



RESEARCH ARTICLE

# Multivariate analysis of genetic variability and divergence in upland cotton (*Gossypium hirsutum* L.) for yield and fiber quality traits

Nileshkumar Tamboli<sup>1</sup>, D H Patel<sup>2</sup>, Chirag P Chandramaniya<sup>3\*</sup>, V B Patel<sup>4</sup>, Denish Savaliya<sup>1</sup>, Manoj C Suthar<sup>1</sup>, Parmar Mukeshkumar<sup>5</sup>, Bhushan Kale<sup>1</sup> & Kaushal Modha<sup>1</sup>

<sup>1</sup>Department of Genetics and Plant Breeding, Navsari Agricultural University, Navsari 396 450, Gujarat, India

<sup>2</sup>Main Cotton Research Station, Navsari Agricultural University, Surat 395 007, Gujarat, India

<sup>3</sup>Department of Genetics and Plant Breeding, Anand Agricultural University, Anand 388 110, Gujarat, India

<sup>4</sup>Department of Basic Science and Humanity, College of Forestry, Navsari Agricultural University, Navsari 396 450, Gujarat, India

<sup>5</sup>Department of Genetics and Plant Breeding, Indian Agricultural and Research Institute, Central Arid Zone Research Institute, Jodhpur 342 003, Rajasthan, India

\*Correspondence email - [chiragpatel91745@gmail.com](mailto:chiragpatel91745@gmail.com)

Received: 14 February 2025; Accepted: 01 August 2025; Available online: Version 1.0: 14 October 2025

**Cite this article:** Nileshkumar T, Patel DH, Chirag PC, Patel VB, Denish S, Manoj CS, Parmar M, Bhushan K, Kaushal M. Multivariate analysis of genetic variability and divergence in upland cotton (*Gossypium hirsutum* L.) for yield and fiber quality traits. Plant Science Today (Early Access). <https://doi.org/10.14719/pst.7734>

## Abstract

Cotton (*Gossypium hirsutum* L.) is a versatile crop with multiple applications. The evaluation of 30 upland cotton genotypes was conducted to measure variability and divergence in seed cotton yield, yield-related attributes and fiber quality traits using multivariate analyses. The investigation, conducted during the *Kharif* season of 2022, was laid out in a randomized block design (RBD) with three replications at the Main Cotton Research Station (MCRS), Navsari Agricultural University (NAU), Surat. Analysis of variance (ANOVA) showed significant genetic variability between the genotypes for all traits. Morphological observations identified GISV-400 as the genotype with the highest seed cotton yield (148.00 g/plant). Fiber quality analysis highlighted GISV-391 as superior in upper half mean length (28.03 mm), fiber strength (28.37 g/tex) and fiber fineness (4.80 µg/inch). Seed cotton yield per plant displayed a significant positive relationship with the number of bolls per plant and boll weight. Principal component analysis (PCA) identified five components with eigenvalues greater than one, collectively explaining 78.66 % of the total variation. The biplot revealed GISV-399, GISV-398, GISV-391, GISV-322, GISV-323, GISV-313 and GISV-402 as the most genetically diverse genotypes. Hierarchical clustering further classified the 30 genotypes into five distinct groups. To identify the best-performing genotypes, the Multi-Trait Genotype-Ideotype Distance Index (MGIDI) was employed for the ranking of genotypes at 10 % selection intensity, four genotypes (GISV-389, GISV-395, GISV-394 and GISV-391) were identified as superior. Overall, seven genotypes (GISV-389, GISV-395, GISV-394, GISV-391, GISV-398, GISV-399 and GISV-400) were consistently identified as high-performing in terms of both yield and fiber quality based on PCA, hierarchical clustering and MGIDI. Overall, the study indicated that the existing variability in tetraploid cotton can be effectively utilized through hybridization and development of mapping populations.

**Keywords:** cotton; hierarchical clustering; MGIDI; multivariate; principal component

## Introduction

The term "cotton" originates from the Arabic word "al qatan" and is widely referred to as "White Gold." As one of the greatest key cash crops, cotton occupies a crucial role in agriculture, contributing significantly to economic and social development. Cotton stands as the world's leading natural fiber and oilseed crop (1, 2). Cotton, often referred to as the "King of Fibers," is a valuable cash crop and a vital raw material for the textile industry. It supports millions of livelihoods globally through its extensive involvement in farming, seed production, marketing, industrial applications and research activities. Cotton belongs to the family Malvaceae and is classified within the tribe Gossypieae. Within this tribe, which comprises eight recognized genera, cotton is categorized under the genus *Gossypium*. Cotton exhibits remarkable genetic diversity, encompassing

more than 50 species, of which 46 are diploid and 7 are tetraploid. Globally, only four cultivated species of cotton are widely recognized: *G. herbaceum*, *Gossypium arboreum*, *G. barbadense* and *G. hirsutum*. Notably, all four species are cultivated in India.

