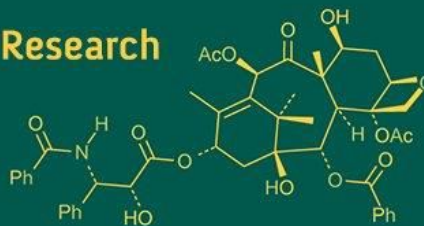
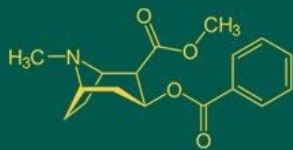


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2025): 5.29
IJABR 2025; SP-9(8): 122-127
www.biochemjournal.com
Received: 26-06-2025
Accepted: 30-07-2025

Cheerla Pavan
M.Sc Scholar, Genetics and
Plant Breeding, Agricultural
College, Bapatla, Andhra
Pradesh, India

Rani Chapara
Scientist, (Genetics and Plant
Breeding), Regional
Agricultural Research Station,
ANGRAU, Lam, Guntur,
Andhra Pradesh, India

M Sudha Rani
Principal Scientist, (Genetics
and Plant Breeding), Regional
Agricultural Research Station,
ANGRAU, Maruteru, Andhra
Pradesh, India

GA Diana Grace
Scientists, (Entomology),
Regional Agricultural
Research Station, ANGRAU
Lam, Guntur, Andhra
Pradesh, India

James M
Research Associate, (Genetics
and Plant Breeding), Regional
Agricultural Research Station,
ANGRAU, Lam, Guntur,
Andhra Pradesh, India

Priyadharshini
Data Analyst, (Agronomy),
Regional Agricultural
Research Station, ANGRAU
Lam, Guntur, Andhra
Pradesh, India

Corresponding Author:
Cheerla Pavan
M.Sc Scholar, Genetics and
Plant Breeding, Agricultural
College, Bapatla, Andhra
Pradesh, India

Unraveling genetic divergence using principal component analysis for yield and fibre quality traits in cotton genotypes

Cheerla Pavan, Rani Chapara, M Sudha Rani, GA Diana Grace, James M and Priyadharshini

DOI: <https://www.doi.org/10.33545/26174693.2025.v9.i8Sb.5099>

Abstract

Cotton (*Gossypium* spp.) is the most widely used natural fibre in the textile industry. Understanding genetic diversity is essential for effective genetic improvement. Principal Component Analysis (PCA) was conducted on 53 cotton genotypes and evaluated using an Augmented Block Design at the Regional Agricultural Research Station, Lam Farm, Guntur. PCA identified four principal components with eigenvalues greater than 1, which were retained for subsequent analysis. The four principal components contributed 70.56% towards variability. Biplot analysis identified GP 49 (L 2387), GP 34 (L 2396) and GP 40 (L 2281) as superior for seed index, lint index and seed cotton yield per plant with lower plant height, making them ideal for yield improvement. The PCA biplot analysis also revealed that GP 49 (L 2387), GP 34 (L 2396), GP 24 (L 2381), GP 25 (L 2267), GP 29 (Yaganti), GP 30 (Srinandi), GP 31 (NDLA 3116-4), GP 32 (Suvin), GP 33 (CCB 29), GP 37 (L 2275) and GP 45 (L 2385) were the most divergent genotypes as they were located far away from the origin and should be tested for their combining ability through Line x Tester mating design or diallel analysis and then could be utilized in future crop improvement programs under heterosis breeding.

Keywords: Cotton, genetic diversity, principal component analysis, seed cotton yield, fibre quality

Introduction

Cotton (*Gossypium* spp.) is a cornerstone of the global textile industry and a critical commercial crop for India. As one of the oldest cultivated plants, it's primarily grown for its fibre and oil in tropical and subtropical regions. India plays a significant role in global cotton production, contributing approximately 20.67% (ICAR-AICRP (Cotton)-Annual Report 2024-25) [5]. It is crucial for increasing the country's economy and is popularly known as "white gold" (Komala *et al.* 2018) [7]. Belonging to the genus *Gossypium* within the Malvaceae family, cotton exhibits extensive genetic diversity. This genus comprises 50 known species: four cultivated, 44 wild diploids, and two wild tetraploids (Percival and Kohel, 1990) [11].

