emergent
Life Sciences Research

Research Article

Selection of diverse parents based on genetic distances and cluster analysis in *Gossypium hirsutum* L.

Vaibhav V. Ujjainkar

Received: 15 December 2025
Accepted: 25 February 2026
Online: 12 March 2026

Authors:
V. V. Ujjainkar ✉
Department of Genetics and Plant Breeding (CAS), College of Agriculture
Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola, India

✉ ujjainkarvv.pdkv@gmail.com

Emer Life Sci Res (2026) 12(1): 11-16

E-ISSN: 2395-6658
P-ISSN: 2395-664X

DOI: http://dx.doi.org/10.31783/ELSR.2026.1211116

Abstract

India holds the title of the largest cotton producer globally and is the sole nation that grows all four cultivated species of cotton. This study was conducted with objectives to estimate the magnitude of genetic diversity among a set of elite germplasm lines and to identify parents for a further cotton improvement program. In the present investigation, a set of thirty elite genotypes of *Gossypium hirsutum* L. were evaluated in a randomized block design with eleven morphological traits. Based on genetic diversity values, the thirty genotypes were categorized into ten diverse clusters despite their geographical origin or ancestry. The clusters I, II, and IX were comprised of the maximum genotypes, while four clusters, *viz*., cluster III, IV, VI, and VIII, were single genotype clusters. The highest magnitude of intra cluster distance was exhibited in cluster II (27.38), the cluster combination of cluster III and cluster VIII exhibited the highest inter cluster distance (36.11). Based on benchmark value of the average inter-cluster distance of 15.73, the seventy parental cross combinations along with their reciprocals were suggested as per the cluster combinations that were above the average genetic distance value. In the present investigation with respect to agro-ecological conditions of Akola, the four genotypes AKH 8828, NH-659, RAH 1066, and SCS 1207 emerged as the potent donors for important quantitative characters for the future cotton improvement program.

Keywords clustering, cotton, genetic diversity, seed cotton yield, selection of parents

Introduction

Cotton, a prime cash crop among several economically significant species from the genus *Gossypium*, is cultivated extensively for its seed-hair fibers. Cotton is recognized as one of the foremost crops globally, boasting a remarkable history of utilization, and can be transformed into numerous biodegradable materials. Cotton is recognized as a predominantly self-pollinated and frequently cross-pollinated cash crop. The name 'cotton' is thought to be derived from the Arabic words 'KUTON' or 'QUTON', and it is classified under the genus *Gossypium* and the family Malvaceae. This plant includes nearly forty-five diploid species and five allotetraploid species that are found naturally in the semiarid and arid regions of Africa, Australia, Central and South America, the Indian subcontinent, Arabia, the Galapagos Islands, and Hawaii. Throughout its evolution and domestication, humans have selected four species of cotton for cultivation, which are distinguished by their long and spinnable seed fibers [1]. Cotton is classified under the genus *Gossypium*, which encompasses around forty-six diploid species and seven allotetraploid species. All diploid cotton species (2n=2x=26) are categorized into eight genome groups (A-G and K), while all tetraploid species (2n=4x=52) fall under a single genome group (AD). *Gossypium* species are found across vast geographical and ecological landscapes, showcasing significant morphological and genetic diversity, largely maintained in germplasm collections and serving as genetic resources for cotton improvement programs around the globe [2-5].

During the cropping season of 2024-25, India's cotton cultivation covered approximately 112-113 lakh hectares, with a production of around 307 lakh tonnes (equivalent to about 240 lakh bales) and a yield of roughly 465 kg/ha. However, forecasts for the 2025-26 season indicate a slight reduction in area (to about 109-112 lakh hectares), while production is expected to remain stable at around 240 lakh bales due to enhanced yields resulting from improved monsoon conditions and irrigation practices. This stability comes despite farmers shifting towards more profitable crops and facing pest challenges such as the Pink Bollworm. Maharashtra is the leading state in terms of area, and efforts are underway to increase yields through new productivity initiatives [6].

Being a vital contributor to the economy, there is a pressing need for cotton cultivars that have high potential for both yield and fiber quality. Thus, it is imperative to develop high-yielding cotton cultivars that are economically viable to produce. For this purpose, a better understanding of the genetic diversity of cotton is essential for the effective exploitation of existing variability. Genetic diversity also plays a crucial role in the manifestation of the phenomenon of heterosis. It's well-known that the hybrids between genotypes from diverse origins exhibit greater heterosis compared to hybrids involving closely related parents. Multivariate analysis has been widely applied as a quantitative approach to identifying diverse genotypes [7-9]. The objective of this study was to evaluate the genetic diversity and relationships among thirty elite cotton genotypes through the use of multivariate Mahalanobis D^2 statistics, hierarchical cluster analysis, and principal component analysis to suggest the possible diverse cross combinations for future cotton hybrid production programs.