The top cotton producing countries include China, India, the United States, Brazil and Pakistan, with India leading in terms of the largest area dedicated to cotton cultivation, covering nearly 40 % of the global total. Remarkably, India pioneered the commercial utilization of heterosis in cotton, achieving a significant milestone with the release of the world's first commercial intra-*hirsutum* hybrid, H-4 (*Gujarat-67* × *American Nectariless*), by the late Dr. C T Patel in 1971 at the Main Cotton Research Station, Gujarat Agricultural University, Surat (3). India ranks second in global cotton production,

contributing approximately 23 % to the world's total output. During 2022-23, cotton cultivation in India spanned about 125.10 lakh hectares, with a production of 341.91 lakh bales and a productivity of 465 kg/ha. Cotton is not only a major fiber crop but also a key oilseed crop worldwide. Although primarily cultivated for its fiber, cotton also plays a significant role in the edible oil market, contributing 15-25 % of the global oilseed supply. Cottonseed oil, derived from cotton seeds, is prized for its premium quality and cholesterol-free properties, making it a popular and healthy dietary option (4).

Since breeding is an evolutionary process, it is crucial to conserve genetic resources, including high-performing cultivars, breeding lines and wild relatives, for the continued improvement of crops (5). Cotton breeding's major goal is to continuously improve the genetics of current plant germplasm to produce new genetic recombinants. The goal is to increase seed cotton yield possible per unit area while maintaining acceptable fiber quality under a variety of agroclimatic situations (6). Genetic diversity, which includes diverse agronomic and morphological features as well as their interactions with biotic and abiotic variables, has been exploited to promote breeding (7). The breadth and character of genetic variability in germplasm provide numerous options for application in successful breeding programs aiming at enhancing various traits. Effective varietal selection is critical in preserving genetic resources while ensuring consistent diversity of desirable attributes (8). The goal of the investigation was to assess the genetic variability and interrelationships among morpho-yield and fiber quality traits in cotton germplasm using correlation analysis and PCA and to identify high-yielding, superior fiber quality genotypes through MGIDI for use in breeding programs.

## Materials and Methods

### Experimental material and field evaluation

The present study was conducted during the *Kharif* season of 2021-22 at the MCRS, NAU, Surat. The station is situated in a tropical zone characterized by hot summers, mild winters and humid monsoon. Geographically, it is positioned at 21° N latitude and 72° E longitude, at an elevation of 12 m above mean sea level. The weather during the growing season was normal and favourable. The experiment consisted of 30 distinct genotypes arranged in a RBD with three replications, aimed at evaluating morphological characteristics. A detail of genotypes and their sources along with pedigree are given in Supplementary Table 1. Each plot contained 12 plants, with 120 cm spacing between rows and 45 cm between individual plants. To ensure healthy crop growth, appropriate and uniform agronomic practices were followed, along with timely plant protection measures.

### Phenotyping for yield, yield attributing and fibre quality traits

For each genotype, five representative plants were randomly selected and marked within each replication. These plants were evaluated for various traits, including days to 50 % flowering (DFF), days to maturity (DM), plant height (PH; cm), monopodia per plant (MPP), sympodia per plant (SYPP), bolls per plant (BPP), boll weight (BW; g), seed cotton yield per plant (SCYP; g), ginning outturn (GOT; %), seed index (SI; g), upper

half mean length (UHML; mm), fiber strength (FS; g/tex) and fiber fineness (FF;  $\mu\text{g}/\text{inch}$ ). Additionally, plot-based observations were conducted for the phenological traits of DFF and DM, ensuring accurate data collection across the study.

For the analysis of fiber quality, 100 g of fiber samples were sent to the Central Institute for Research on Cotton Technology (CIRCOT) laboratory at NAU, Surat. The pooled fiber samples were evaluated for fiber quality characteristics under controlled conditions of  $65 \pm 2$  % relative humidity and  $27 \pm 2$  °C. The analysis was conducted using the high-volume instrument (HVI) mode, adhering to the ASTM D-5867-2012 standard to ensure precise and reliable results.