Globally, four main cotton species are cultivated, including two diploids (*Gossypium herbaceum* and *G. arboreum*) and two New World tetraploids (*G. hirsutum* and *G. barbadense*). However, the genetic base of cultivated cotton genotypes remains narrow (McCarty *et al.* 2004; Shakeel *et al.* 2015) [9, 14]. To broaden this genetic base, especially through hybrid breeding programs, an understanding of genetic divergence among available germplasm is essential. Genetic variation among cotton genotypes with respect to morphological and fiber quality traits has been studied to facilitate their improvement (Ruilian *et al.* 2005) [8]. Researchers have employed principal component analysis (PCA) to assess similarity among genotypes based on these traits and to classify them (Sarwar *et al.* 2021) [13].

Principal Component Analysis (PCA) and cluster analysis are widely utilized to explore genetic similarities and differences across genotypes, grouping them into distinct clusters (Naik *et al.* 2016) [10]. PCA is particularly effective at partitioning total variation, making it highly suitable for identifying elite parental lines by highlighting the key traits that drive divergence. It also aids in estimating phenotypic variability and genetic diversity among existing genotypes supporting informed selection decisions.

By revealing which traits contribute most to variability, PCA provides essential insights for choosing genotypes with both agronomic and genetic significance (Isong *et al.* 2017) [6].

The main objective of the present study is to assess the genetic diversity in yield and fibre quality traits in 53 cotton genotypes and to analyze the associations among them.

Materials and Methods

Plant Material and Site Characteristics: 50 cotton genotypes along with three checks (NDLH 2051-1, Sivanandhi and CICR 23 Bt) varieties were evaluated during *khari* season 2024-2025 at experimental area of Regional Agricultural Research Station, Lam farm, Guntur, Andhra Pradesh which was present at 16° 20' North latitude and 80° 26' East longitude and an altitude of 31.5 m above mean sea level. The experiment was sown on 1st August to evaluate genotypes for yield and fibre parameters. The list of the genotypes was presented in (Table 1).

Observations Recorded: Traits like days to fifty percent flowering (plot basis), plant height (cm), number of

monopodia per plant, number of sympodia per plant, number of bolls per plant, boll weight (g), lint index (g), seed index (g), ginning out turn (%), seed cotton yield per plant (g) and for fibre quality traits like Upper Half Mean Length (UHML) (mm), uniformity index (%), micronaire (µg/inch) and tenacity (g/tex) data was recorded on plot basis.

Experimental Design: The experiment was conducted using Augmented Block Design (Federer, 1956) [4] in six blocks. In first five blocks, 12 entries (nine genotypes + three checks) and in the last block eight entries (five genotypes + three checks) were planted randomly in the blocks. Each entry was sown in two rows with a row length of 6.3 m each with a spacing of 105 x 60 cm.

Statistical Analysis: The means of 14 traits were subjected to Analysis of Variance (ANOVA) following the statistical procedure outlined by Falconer (1964) [2]. Subsequently, the mean data was used to perform Principal Component Analysis (PCA) using 'factoextra', 'factoextra' package in R studio version R.4.4.2.