Methodology

The experimental material comprised thirty genotypes of upland cotton, grown in a Randomized Block Design (RBD) fashion with three replications as per the recommended package and practices for a healthy plant stand at the Farm of AICRIP Centre, Cotton Research Unit, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola (MS). Data were recorded for eleven important yield and yield contributing characters viz., plant height (cm), days to 50 percent flowering, days to 50 per cent boll bursting, number of monopodia and sympodial branches per plant, number of bolls per plant, boll weight (g), seed index (g), lint index, ginning outturn (%) and seed cotton yield per plant (g). The experiment was sown in a 60x30 cm spacing with a plot size of 2.4 x 3 meters.

After performing ANOVA following Panse and Sukhatme [10], the data were subjected to Mahalanobis D^2 analysis (Tocher's method) as given by Rao [11]. Agglomerative hierarchical clustering technique (Ward's minimum variance) was followed for cluster analysis as given by Anderberg [12]. Based on the inter-cluster distances, the parents were suggested for future hybrid production program.

Results and Discussion

The analysis of variance (ANOVA) revealed the highly significant differences among thirty genotypes used as experimental material for the present investigation for all eleven quantitative traits (Table 1), indicating the existence of diversity among the genotypes for the characters studied.

Table 1. Analysis of variance (ANOVA) for eleven morphological traits in cotton

Characters	Source of Variation		
	Replication (df-2)	**Cultivars (df-29)**	**Error (df-58)**
Plant Height (cm)	4.58	168.35**	11.05
Days to 50% Flowering	0.71	26.18**	2.85
Days to 50% Ball Bursting	6.43	27.67**	2.81
No of Monopodia	0.05	0.25**	0.08
No of Sympodia	1.94	9.74**	3.49
No. of bolls per plant	4.72	37.27**	1.67
Boll weight (g)	0.02	0.15*	0.08
Seed Index (g)	1.48	0.86**	0.28
Lint Index	0.30	0.48**	0.09
Ginning Outturn (%)	0.33	8.60**	0.60
Seed Cotton Yield per plant (g)	1.80	8.42**	2.68

(** significant at 1% degrees of freedom)

As per the procedure explained in Tocher's method, the thirty cotton genotypes were grouped into ten distinct clusters, ensuring that the average intra-cluster D^2 values were lower than the inter-cluster D^2 values. The grouping of genotypes into the ten clusters was presented in Table 2. The distribution of thirty cotton genotypes into ten clusters was at random, with the maximum number of genotypes in clusters I, II, and IX (five genotypes each), followed by four genotypes in cluster VII and cluster X, while cluster V comprised three genotypes, and the remaining clusters i.e., III, IV, VI, and VIII comprised of single genotype. The observed clustering pattern in the present investigation indicates that there is not necessarily a relationship between genetic diversity and geographical origin. Instead, this diversity may stem from a variety of influences such as natural selection, the interchange of breeding materials, genetic drift, and environmental changes. Therefore, when selecting genotypes for hybridization, it is essential to focus on genetic diversity rather than geographical distribution. Similar kinds of reports in cotton crop while experimenting on genetic diversity and prediction of potential for hybridization program by many cotton breeders [13-16].

Table 2. Clustering of thirty cotton genotypes in ten clusters

Cluster	No. of genotypes	Genotypes included
I	5	BS 3, CHH 14-4, CNH 7022, NDLH 2010, GTHV 13/17
II	5	AKH 08-8, NH 615, CNH 61, NDLH 2015, GBHV 183
III	1	AKH 8828
IV	1	SCS 1207
V	3	CCH 14-3, L 762, CSH 2
VI	1	RAH 1066
VII	4	AKH 08-3, RAH 1003, IH 70-1, ARBH 1451
VIII	1	NH-659
IX	5	CPD 1451, CPD 1452, GBHV 188, CSH 1, ARBH 1452
X	4	SCS 1061, BS 26, L 770, H 1472

In the present investigation, the intra-cluster average D^2 values ranged from 0.00 to 27.38 (Table 3). Among the ten clusters, Cluster II exhibited the highest intra-cluster distance (27.38), indicating greater genetic variability among its genotypes, while the lowest measurable distance within a cluster was observed in Cluster VII, i.e., 8.73, while other clusters, viz., III, IV, VI, and VIII, exhibited zero intra cluster distances as being single genotype clusters. The maximum inter-cluster distance was observed between Cluster III and Cluster VIII (36.11), followed by Cluster VI and Cluster VIII (32.97) and Cluster IV and Cluster VIII (31.69), suggesting a high degree of genetic divergence between these clusters/groups. These results accentuate the critical nature of carefully choosing both the relevant clusters and the individual genotypes within them while organizing a cotton improvement program through hybridization. The crosses that involve genotypes from diverse clusters are more likely to exhibit significant heterosis and the chances of developing transgressive segregants.