### Statistical analysis of data

The data collected for the studied traits were subjected to ANOVA and the critical difference (CD) was calculated to identify significantly different genotypes using the R software. Correlation analysis was performed for each environment using the *corrplot* function of the *metan* package (9) in R. PCA and biplots were generated using paleontological statistics (PAST) version 4.16 c to detect high- and low-yielding genotypes. Additionally, hierarchical clustering analyses were conducted in R, utilizing the *heatmap* package for visualizing relationships among genotypes. The MGIDI, a novel method proposed (10), was employed to evaluate the proximity of genotypes to an ideal ideotype across multiple traits. The MGIDI was computed using the *metan* R package, with trait-specific relative values (RVs) serving as the basis for the analysis, providing a robust framework for multi-trait genotype evaluation.

## Results and Discussion

### Analysis of variance and variability among the morpho-fiber quality traits

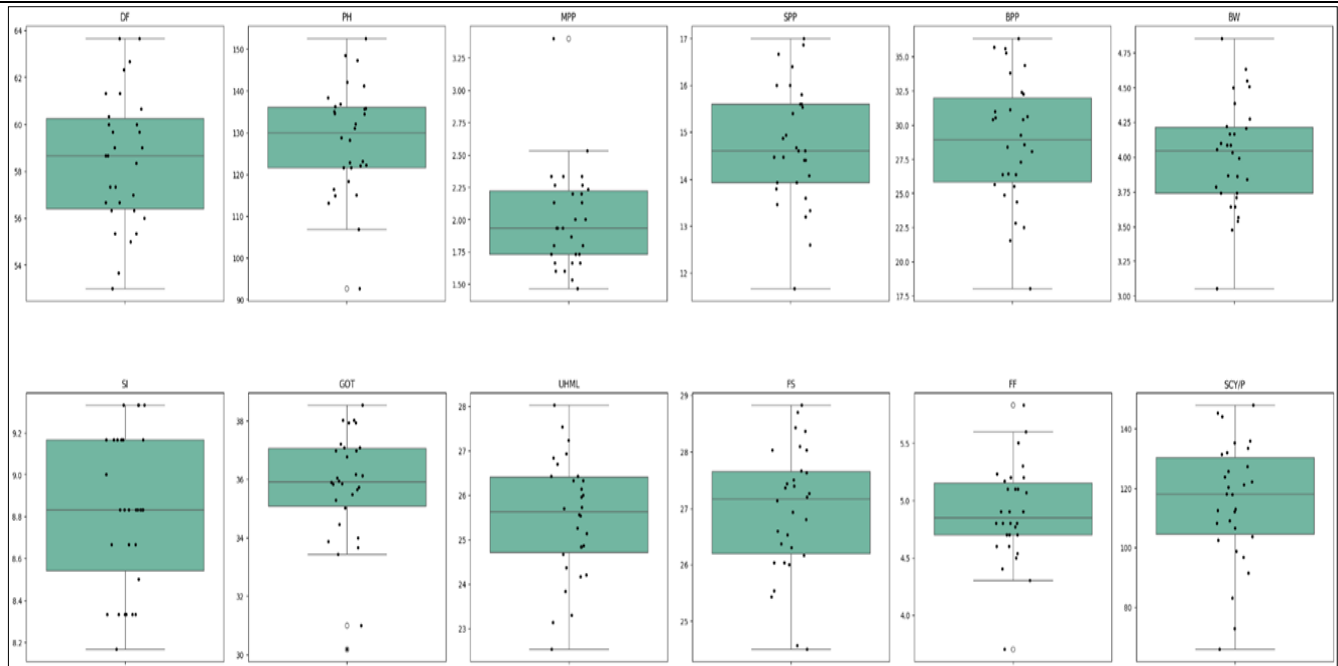
Genetic variability encompasses the total genetic differences present among individuals within a population or community. In this study, the analysis of variance showed that the mean square for genotypes was highly significant ( $p \leq 0.01$ ) for morpho-fiber quality traits Table 1. These findings are consistent with the results reported (11-13). The significant differences among genotypes for various traits observed in the ANOVA highlight the presence of substantial variation within the studied genotypes. The significant variability observed for each character suggests the existence of ample genetic diversity, providing a robust foundation for the selection of promising advanced lines in cotton improvement efforts. Plant breeders consistently prioritize harnessing the genetic diversity found within existing germplasm to attain desired traits (14).