Table 1: The list of the genotypes was presented

S. No.	Genotypes	Entries	S. No.	Genotypes	Entries
1	GP 1	LHBT 2203	28	GP 28	L 2273
2	GP 2	LHBT 24	29	GP 29	Yaganti
3	GP 3	LHBT 8	30	GP 30	Srinandi
4	GP 4	NDLH 3104-4	31	GP 31	NDLA 3116-4
5	GP 5	LHBT 9	32	GP 32	Suvin
6	GP 6	NDLH 2985	33	GP 33	CCB 29
7	GP 7	LHBT 26	34	GP 34	L 2396
8	GP 8	NDLH 3116-4	35	GP 35	L 2384
9	GP 9	LHBT 6	36	GP 36	L 2274
10	GP 10	LHBT 22	37	GP 37	L 2275
11	GP 11	LHBT 7	38	GP 38	L 2277
12	GP 12	ARCV 111	39	GP 39	L 2278
13	GP 13	WGCV Bt 26	40	GP 40	L 2281
14	GP 14	LHBT 2390	41	GP 41	L 2265
15	GP 15	LHBT 2392	42	GP 42	L 2386
16	GP 16	LHBT 2385	43	GP 43	L 2389
17	GP 17	LHBT 2383	44	GP 44	L 2390
18	GP 18	LHBT 2395	45	GP 45	L 2385
19	GP 19	Lam Bt 2208	46	GP 46	L 2395
20	GP 20	ARCV 99	47	GP 47	L 2388
21	GP 21	LHBT 3	48	GP 48	L 2383
22	GP 22	Lam Bt 2209	49	GP 49	L 2387
23	GP 23	LHBT 4	50	GP 50	L2382
24	GP 24	L 2381	51	Check 1	CICR 23 Bt
25	GP 25	L 2267	52	Check 2	NDLH 2051-1
26	GP 26	L 2268	53	Check 3	Sivanandhi
27	GP 27	L 2270			

Results and Discussion

ANOVA revealed significant differences among the 53 cotton genotypes for all the traits studied, indicating the presence of substantial genetic variability and diversity within the material. Mean data of all the 14 traits were analyzed by principal component analysis in R Studio to investigate the genetic divergence among 53 cotton genotypes. The results suggested the importance of the first four PCs with eigen values greater than or equal to one in discriminating the germplasm collection. The first two PC's

were considered to represent the biplot as it was capable to explain 49.73% of the total variation. Hence, the other two PC's (PC₃ and PC₄) were ignored for biplot analysis. The results revealed that four canonical roots accounted for 70.56% of total divergence. The PC₁ contributed maximum towards divergence (26.47%) with eigen value of 3.70. The second, third and fourth canonical vectors contributed 23.25%, 11.54% and 9.28% respectively to total divergence with eigen values of 3.25, 1.61 and 1.30 respectively (Table 2).

Table 2: Canonical vectors for 14 characters in 53 cotton genotypes

Parameter	PC ₁	PC ₂	PC ₃	PC ₄
Eigen Value (Root)	3.70	3.25	1.61	1.30
Variance (%)	26.47	23.25	11.54	9.28
Cumulative Variance (%)	26.47	49.73	61.27	70.56
Characters				
Days to fifty percent flowering	12.19	5.32	3.12	8.07
Plant height (cm)	1.99	6.30	9.91	28.75
Number of monopodia per plant	7.97	0.27	17.15	5.17
Number of sympodia per plant	0.02	3.51	0.01	49.41
Number of bolls per plant	14.82	1.98	10.37	1.16
Boll weight (g)	4.43	7.38	13.44	1.22
Seed index (g)	6.77	15.36	2.86	0.29
Lint index (g)	8.29	9.91	1.78	0.13
Ginning out turn (%)	1.29	0.32	28.00	3.35
Seed cotton yield per plant (g)	11.63	9.66	0.42	0.05
Upper half mean length (mm)	10.15	12.19	2.29	0.01
Micronaire (µg/inch)	7.33	11.78	2.02	0.91
Tenacity (g/tex)	4.63	8.80	7.79	0.02
Uniformity Index (%)	8.48	7.22	0.85	1.46

In the first Principal Component (PC₁), number of bolls per plant made the largest contribution to total divergence (14.82%), followed by days to 50% flowering (12.19%), seed cotton yield per plant (11.63%), upper half mean length (10.15%), uniformity index (8.48%), lint index (8.29%) and number of monopodia per plant (7.97%). Traits with minimal influence on PC₁ included number of sympodia per plant (0.02%), ginning out turn (1.29%), and plant height (1.99%) (Table 2).

