South Asian Journal of Agricultural Sciences 2025; 5(2): 372-375

E-ISSN: 2788-9297
P-ISSN: 2788-9289
Impact Factor (RJIF): 5.57
www.agrijournal.org
SAJAS 2025; 5(2): 372-375
Received: 15-10-2025
Accepted: 19-11-2025

Principal Component Analysis (PCA) in cotton (*Gossypium hirsutum* L.) for estimation of diversity of yield and fibre quality

Vaibhav V Ujjainkar

Dr. Vaibhav V Ujjainkar
Professor of Genetics and Plant Breeding (CAS) & Head of Section (Agril. Botany)
College of Agriculture, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola, Maharashtra, India

DOI: https://www.doi.org/10.22271/27889289.2025.v5.i2e.235

Abstract

Cotton is essential to Indian agriculture, serving as a key cash crop offering raw materials for the global textile industry, provides significant income and job opportunities for millions. Genetic diversity is essential component for any breeding adventure; the genetic diversity in terms of phenotypic expression offers the opportunity to select desirable genotypes that possess multiple advantageous traits simultaneously. By converting extensive datasets comprising morpho-qualitative traits into significant patterns, Principal Component Analysis (PCA) offers breeders statistical assurance to make well-informed choices, resulting in more efficient selection and accelerated advancement of breeding outcomes. With this view, the set of thrity-four genotypes of *Gossypium hirsutum* were tested in randomozed block design with tree replications for fourteen morphological traits including fibre quality components. Analysis of variance exhibited the presence of substantial degree of variability in experimental material which was confirmed by Principal Components (PCs). The 80.46 per cent of total variability present in genepool was accounted by seven principal components, the PC-I contributed the maximum of 27.19 per cent of variability. The trait viz., ginning outturn and lint index were major contributors towards divergence based on first two PCs, whereas, genotypes viz., G-22, G-16, G-29 and G-20 were found to be better and having potential to be included in cotton improvement program.

Keywords: Principal component analysis, *Gossypium hirsutum,* cotton, genetic diversity, fibre attributes

Introduction

Cotton stands as one of the most significant fibre and cash crops in India, playing a pivotal role in the nation's industrial and agricultural economy. It supplies the essential raw material (cotton fibre) for the cotton textile sector. In India, cotton directly supports the livelihoods of six million farmers, while around forty to fifty million people are involved in cotton trading and its processing. The country features ten primary cotton-growing states, which are organized into three distinct zones: the north zone, central zone, and south zone. The north zone comprises Punjab, Haryana, and Rajasthan. The central zone includes Madhya Pradesh, Maharashtra, and Gujarat. The south zone consists of Andhra Pradesh, Telangana, Karnataka, and Tamil Nadu. Additionally, cotton cultivation has seen growth in the eastern state of Orissa. Cotton is also cultivated in smaller regions of non-traditional states like Uttar Pradesh, West Bengal, and Tripura. Four principal cultivated species of cotton include *Gossypium hirsutum* L., *G. barbadense* L., *G. arboreum* L., and *G. herbaceum* L. The *G. hirsutum* group of cultivars and hybrids makes a significant contribution to cotton cultivation areas, characterized by their high yield and wide adaptability. Conversely, *G. barbadense* is cultivated for its exceptionally long, strong, and fine fiber (Ujjainkar et al, 2020) [12].

Understanding the nature and extent of genetic diversity within the germplasm is a crucial prerequisite for the success of any breeding program. Therefore, it is essential to assess the variation found in the germplasm, as hybrids derived from diverse lines exhibit a more significant heterotic effect compared to those from closely related lines. Plant breeders identify genotypes with favourable traits by examining the phenotype, utilizing statistical methods grounded in morphological observations.

Principal components analysis (PCA) can be described as a method for reducing data, specifically for quantitative datasets. PCA takes multiple correlated variables and transforms

Correspondence Author:
Dr. Vaibhav V Ujjainkar
Professor of Genetics and Plant Breeding (CAS) & Head of Section (Agril. Botany)
College of Agriculture, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola, Maharashtra, India

them into a new set of uncorrelated variables for further examination. The new variables are linear combinations derived from the original ones. This process is founded on the calculation of eigenvalues and mutually independent eigenvectors (principal components), which are arranged in descending order according to their variance. Such components yield scatter plots of observations that are optimal for analysing the underlying variability and correlation (Ujjainkar and Marawar, 2020) [12]. Moreover, the PCA is one of the effective statistical technique for simplifying complex datasets, distilling numerous traits into principal components, identifying superior genotypes, choosing parents for hybridization, comprehending the relationships between traits (such as yield versus height), evaluating genetic diversity, and shaping selection strategies for high-yielding, stress-tolerant varieties, thus making the breeding process more efficient and targeted. In this paper, the PCA technique has applied on thirty-four cotton genotypes to find out the degree of variability and contribution of each character towards divergence.