### Phenological and morphological traits

Descriptive mean value of morpho-fiber quality traits was visualized using a box plot Fig. 1 and Supplementary Table 2. Early maturing genotypes provide considerable benefits, particularly in areas prone to recurrent terminal droughts. They can also avoid peak pest and disease outbreaks, adapt to unexpected weather conditions and ensure the timely planting of subsequent crops. Early flowering is desirable, with DFF ranging from 53.00 to 63.67 days (average 58.50). GISV 398 was the earliest flowering in 53 days, while GISV 401 had the latest

**Table 1.** Analysis of variance and mean performance of quantitative characters in 30 genotypes

Characters	Replications (Df: 2)	Genotypes (Df: 29)	Error (Df: 58)	Mean	Range	S. Em. $\pm$	CD at 5 %	CV %
Days to 50 % flowering	1.03	23.86 **	9.39	58.50	53.00-63.67	1.77	5.01	5.23
Plant height (cm)	5.00	517.31 **	158.93	128.34	92.60-152.60	7.28	20.60	9.82
Monopodia per plant	0.07	0.45 **	0.04	2.00	1.47-3.40	0.12	0.33	10.21
Sympodia per plant	2.58	4.90 **	2.31	14.73	11.67-17.00	0.88	2.49	10.33
Bolls per plant	31.17	63.67 **	9.87	28.87	18.04-36.34	1.81	5.13	10.88
Boll weight (g)	0.22	0.47 **	0.07	4.01	3.07-4.87	0.16	0.45	6.9
Ginning out turn (%)	1.17	11.42 **	1.98	35.74	30.19-38.55	0.81	2.30	3.93
Seed index (g)	0.08	0.40 **	0.18	4.01	8.17-9.33	0.16	0.45	4.82
Upper half mean length (mm)	0.94	5.37 **	0.11	25.49	22.53-28.03	0.19	0.54	1.3
Fibre strength (g/tex)	0.66	3.81 **	0.15	26.96	24.50-28.83	0.22	0.63	1.4
Fibre fineness (mv)	0.11	0.52 **	0.08	4.89	3.70-5.83	0.16	0.47	5.82
Seed cotton yield per plant (g)	276.80	1228.98**	147.33	115.27	72.90-148.00	7.01	19.84	10.53

**Fig. 1.** Box plots displaying mean performance of the studied traits.

flowering time at 63.67 days. The significance of PH in cotton varies depending on the specific breeding objective. For machine harvesting, shorter sympodial varieties are typically preferred. However, under suboptimal management conditions, taller genotypes with more monopodial branches may be better suited. PH showed considerable variation among the genotypes, ranging from 92.60 cm (GISV-313) to 152.60 cm (GISV-391), with an average of 128.34 cm. Typically, dwarf plants like GISV-313 are ideal for high-density planting, which can enhance yield potential.

Monopodial (vegetative) and sympodial (fruiting) branches play a vital role in cotton, as they influence branching pattern, plant structure and fruiting efficiency. Monopodial branches contribute to the main stem and primary branches, whereas sympodial branches give rise to secondary and fruiting branches. Proper management and balance of their development are essential for optimizing crop yield and enhancing overall cotton productivity. In addition, monopodial genotypes can be planted with wider spacing, which reduces the need for expensive hybrid seeds and, as a result, lowers production costs. Generally, a lower number of monopodial branches per plant is preferred. The average number of monopodia ranged from 1.47 (GISV-319) to 3.40, with GISV-398 exhibiting the highest count. The overall mean number of MPP was 2.00. The number of sympodial branches per plant stretched from 11.67 (GISV-322) to 17.00 (GISV-39), with an average of

14.73. Out of the 30 genotypes, 13 exhibited a higher number of sympodial per plant (SYPP) than the mean.

Bolls per plant (BPP) play a significant role in seed cotton yield. The 30 germplasms displayed a wide variation in the number of BPP, ranging from 18.04 (GISV-322) to 36.34 (GISV-269). The overall average for BPP was 28.87, with 14 genotypes exceeding this average. Seed cotton yield is determined by the balance between flower production and the retention or shedding of flowers and bolls. Selecting germplasms with a higher number of effective bolls is crucial for maximizing yield. Genotypes with greater boll weight (BW) are key contributors to higher cotton yield and more efficient picking with reduced waste. A larger BW reflects the potential for increased cotton fiber production, thereby enhancing the overall yield per plant and per unit area. Breeding and selecting genotypes with higher boll weight is a crucial approach for enhancing cotton yield and productivity. BW ranged from 3.07 g to 4.87 g, with an average of 4.01 g. Genotype GISV-386 demonstrated exceptionally high BW (4.87 g), while GISV-324 exhibited lower BW (3.07 g).