In the second Principal Component (PC₂), the trait seed index had the strongest impact (15.36%), followed by upper half mean length (12.19%), micronaire (11.78%), lint index (9.91%), seed cotton yield per plant (9.66%), tenacity (8.80%), boll weight (7.38%) and uniformity index (7.22%). The lowest contributors to PC₂ were number of monopodia per plant (0.27%) and ginning out turn (0.32%) (Table 2).

In third Principal Component (PC₃) the trait ginning out turn contributed the highest proportion to the total divergence with a percentage of 28%, followed by number of monopodia per plant, boll weight and number of bolls per plant with the proportions of 17.15%, 13.44% and 10.37% respectively. The traits, number of sympodia per plant (49.41%) and plant height (28.75%) contributed the highest proportion to the total divergence to the PC₄ respectively (Table 2).

The highest contribution for the PC₁ was observed by the genotype GP 32 (Suvini) with the proportion of 26.87% followed by the genotypes GP 33 (CCB 29), GP 24 (L 2381) and GP 25 (L 2267) with the percentage of 19.94%, 9.70% and 6.32% respectively. The lowest were contributed by the genotypes GP 31 (NDLA 3116-4), GP 9 (LHBT 6) and GP 5 (LHBT 9) respectively (Table 3).

The genotype GP 30 (Srinandi) made the highest contribution to PC₂, accounting for 17.35%, followed by GP 29 (Yaganti) (17.22%), GP 31 (NDLA 3116-4) (13.94%), GP 49 (L 2387) (4.14%) and GP 34 (L 2396) (4.11%). In

contrast, the lowest contributions were observed by genotypes GP 11 (LHBT 7), CICR 23 Bt, and GP 14 (LHBT 2390) (Table 3).

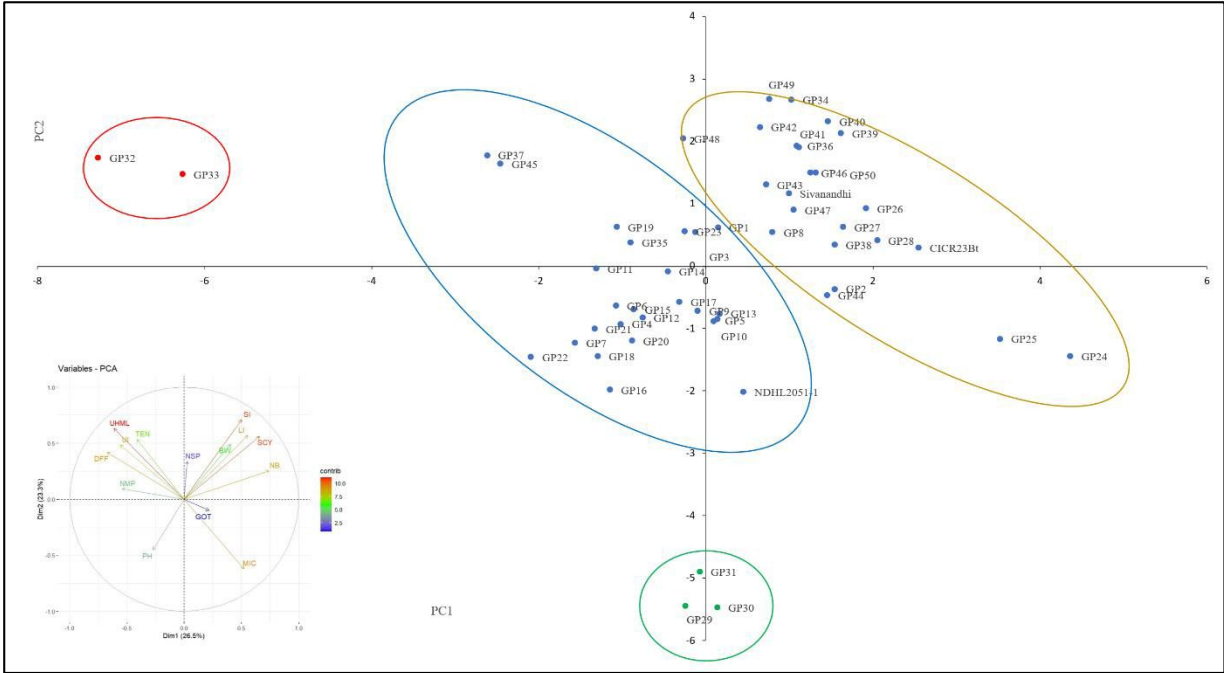
Based on the biplot analysis of the traits and individual genotypes, the genotypes GP 49 (L 2387), GP 34 (L 2396), GP 40 (L 2281) and GP 39 (L 2278) exhibited better seed index, lint index and seed cotton yield per plant and lowest plant height and these genotypes could be used for improving these traits. The genotypes GP 32 (Suvini) and GP 33 (CCB 29) showed better fibre quality traits like UHML, uniformity index, tenacity and days to fifty percent flowering coupled with lower micronaire value which was in desirable direction. These genotypes could be selected for improving the fibre quality traits. The genotypes GP 24 (L 2381), GP 25 (L 2267), GP 29 (Yaganti), GP 30 (Srinandi) and GP 31 (NDLA 3116-4) reported high micronaire value meant that the genotypes could be considered for coarse fibre genotypes (Figure 1).