Material and Methods

In present investigation, the thirty-four elite genotypes collected from major cotton growing states of India grown in randomized block design with three replications. The recommended cultivation package and practices were followed to grow healthy plant stand.

The analysis of variance was performed to test the significance of differences among the genotypes for all fourteen quantitative traits under study as per standard method (Panse and Sukhatme, 1967; Singh and Chaudhary, 1985) [5, 11]. The mean data was used for Principal Component Analysis (PCA) using GrapesAgri1 (1.1.0) software (Gopinath, 2021) [3]. However, the grouping of genotypes into different clusters and estimation of inter and intra cluster distances was done with the help of method described by Rao (1952) [6]

Result and Discussion

The data obtained from field experiment was subjected to the analysis of variance for fourteen quantitative traits of cotton under study. The results are presented in Table 1 indicated the highly significant differences among the genotypes for all the fourteen traits under study indicating substantial genetic variability.

Table 1: Analysis of Variance for fourteen quantitative traits in cotton

Characters	Sources of variation		
	Replication (df= 2)	Genotypes (df= 33)	Error (df= 66)
Days to 50 per cent flowering	49.00	56.14**	19.70
Plant Height (cm)	136.40	743.78**	73.74
No. of monopodia plant^{-1}	0.48	1.35**	0.22
No. of sympodia plant^{-1}	30.83	29.84**	4.72
No. of bolls plant^{-1}	7.82	36.15**	7.40
Boll weight (g)	0.03	0.49**	0.03
No. of seeds boll^{-1}	36.75	28.11**	4.23
Ginning outturn (%)	6.54	18.38**	2.29
Seed Index (g)	0.04	0.64**	0.01
Lint Index (%)	0.09	0.85**	0.07
2.5% span length (mm)	1.55	4.41**	0.19
Fibre strength (g/tex)	2.14	14.96**	0.29
Micronaire Value (mg/inch)	0.74	0.30**	0.02
Seed cotton yield plant^{-1} (g)	47.58	95.55**	10.59

The mean data of all the fourteen traits including fibre quality were analysed by principal component analysis to investigate the genetic divergence among thirty five cotton genotypes. The results suggested the importance of the first five PCs with eigen values greater than to one in discriminating the germplasm collection. The results revealed that seven canonical roots accounted for 80.46% of total divergence. The PC-I contributed maximum towards divergence (27.19%) with eigen value of 3.81. The second, third and fourth canonical vectors contributed 12.41%, 10.62% and 9.72% respectively to total divergence with eigen values of 1.74, 1.49 and 1.36 respectively (Table. 2)

Table 2: Canonical vectors for fourteen characters in thirty-four cotton genotypes

Parameter	PC-I	PC-II	PC-III	PC-IV	PC-V	PC-VI	PC-VII
Eigenvalues (Roots)	3.81	1.74	1.49	1.36	1.08	0.99	0.81
Variance (%)	27.19	12.41	10.62	9.72	7.72	7.05	5.76
Cumulative Variance (%)	27.19	39.60	50.22	59.94	67.66	74.71	80.46
Characters							
Days to 50 per cent flowering	6.21	4.21	5.18	1.84	9.03	19.62	0.28
Plant Height (cm)	5.21	0.29	36.04	0.10	1.44	2.81	0.05
No. of monopodia plant^{-1}	2.23	14.51	4.09	5.94	1.28	25.54	13.45
No. of sympodia plant^{-1}	9.07	1.69	6.63	2.40	1.57	2.37	2.42
No. of bolls plant^{-1}	10.22	0.05	0.25	18.03	0.21	13.58	0.46
Boll weight (g)	13.00	1.99	4.27	0.11	0.85	0.64	7.46
No. of seeds boll^{-1}	0.35	2.70	22.01	15.97	7.27	0.00	25.96
Ginning outturn (%)	13.44	13.77	0.10	1.06	2.01	0.13	10.13
Seed Index (g)	3.26	6.96	4.35	0.00	42.57	2.63	15.59
Lint Index (%)	14.77	17.39	1.58	0.84	4.58	0.36	0.94
2.5% span length (mm)	0.24	1.71	7.07	50.19	0.11	3.04	4.87
Fibre strength (g/tex)	6.47	14.77	3.36	2.29	13.21	4.33	11.12
Micronaire Value (mg/inch)	3.03	19.64	2.66	0.59	2.02	22.90	2.28
Seed cotton yield plant^{-1} (g)	12.51	0.33	2.42	0.65	13.84	2.04	5.02