#### Yield and yield attributing traits

The primary goal of cotton breeders is to develop varieties and hybrids with enhanced yield potential while maintaining acceptable fiber quality. This focus on improving both yield and fiber traits serves two key purposes: increasing cotton production to meet the textile industry's demands and

promoting agricultural sustainability. Understanding the variability in yield, a complex polygenic trait, is essential for maximizing crop harvests. In the current study, seed cotton yield per plant varied significantly, ranging from 72.90 to 148.00 g across all genotypes. GISV 400, GISV 375, GISV 389, GISV 269 and GISV 399 exhibited high seed cotton yields and could serve as valuable parents for the development of high-yielding hybrids.

Ginning out turn (GOT) is a vital factor in the cotton industry, as it directly influences the quantity of lint obtained after ginning. A high GOT is beneficial for cotton producers as it enhances the market value of lint and boosts profitability. The GOT ranged from 30.19 % (GISV-216) to 38.55 % (GISV-398), with the overall mean at 31.29 %. These variations underscore the significance of GOT assessment in cotton, as it directly affects lint yield and overall cotton quality. Seed index plays a key role in determining yield. The average seed index of the 30 genotypes was 4.01 g, ranging from 8.17 g (GISV-323) to 9.33 g (GISV-380, GISV-384, GISV-391 and GISV-393). Larger seeds have a reduced surface area-to-seed mass ratio, which results in lower seed lint content. Therefore, seed index is considered a contrasting trait in cotton, with a lower seed index being more desirable.

### Fibre quality traits

The UHML (Upper Half Mean Length) of cotton is a vital parameter in the textile industry for determining fiber quality. Longer UHML values are highly desirable, as they indicate stronger and more uniform fibers, which are ideal for manufacturing premium-quality textiles. Cotton with a UHML of less than 24 mm tends to exhibit diminished spinning efficiency, making longer fiber lint the preferred choice for cotton-based industries. In this study, UHML values ranged from 22.53 mm (GISV-402) to 28.03 mm (GISV-391), with an overall average of 25.49 mm. The observed variation in UHML can be attributed to genetic differences among cotton species and the influence of diverse environmental conditions.

Cotton varieties with superior tensile strength are highly valued in the textile industry for producing durable and high-performance fabrics. In this study, fiber strength varied from 24.50 g per tex (GISV-402) to 28.83 g per tex (GISV-322), with an overall average of 26.96 g per tex, consistent with findings reported in previous research. Micronaire, a crucial indicator of fiber fineness, is most desirable at lower values as it directly impacts the texture of the fiber determining whether it feels soft or coarse, silky or harsh. Environmental factors significantly influence these characteristics. In this study, the highest micronaire value was recorded in genotype GISV-387 (5.83  $\mu$ g per inch), reflecting relatively coarser fiber compared to other genotypes. Conversely, the lowest micronaire value was observed in genotype GISV-375 (3.70  $\mu$ g per inch), indicating exceptionally fine fiber. The overall mean for fiber fineness across the genotypes was 4.89  $\mu$ g per inch.

Based on the mean performance, the genotypes GISV 400, GISV 375, GISV 389, GISV 269 and GISV 399 recorded high seed cotton yield thus it can serve as a potential donor for future breeding programs aimed at developing high-yielding hybrids. The genotype GISV-391 possessed long fibre length and average fibre strength with length and strength ratio nearly to one, which is desired by the textile industry and thus this genotype could be utilized for the development of suitable genotypes/hybrids for textile industry. Hence, these genotypes hold potential for utilization in crop improvement programs aimed at enhancing specific desirable traits.

### Correlation coefficient analysis

Correlation heat maps were created to recognize the genotypic association between morpho-fiber quality variables with seed yield per plant Fig. 2. SCYP was highly significant and positively correlated with BPP (0.85) and BW (0.48) (Fig. 3). This suggests that these characteristics can be enhanced concurrently with SCYP through direct selection. Therefore, selecting these traits