The analysis thus identified that the yield contributing variables *i.e.*, number of sympodia per plant number of bolls per plant, boll weight, lint index and seed cotton yield per plant were significantly loaded in PC₁ and contributed more towards variability. The PC₂ contributed for fibre quality traits like UHML, tenacity and micronaire (Figure 2).

The PCA biplot analysis of 53 cotton genotypes revealed that GP 49 (L2387), GP 34 (L 2396), GP 24 (L 2381), GP 25 (L 2267), GP 29 (Yaganti), GP 30 (Srinandi), GP 31 (NDLA 3116-4), GP 32 (Suvini), GP 33 (CCB 29), GP 37 (L 2275) and GP 45 (L 2385) were identified as the most divergent genotypes as they were located far away from the origin. Hence, these genotypes could be effectively used in hybridization programmes for crop improvement (Figure 1). Similar findings were reported by Shakeel *et al.* (2015) [14], Farooq *et al.* (2017) [3], Sarwar *et al.* (2021) [13], Sahar *et al.* (2021) [12] and Chapara *et al.* (2022) [1].

Table 3: percent contribution of each genotype towards the first two principal components

Genotypes	PC ₁	PC ₂	Genotypes	PC ₁	PC ₂
GP 1 (LHBT 2203)	0.011	0.222	GP 28 (L 2273)	2.145	0.098
GP 2 (LHBT 24)	1.221	0.079	GP 29 (Yaganti)	0.029	17.228
GP 3 (LHBT 8)	0.007	0.171	GP 30 (Srinandi)	0.01	17.356
GP 4 (NDLH 3104-4)	0.524	0.501	GP 31 (NDLA 3116-4)	0.002	13.944
GP 5 (LHBT 9)	0.005	0.451	GP 32 (Suvin)	26.875	1.74
GP 6 (NDLH 2985)	0.587	0.238	GP 33 (CCB 29)	19.948	1.259
GP 7 (LHBT 26)	1.24	0.885	GP 34 (L 2396)	0.54	4.119
GP 8 (NDLH 3116-4)	0.328	0.17	GP 35 (L 2384)	0.408	0.082
GP 9 (LHBT 6)	0.004	0.305	GP 36 (L 2274)	0.601	2.14
GP 10 (LHBT 22)	0.011	0.422	GP 37 (L 2275)	3.481	1.825
GP 11 (LHBT 7)	0.867	0.001	GP 38 (L 2277)	1.217	0.065
GP 12 (ARCV 111)	0.29	0.403	GP 39 (L 2278)	1.338	2.619
GP 13 (WGCV Bt 26)	0.015	0.337	GP 40 (L 2281)	1.097	3.124
GP 14 (LHBT 2390)	0.103	0.005	GP 41 (L 2265)	0.636	2.092
GP 15 (LHBT 2392)	0.375	0.28	GP 42 (L 2386)	0.215	2.877
GP 16 (LHBT 2385)	0.666	2.287	GP 43 (L 2389)	0.266	0.991
GP 17 (LHBT 2383)	0.049	0.193	GP 44 (L 2390)	1.075	0.131
GP 18 (LHBT 2395)	0.848	1.221	GP 45 (L 2385)	3.063	1.551
GP 19 (Lam Bt 2208)	0.574	0.224	GP 46 (L 2395)	0.797	1.292
GP 20 (ARCV 99)	0.394	0.828	GP 47 (L 2388)	0.567	0.468
GP 21 (LHBT 3)	0.895	0.588	GP 48 (L 2383)	0.037	2.415
GP 22 (Lam Bt 2209)	2.219	1.237	GP 49 (L 2387)	0.293	4.147
GP 23 (LHBT 4)	0.032	0.18	GP 50 (L 2382)	0.886	1.296
GP 24 (L 2381)	9.701	1.221	CICR 23 Bt	3.313	0.048
GP 25 (L 2267)	6.324	0.792	NDLH 2051-1	0.104	2.357
GP 26 (L 2268)	1.878	0.493	Sivanandhi	0.505	0.777
GP 27 (L 2270)	1.383	0.224			