In the first Principal Component (PC-I), lint index (%) was found to be the largest contribution to total divergence (14.77%), followed by ginning outturn (13.44%), boll weight (13.00%), seed cotton yield per plant (12.51%), number of bolls per plant (10.22%), number of sympodia per plant (9.07%), fibre strength (6.47%), days to 50 per cent flowering (6.21%) and plant height (5.21%). Traits with minimal influence on PC-I included seed index

(3.26%), micronaire value (3.0%), number of monopodia per plant (2.23%), number of seeds per boll (0.35%) and 2.5 per cent span length (0.24%) as depicted in Table 2. While, in respect of second Principal Component (PC-II), the fibre quality trait viz., micronaire value had the strongest impact (19.64%), followed by lint index (17.39%), fibre strength (14.77%), number of monopodia^{-1} (14.51%), ginning outturn (13.77%), seed index (6.96%) and days to 50 per cent flowering (4.21%). The lowest contributors to PC-II were number of seeds per boll (2.70%), boll weight (1.99%), 2.5 per cent span length (1.71%), number of sympodial branches per plant (1.69%), seed cotton yield per plant (0.33%), plant height (0.29%) and number of bolls per plant (0.05%). This result was in accordance with work reported by Khan et al., 2015 [4] and Rathinavel, 2019 [7].

In a biplot graph produced through principal component analysis, the base connecting marked points of various attributes is identified as the 'traits vector.' The cosine angle between these attributes indicates the correlation among all the attributes under investigation. In present study, based on the biplot analysis of the traits and individual genotypes, the genotypes viz., G-16, G-29 and G-20 are showing better performance for traits like lint index, plant height and seed cotton yield per plant highlighted the importance of these genotypes for improvement of yield. For boll characteristics, G-22 may be promising aligned with traits like plant height, seed cotton yield, sympodial branches and boll weight. However, in respect of fibre quality traits viz., micronaire value and 2.5 per cent span length, genotypes G-12, G-2, G-8 and G-11 may be used in improvement program (Figure 1). The importance of these traits towards contribution in yield is highlighted by cotton breeders viz., Shakeel et al. (2015) [10], Farooq et al. (2017) [2], Sarwar et al. (2021) [9], Sahar et al. (2021) [8] and Chapara et al. (2022) [1] in their studies.