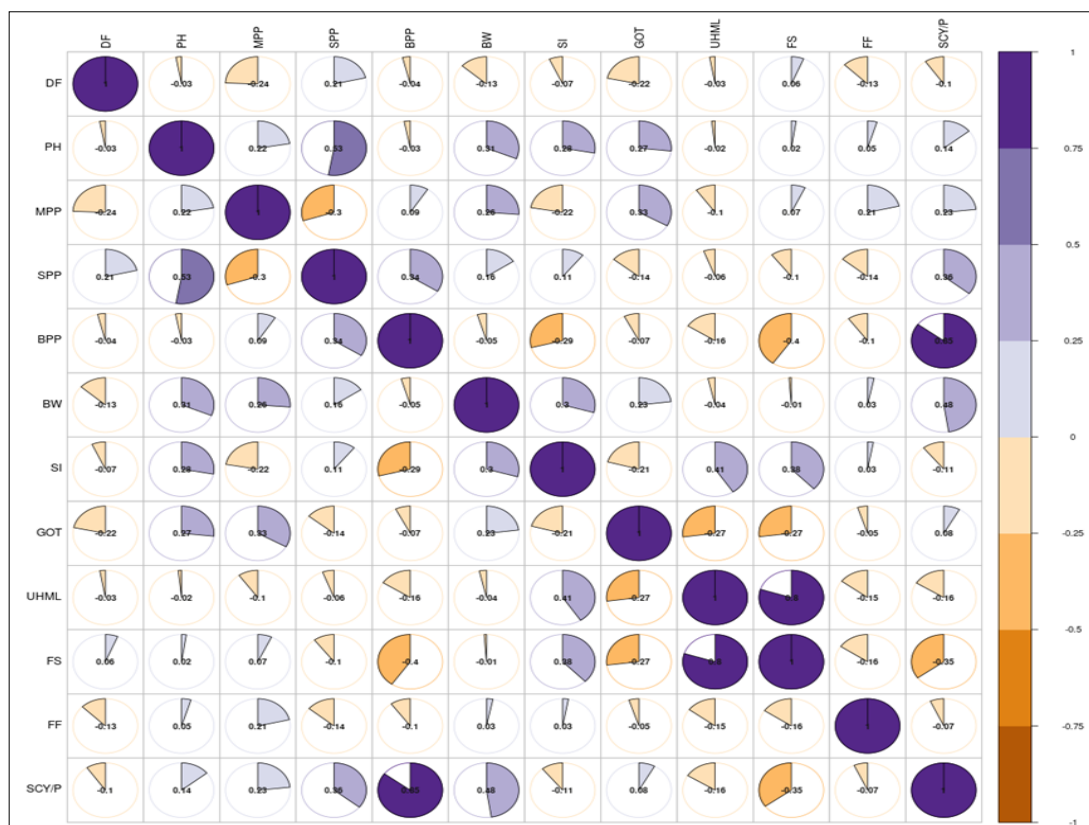
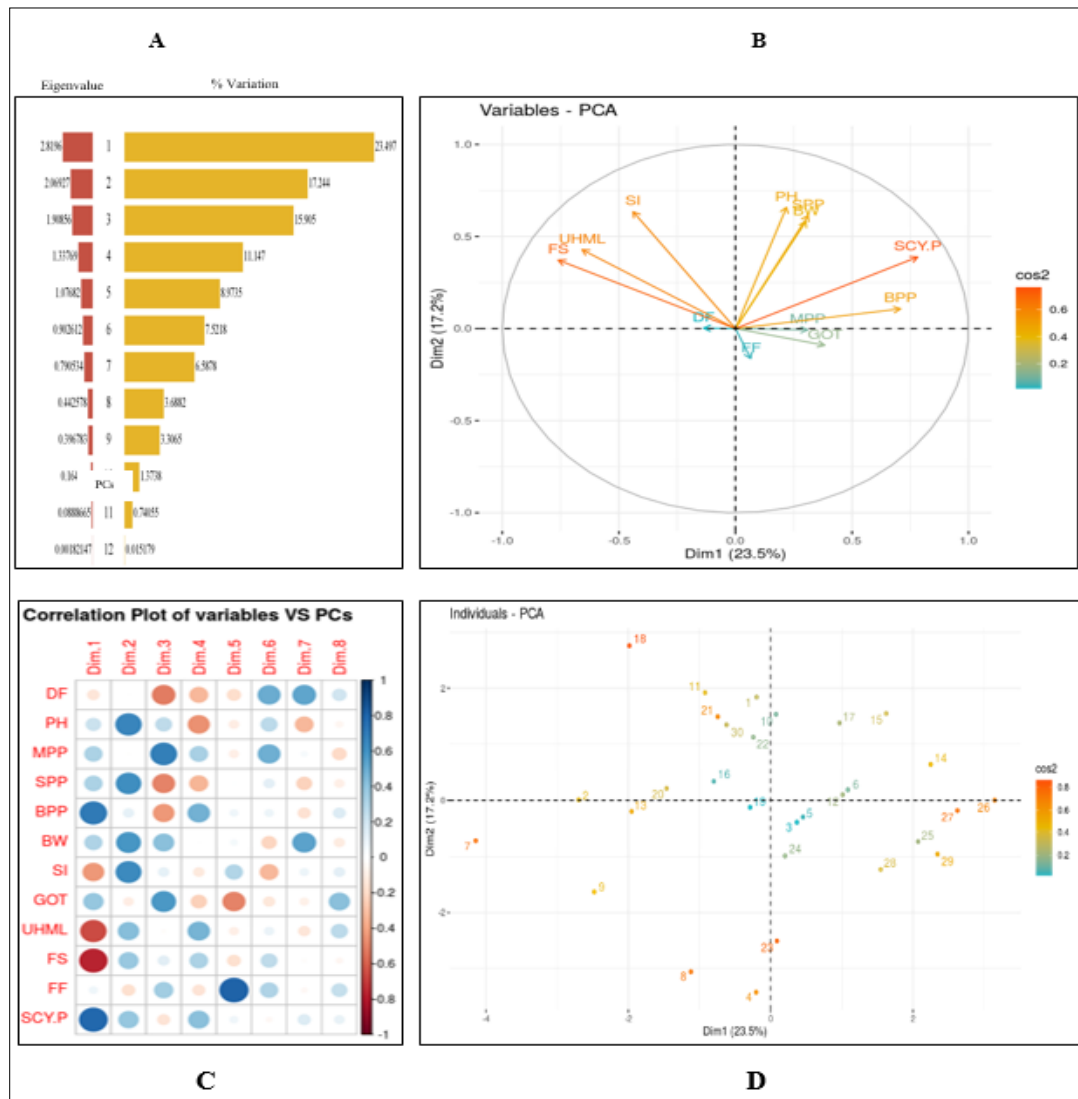