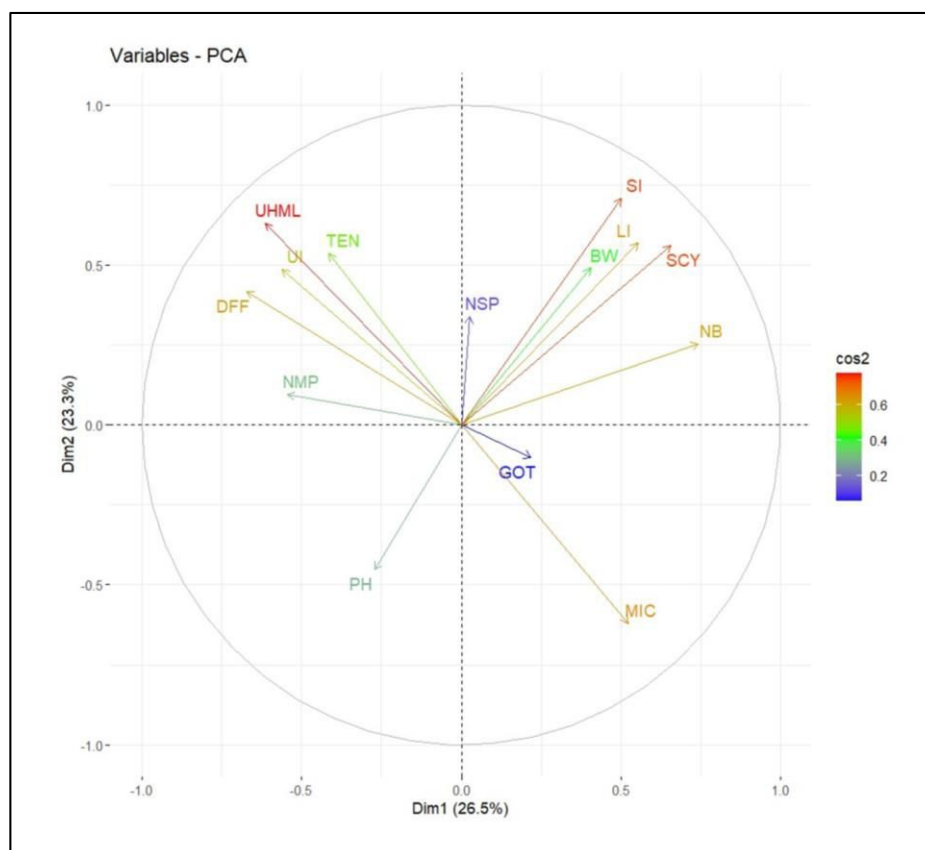


Fig 2: PCA biplot representing distribution of 14 variables in 53 cotton genotypes