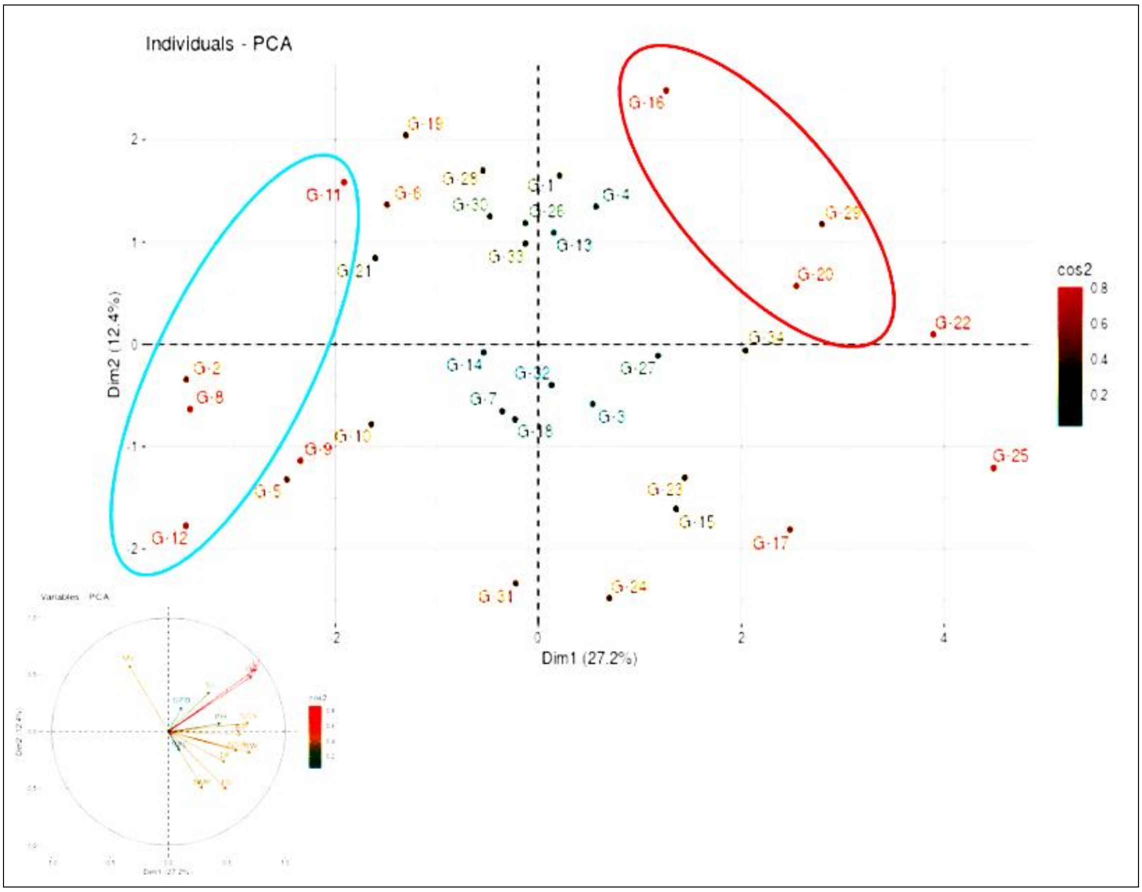

Fig 1: PCA biplot representing distribution of thrity-four cotton genotypes

Conclusion

In present investigation for cotton improvement experiments for yield and fibre quality, the PCA helps identify the potential parents. It might show that genotypes with high lint index, ginning outturn, boll weight, seed cotton yield per plant, number of bolls per plant, number of sympodia per plant (which load heavily on PC-I) are good candidates for crossing to develop new candidates in cotton genepool. Further, it can be concluded that Principal Component Analysis (PCA) is vital tool to handle the multidimensional data based on morphological traits in cotton including fibre quality aspects. Based on the interaction of the genotypes with vector, principal component score, the genotypes viz., G-22, G-16, G-29 and G-20 can be selected for the characters seed yield per plant and its contributing traits like plant height, boll weight, sympodial branches. These genotypes may aid to cotton breeders in formulating the crop improvement program.

Acknowledgement

Author is thankful to Senior Research Scientists, Cotton Research Unit, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola 444 104 (MS) for availing the facilities to conduct the field experiment

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. 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. Advances in Agriculture & Biology Bioflux. 2017;9(3):111–118.
3. Gopinath Pratheesh P. grapesAgri1: Collection of Shiny apps for data analysis in agriculture. Zenodo. 2021. doi:10.5281/zenodo.5106216.
4. Khan FUZ, Rehman SU, Abid MA, Malik W, Hanif CM, Bilal M, Qanmber G. Exploitation of germplasm for plant yield improvement in cotton (*Gossypium hirsutum* L.). Journal of Green Physiology, Genetics and Genomics. 2015;1(1):1–10.
5. Panse VG, Sukhatme PV. Statistical methods for agricultural workers. New Delhi: ICAR Publication; 1967.
6. Rao CR. Advanced statistical methods in biometrical research. New York: John Wiley and Sons Inc.; 1952.
7. Rathinavel K. Agro-morphological characterization and genetic diversity analysis of cotton germplasm (*Gossypium hirsutum* L.). International Journal of Current Microbiology and Applied Sciences. 2019;8(2):2039–2057. doi:10.20546/ijcmas.2019.802.237.
8. Sahar A, Zafar MM, Razzaq A, Manan A, Haroon M, Sajid S. Genetic variability for yield and fibre related traits in genetically modified cotton. Journal of Cotton Research. 2021;4(1):1–9.
9. 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.
10. 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.
11. Singh RK, Chaudhary BD. Biometrical methods in quantitative genetics analysis. Ludhiana: Kalyani Publishers; 1985.
12. Ujjainkar VV, Marawar MW. Genetic diversity: a tool for assessing potential of genepool. Remarking An Analisation. 2020;4(11):E38–E44.
13. Ujjainkar VV, Patil VD, Moharil MP. Diversity assessment of parental lines through biochemical and molecular markers in upland cotton (*Gossypium hirsutum* L.). International Journal of Genetics. 2020;12(6):740–744.
14. Ujjainkar VV, Rathod TH. Character association study in cotton (*Gossypium hirsutum* L.). International Journal of Research and Analytical Reviews. 2022;9:678–683.
15. Ujjainkar VV. Study of environmental influence on genetic diversity estimates (D^2) of parental lines in upland cotton (*Gossypium hirsutum* L.). International Journal of Current Microbiology and Applied Sciences. 2022;11(7):212–221. doi:10.20546/ijcmas.2022.1107.026.