Fig. 2. Correlation matrices of the studied traits.





**Fig. 3.** Summary of panels (A) Depict butterfly bar charts showing the variable percentage contribution of each principal component (PC) as well as the eigenvalue. (B) Quality of representation of different traits ( $\cos^2$ ). Panels (C) and (D) display biplots involving PC1 and PC2, illustrating the allocation of 12 traits and 30 genotypes respectively. Scattered numbers across the plot indicates serial number of lines enlisted in materials and methods.

will help identify genotypes with improved SCYP. Similar outcomes have been reported (15-21). PH ( $r = 0.14$ ), MPP ( $r = 0.23$ ), SYP ( $r = 0.36$ ) and GOT ( $r = 0.08$ ) revealed non-significant positive association with SCYP. Similar results for monopodia per plant were also detected (22, 23). However, SCY/P were negative association with DF ( $r = -0.10$ ), SI ( $r = -0.11$ ), UHML ( $r = -0.16$ ), FS ( $r = -0.35^*$ ) and FF ( $r = -0.07$ ). Among these, the negative correlation with FS ( $r = -0.35^*$ ) is the most significant, suggesting that higher fiber strength may come at the expense of seed cotton yield. This indicates a potential trade-off between yield and certain quality traits, which should be carefully balanced in breeding plans aimed at improving both yield and fiber quality. Akin outcomes were stated for DF and SI (24); BW (25).

#### Principal component and biplot analysis

Significant genetic diversity was identified among cotton components in this study. PCA, an advanced multivariate analysis technique, was employed to simplify and interpret complex high-dimensional datasets (26, 27). We applied PCA to enhance our ability to differentiate between the measured traits based on their relationships. This method enabled the identification of key traits that contributed the most to overall variability, providing deeper insights into trait interactions. Among the twelve principal

components (PCs), five components exhibited eigenvalues greater than 1, accounting for 78.74 % of the cumulative variability for the traits under investigation Supplementary Table 3 and Fig. 3A.

The cumulative contribution rate was 76.74 %. Principal component I (PC I) had an eigenvalue of 2.819, contributing 23.49 % of the total variability. Germplasm in PC I had the most significant positive impact on BPP and SCY/P (Supplementary Table 3). Principal component II (PC II) exhibited an Eigenvalue of 2.069, explaining 17.24 % of the variability. The germplasm in PC II showed positive effects on PH, SPP, BW and SI. Principal component III (PC III) had an Eigenvalue of 1.908, accounting for 15.90 % of the variability. Germplasm in PC III positively influenced monopodia per plant. Principal component IV (PC IV) showed an Eigenvalue of 1.337, contributing 11.14 % of the total variability, with a favorable impact on UHML and FS. Principal component V (PC V) had an Eigenvalue of 1.076, explaining 8.97 % of the variability, with positive results for FF. Germplasm lines exhibiting maximum positive PC scores and common presence in PC1, PC2, PC3, PC4 and PC5 are lines GISV-389, GISV-398, GISV-399 and GISV-400 (Supplementary Table 4). Selecting these lines can contribute significantly to the further advance of new high-yielding varieties/hybrids.