Conclusions

Diverse genotypes could be a potential source for transfer of desirable genetic combination for yield improvement. Principal Component Analysis (PCA) identified four principal components with eigen value more than one and attributed 70.56% of cumulative variance. The first two PC's were considered, as it was capable to explain 49.73% of the total variation. The trait number of bolls per plant contributed the most to total divergence with the proportion of 14.82%, followed by days to fifty percent flowering, seed cotton yield per plant and upper half mean length with the proportions of 12.19%, 11.63% and 10.15%, respectively, in PC₁. In second Principal Component (PC₂), seed index contributed the highest proportion to the total divergence with a percentage of 15.36%, followed by upper half mean length, micronaire, lint index and seed cotton yield per plant with the proportions of 12.19%, 11.78%, 9.90% and 9.66%, respectively. The biplot analysis revealed that GP 49 (L2387), GP 34 (L 2396), GP 24 (L 2381), GP 25 (L 2267), GP 29 (Yaganti), GP 30 (Srinandi), GP 31 (NDLA 3116-4), GP 32 (Suvin), GP 33 (CCB 29), GP 37 (L 2275) and GP 45 (L 2385) were identified as the most divergent genotypes as they were located far away from the origin. Hence, these genotypes could be effectively used in hybridization programmes for crop improvement.

References

1. Chapara R, Reddy KV, Rani MS, Lakshmi BS, Roja V, Pranaya J. Variability studies and genetic divergence in cotton (*Gossypium hirsutum* L.) germplasm using multivariate analysis. *Electronic Journal of Plant Breeding*. 2022;13(4):1305-1311.
2. Falconer DS. *Introduction to quantitative genetics*. New Delhi: Pearson Education India; 1964.
3. Farooq J, Rizwan M, Sharif I, Saleem S, Chohan SM, Kainth RA. Genetic diversity studies in some advanced lines of *Gossypium hirsutum* L. for yield and quality related attributes using cluster and principal component analysis. *AA Bioflux*. 2017;9(3):111-118.
4. Federer WT. *Augmented designs*, Hawaiian Planters' Record. 1956;55:191-208.
5. ICAR-All India Coordinated Research Project on Cotton. *Annual Report 2024-25*. Nagpur: ICAR; 2025.
6. Isong A, Balu PA, Ramakrishnan P. Association and principal component analysis of yield and its components in cultivated cotton. *Electronic Journal of Plant Breeding*. 2017;8(3):857-864.
7. Komala M, Ganesan NM, Kumar M. Genetic variability, heritability and correlation analysis in F₂ populations of ratoon upland cotton hybrids. *International Journal of Agriculture, Environment and Biotechnology*. 2018;11(6):815-827.
8. Ruilian L, Zheng L, Jinxiang C. Comparison of fibre quality characteristics of different varieties of hybrid cotton. *Journal of Hunan Agricultural University*. 2005;31(5):492-495.
9. McCarty JC, Jenkins JN, Wu J. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. *Crop Science*. 2004;44(4):1226-1230.
10. Naik BM, Satish Y, Babu DR. Genetic diversity analysis in American cotton (*Gossypium hirsutum* L.). *Electronic Journal of Plant Breeding*. 2016;7(4):1002-1006.
11. Percival AE, Kohel RJ. Distribution, collection, and evaluation of *Gossypium*. *Advances in Agronomy*. 1990;44:225-256.

12. Sahar A, Zafar MM, Razzaq A, Manan A, Haroon M, Sajid S, *et al.* Genetic variability for yield and fibre related traits in genetically modified cotton. *Journal of Cotton Research*. 2021;4(1):19.
13. Sarwar G, Nazir A, Rizwan M, Shahzadi E, Mahmood A. Genetic diversity among cotton genotypes for earliness, yield and fibre quality traits using correlation, principal component and cluster analyses. *Sarhad Journal of Agriculture*. 2021;37(1):307-314.
14. Shakeel A, Talib I, Rashid M, Saeed A, Ziaf K, Saleem MF. Genetic diversity among upland cotton genotypes for quality and yield related traits. *Pakistan Journal of Agricultural Sciences*. 2015;52(1):73-77.