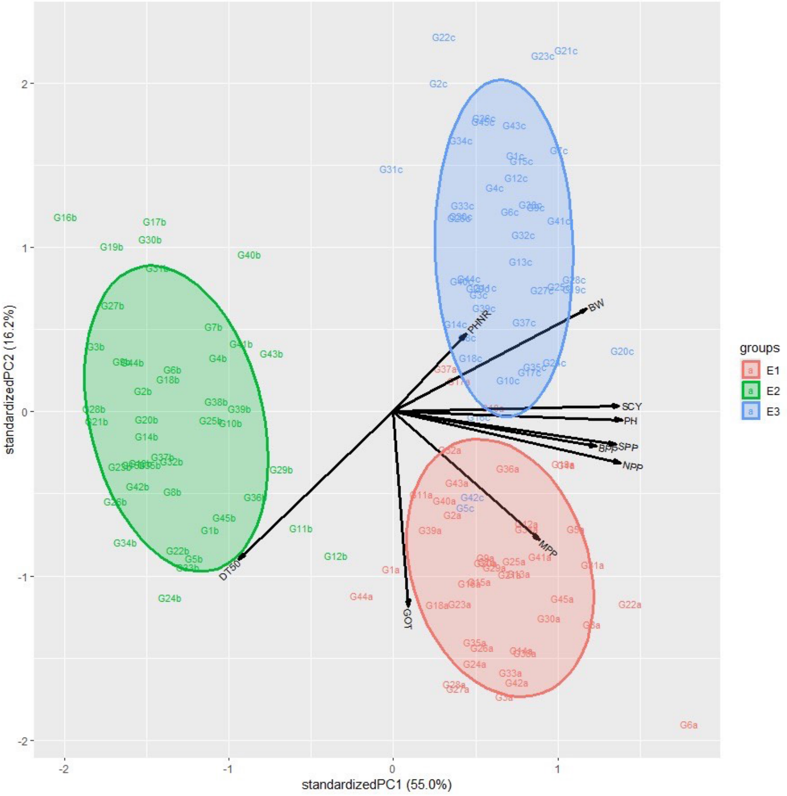

Figure 2. Biplot diagram between PC1 and PC2 shows similarity and dissimilarity relationships between the studied traits in the 45 hybrids.

observed variation in the dataset. This underscores the importance of breeding for resistance to biotic stresses, as they can significantly impact yield stability across years. Furthermore, the distinct clustering patterns emphasize the need to consider genotype × environment interactions in breeding programs, as the optimal combination of traits for high yield might vary depending on the presence or absence of specific biotic stresses. This environmental clustering has important implications for breeding. Although the correlation analysis revealed significant positive associations between traits such as SCY and sympodial branches per plant, the biplot suggests that these correlations might differ across years. A trait combination beneficial in 2021 (E1) might not be as effective in 2022 (E2) or 2023 (E3). This highlights the need to consider genotype × environment interactions in cotton breeding programs.

The position of individual genotypes within each cluster further highlights how specific genotypes responded differently to each year's conditions. Some genotypes might perform consistently across all years, whereas others might show a strong preference for a particular year. This analysis can help identify genotypes that are stable across environments (those closer to the center of the plot) and those highly adapted to specific years (those located at the edges of the clusters). Based on the biplot, genotypes such as Solar 75, Ankur 3244, and MH 5302, which are located closer to the center and exhibit less variation across environments, can be considered relatively stable across the three years. PCA analysis, particularly biplot visualization, serves as a good predictor for separating genotypes into groups based on their performance and environmental responses. This is consistent with previous findings in cotton (Nandhini et al., 2018; Rathinavel, 2019; Yehia and El-Hashash, 2021) where PCA effectively captured the extent of variation in traits between genotypes. The traits strongly associated with PC1 and PC2, such as plant height, nodes per plant, sympodial branches per plant, boll number, boll weight, and DT50, offer valuable insights for selection and improvement strategies in cotton breeding programs. By considering these key traits and their interactions with environmental conditions, breeders can develop superior genotypes with enhanced yield potential and stability across varying environments.

CONCLUSIONS

This study provides a comprehensive analysis of yield-determining factors in Bt cotton hybrids, revealing a complex interplay between genetics and environment. By integrating correlation and PCA, we demonstrated the significant influence of key architectural and boll-related traits including plant height, nodes, sympodial branches, and boll number, as major drivers of yield variation, highlighting the importance of direct selection for these traits in breeding programs. Furthermore, biplot visualization revealed distinct genotype clustering patterns across three years with contrasting environmental conditions underscoring the significant impact of biotic stresses (whitefly, CLCuD, and pink bollworm) and genotype × environment interactions. Based on the biplot, genotypes such as Solar 75, Ankur 3244, and MH 5302 can be considered relatively stable across the three years but multilocation analysis could provide a clear view about their performance. These findings provide critical insights for cotton breeders, suggesting direct selection for key traits, considering earliness and environmental stability for developing high yielding and resilient cultivars that can thrive in the face of diverse environmental challenges and contribute to increased and sustainable cotton production in India.