Table 3. Inter cluster distances among ten clusters

Cluster	I	II	III	IV	V	VI	VII	VIII	IX	X
I	**17.06**	8.75	19.30	18.58	9.27	14.96	9.03	19.39	6.66	10.00
II		**27.38**	20.44	16.27	13.84	19.20	8.39	16.43	8.47	11.50
III			**0.00**	8.72	15.10	8.32	17.32	36.11	24.27	10.23
IV				**0.00**	14.85	13.80	16.16	31.69	21.81	9.88
V					**13.27**	10.50	14.70	25.39	12.89	8.57
VI						**0.00**	16.01	32.97	20.14	9.49
VII							**8.73**	22.35	12.09	9.12
VIII								**0.00**	13.53	26.81
IX									**24.06**	14.74
X										**13.69**

(Diagonal values indicates intra-cluster distance; Average D^2 distance – 15.73)

The possible potent cross combinations may be suggested using the criteria that the inter-cluster distances between clusters may be above the average inter-cluster distance (D=15.73) (Table 4). Therefore, the following crosses between genotypes from clusters having higher inter-cluster distances than the average distance are expected to produce desirable transgressive segregants and could be promising for future breeding efforts. The tentative parental combinations are suggested for the hybridization program; the breeder should take care to select the male and female parents based on need or may use these proposed combinations in a reciprocal way to maximize the potential of these lines. A similar kind of research has been reported by cotton workers [17-19].

Table 4. Proposed diverse parental combinations above mean statistical distance (D=15.73)

Cluster combination	Average inter-cluster distance	No of combinations	Possible cross combination
III x VIII	36.11	01	AKH 8828 x NH-659
VI x VIII	32.97	01	RAH 1066 x NH-659
IV x VIII	31.69	01	SCS 1207 x NH-659
VIII x X	26.81	04	NH-659 x SCS 1061, NH-659 x BS 26, NH-659 x L 770, NH-659 x H 1472
V x VIII	25.39	03	CCH 14-3 x NH-659, L 762 x NH-659, CSH 2 x NH-659
III x IX	24.27	05	AKH 8828 x CPD 1451, AKH 8828 x CPD 1452, AKH 8828 x GBHV 188, AKH 8828 x CSH 1, AKH 8828 x ARBH 1452
VII x VIII	22.35	04	AKH 08-3 x NH-659, RAH 1003 x NH-659, IH 70-1 x NH-659, ARBH 1451 x NH-659
IV x IX	21.81	05	SCS 1207 x CPD 1451, SCS 1207 x CPD 1452, SCS 1207 x GBHV 188, SCS 1207 x CSH 1, SCS 1207 x ARBH 1452
II x III	20.44	05	AKH 08-8 x AKH 8828, NH 615 x AKH 8828, CNH 61 x AKH 8828, NDLH 2015 x AKH 8828, GBHV 183 x AKH 8828

Continued

VI x IX	20.14	05	RAH 1066 x CPD 1451, RAH 1066 x CPD 1452, RAH 1066 x GBHV 188, RAH 1066 x CSH 1, RAH 1066 x ARBH 1452
I x VIII	19.39	05	BS 3 x NH-659, CHH 14-4 x NH-659, CNH 7022 x NH-659, NDLH 2010 x NH-659, GTHV 13/17 x NH-659
I x III	19.30	05	BS 3 x AKH 8828, CHH 14-4 x AKH 8828, CNH 7022 x AKH 8828, NDLH 2010 x AKH 8828, GTHV 13/17 x AKH 8828
II x VI	19.20	05	AKH 08-8 x RAH 1066, NH 615 x RAH 1066, CNH 61 x RAH 1066, NDLH 2015 x RAH 1066, GBHV 183 x RAH 1066
I x IV	18.58	05	BS 3 x SCS 1207, CHH 14-4 x SCS 1207, CNH 7022 x SCS 1207, NDLH 2010 x SCS 1207, GTHV 13/17 x SCS 1207
III x VIII	17.32	01	AKH 8828 x NH-659
II x VIII	16.43	05	AKH 08-8 x NH-659, NH 615 x NH-659, CNH 61 x NH-659, NDLH 2015 x NH-659, GBHV 183 x NH-659
II x IV	16.27	05	AKH 08-8 x SCS 1207, NH 615 x SCS 1207, CNH 61 x SCS 1207, NDLH 2015 x SCS 1207, GBHV 183 x SCS 1207
IV x VII	16.16	04	SCS 1207 x AKH 08-3, SCS 1207 x RAH 1003, SCS 1207 x IH 70-1, SCS 1207 x ARBH 1451
VI x VIII	16.01	01	RAH 1066 x NH-659
Total Parental combinations suggested- 70			

Conclusion

In the present investigation, it has been observed that ANOVA is not sufficient to understand the variability among genotypes; the genetic differences exhibited in statistical distances are a helpful tool for understanding diversity and employing it for finding diverse candidates of the population. It can be concluded that the high yielding and genetically diverse parents are likely to contribute the maximum potential in hybridization programs. As a result, the D^2 analysis is used to identify diverse parental combinations for hybridization. Besides this, genotypes AKH 8828, NH-659, RAH 1066, and SCS 1207 may be useful as donors for various characters.

References

[1] M. Reddychennaiahgari, V. Ujjainkar, A. Akhare and A. Kharkar **(2025)**. Relation of seed vigour to yield performance in different seed lots of cotton (*Gossypium hirsutum* L.). Int. J. Res. Pub. Rev., **6:** 5636-5642.

[2] C. L. Brubaker, A. H. Paterson and J. F.Wendel **(1999)**. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome, **42:** 184-203.

[3] P. Fryxell **(1979)**. The natural history of the cotton tribe (Malvaceae, Tribe Gossypieae); Texas A & M University Press: College Station, TX, USA., pp245.

[4] J. F. Wendel, C. Brubaker, I. Alvarez, R. Cronn and J. M. Stewart **(2009)**. Evolution and natural history of the cotton genus. *In* A. H. Paterson (Ed.), Genetics and Genomics of Cotton, Springer Science and Business Media LLC, New York, pp3-22.

[5] J. F. Wendel and R. C. Cronn **(2003)**. Polyploidy and the evolutionary history of cotton. Adv. Agron., **78:** 139-186.

[6] Anonymous **(2025)**. Report on Cotton and product updates., USDA Foreign Agriculture Service.

[7] R. Asha, M. L. Ahamed, D. R. Babu and P. A. Kumar **(2013)**. Multivariate analysis in upland cotton (*Gossypium hirsutum* L.), Madras Agric. J., **100:** 333-335

[8] V. V. Ujjainkar **(2022)**. Similarity coefficient analysis using UPGMA based on polymorphism of isozyme and DNA markers in Cotton. Int. J. Res. Appl. Sci. Eng. Technol., **10:** 961-977.

[9] V. Ujjainkar **(2025)**. D^2 statistics technique: methodology and applications in plant breeding. Int. J. Res. Rev., **12:** 127-139.

[10] V. G. Panse and P. V. Sukhatme **(1978)**. Statistical method for agricultural workers ICAR Publication. New Delhi

[11] C. R. Rao **(1952)**. Advanced statistical methods in biometrical research. John Wiley and Sons Inc., New York.

[12] M. R. Anderberg **(1973)**. Cluster analysis for application. Academic Press, New York

[13] J. J. Gauswami., M. G. Valu and G. N. Odedara **(2021)**. Correlation, path analysis and D^2 analysis study of seed cotton yield and fibre quality traits in American cotton (*Gossypium hirsutum* L.). J. Pharmocogn. Phytochem., **10:** 222-230

[14] B. M. Naik, Y. Satish and D. R. Babu **(2016)**. Genetic diversity analysis in American cotton (*Gossypium hirsutum* L.)., Electron. J. Plant Breed., **7:** 1002-1006.

[15] S. D. Pathan, S. N. Devkule, H. V. Kalpande, A. M. Dethe and V. N. Chinchane **(2025)**. Character association study in cotton (*Gossypium hirsutum* L.). Int. J. Adv. Biochem. Res., **SP-9:** 151-157

[16] V. V. Ujjainkar **(2022)**. Study of environmental influence on genetic diversity estimates (D^2) of parental lines in upland cotton (*Gossypium hirsutum* L.). Int. J. Curr. Microbiol. App. Sci., **11:** 212-221.

[17] C. Pavan, R. Chapara, M. S. Rani, G. A. D. Grace, M. James and Priyadharshini **(2025)**. Genetic diversity studies for seed cotton yield and fiber quality parameters in diverse cotton genotypes. Plant Arch., **25:** 2645-2651.

[18] A. Manivannan and B. Dharajothi **(2019)**. Multivariate Analysis based Characterization of Cotton Germplasm. Int. J. Curr. Microbiol. App. Sci., **8:** 480-487.

[19] S. Vachhani, S. R. Patel, S. G. Patel, H. N. Patel, A. Dinisha and A. Nayak **(2025)**. Multivariate analysis in cotton. Pharma Innovation., **14:** 8-11