REFERENCES

Abasianyanga, I., P.A. Balu, and P. Ramakrishnan. 2017. Association and principal component analysis of yield and its components in cultivated cotton. Electronic J. Plant Breeding. 8(3):857–864. https://doi.org/10.5958/0975-928X.2017.00140.5

Abdel-Monaem, M.A., W.A.E. Abido, Á. Hadházy, M.H. Ghoneima, Y.M. EL-Mansy, and M.W. EL-Shazly. 2020. Genetic divergence among Egyptian cotton genotypes under water deficit conditions. Acta Ecologica Sinica. 42(2):11–18. https://doi.org/10.1016/j.chnaes.2020.11.007

Amein, M.M.M., M.I. Masri, H.H. EL-Adly, and S.S. Attia. 2020. Correlation and path coefficient analysis for yield components traits in Egyptian cotton genotypes (*Gossypium barbadense* L.). Plant Arch. 20(2):803–806.

Chaudhari, M., G. Faldu, and H.J.A.I.B. Ramani. 2017. Genetic variability, correlation and path coefficient analysis in cotton (*Gossypium hirsutum* L.). Adv. Biomed. Res. 8(6):226–233.

Cotton Corporation of India. 2023. Area, production and productivity of cotton in India. Available online at https://cotcorp.org.in/statistics.aspx (verified 20 Nov 2024).

Isong, A., P.A. Balu, and P. Ramakrishnan. 2017. Association and principal component analysis of yield and its components in cultivated cotton. Electronic J. Plant Breeding. 8(3):857–64. https://doi.org/10.5958/0975-928X.2017.00140.5

Jangid, K., O. Sangwan, C.M. Sagar, and P. Kumar. 2022. Correlation and path analysis for seed cotton yield and its contributing traits under irrigated conditions of Sirsa in desi cotton. J. Agrisearch. 9(2):129–132.

Jarwar, A.H., X. Wang, M.S. Iqbal, Z. Sarfraz, L. Wang, Q. Ma, and F. Shuli. 2019. Genetic divergence on the basis of principal component, correlation and cluster analysis of yield and quality traits in cotton cultivars. Pakistan J. Bot. 51(3):1143–1148.

Kumbhar, Z.M., A.J. Wajid, J.K. Sootaher, M.I. Baloch, A.A. Gadahi, K.K. Menghwar, M.S. Chang, and M. Kachi. 2020. Studies on correlation and heritability estimates in upland cotton (*Gossypium hirsutum* L.) genotypes under the agro-climatic conditions of Tandojam, Sindh, Pakistan. Pure Appl. Biol. 9(4):2272–2278. https://doi.org/10.19045/bspab.2020.90241

Manonmani, K., L. Mahalingam, D. Malarvizhi, N.S Ritharan, and N. Premalatha. 2019. Genetic variability, correlation and path analysis for seed cotton yield improvement in upland cotton (*Gossypium hirsutum* L.). J. Pharmacognosy Phytochem. 8(4):1358–1361.

Memon,S., A.W.B.N. Gandahi, T.A. Yasir, S.M. Sarki, N.S. Wasaya, T.M. Khan, M.A. Iqbal, R. Ullah, M.A. Bhutta, S. Ahmad, and M.Z. Ishaq. 2022. Assessment of yield contributing quantitative traits in upland cotton (*Gossypium hirsutum* L). Sarhad J. Agric. 38(1):353–359. https://doi.org/10.17582/journal.sja/2022/38.1.353.359

Nandhini, K., P.A. Balu, and A. Isong. 2018. Genetic analysis and inheritance studies in f^2 population of upland cotton (*G. hirsutum* L.). Internat. J. Pure Appl. Biosci. 6(2):1499–1505.

Rathinavel, K. 2019. Agro-morphological characterization and genetic diversity analysis of cotton germplasm (*Gossypium hirsutum* L.). Intern. J. Curr. Microbiol. Appl. Sci. 8(2):2039–2057. https://doi.org/10.20546/ijcmas.2019.802.237

Satish, Y., M.S. Rani, and R. Chapara. 2020. Correlation and path co-efficient analysis for yield and yield component traits in upland cotton (*Gossypium hirsutum* L.). Internat. J. Chem. Stud. 8(2):2742–2746. https://doi.org/10.22271/chemi.2020.v8.i2ap.9166

Shah, S.A.I., S.J. Khan, K. Ullahand, and O.U. Sayal. 2018. Genetic diversity in cotton germplasm using multivariate analysis. Sarhad J. Agric. 34:130–135. https://doi.org/10.17582/journal.sja/2018/34.1.130.135

Vinodhana, N.K., and P. Gunasekaran. 2019. Evaluation of genetic diversity in cotton (*Gossypium barbadense* L.) germplasm for yield and fibre attributes by principal component analysis. Internat. J. Curr. Microbiol. Appl. Sci. 8(4):2614–2621. https://doi.org/10.20546/ijcmas.2019.804.304

Yehia, W.M.B., and E.F. El-Hashash. 2021. Correlation and multivariate analysis across non-segregation and segregation generations in two cotton crosses. Egyptian J. Agric. Res. 99(4):380–390. https://doi.org/10.21608/ejar.2021.81571.1117