The biplot, also referred to as a polygon, generated during principal component analysis, displayed the 30 germplasm based on their genetic characteristics. The germplasms GISV-399, GISV-398, GISV-391, GISV-322, GISV-323, GISV-313 and GISV-402 were located at the vertices of the polygon, indicating the highest genetic diversity. These highly diverse genotypes have the potential to be valuable in future cotton improvement programs. Conversely, several germplasms were positioned close to the origin of the polygon, suggesting a higher degree of genetic similarity and minimal divergence due to their narrow genetic base. Germplasms such as GISV-392, GISV-318, GISV-323, GISV-387, GISV-382 and GISV-397 were found nearer to the origin, indicating lower genetic diversity and suggesting their limited utility for broadening the genetic base in upcoming breeding efforts. The cos 2 (squared cosines or squared coordinates) values are used to estimate the quality of representation of the variables on factor map. A high cos 2 value indicated the good representation of the variable on the principal component whereas low cos 2 indicated that the variable is not perfectly represented by the PCs Fig. 3B.

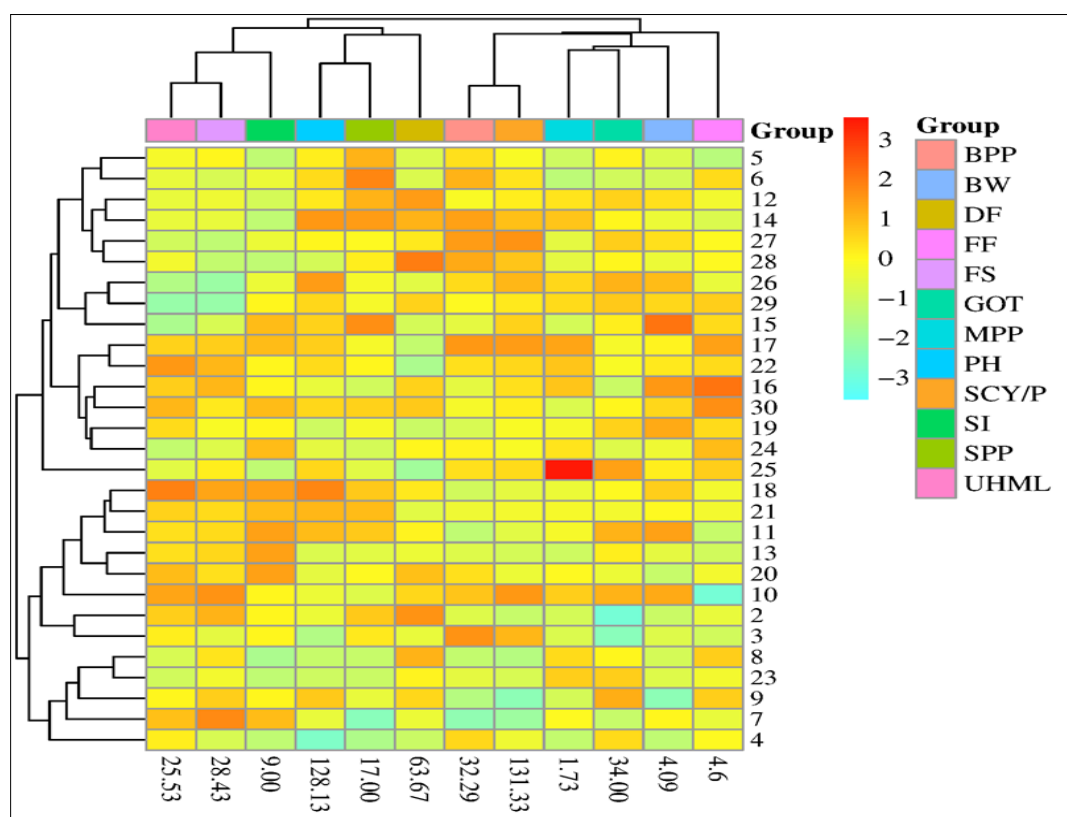
The biplot analysis effectively differentiated the associated traits based on their correlations: positively associated characteristics ( $> 90^\circ$ ), independent attributes ( $= 90^\circ$ ) and negatively associated traits ( $> 90^\circ$ ). In character biplot for the GY exhibited a strong association with SPP, BPP and BW as indicated by the very low angle between their corresponding lines Fig. 3C. SCYP showed a marked negative correlation with SI, UHML and FS, as indicated by the angle between their corresponding vectors being greater than  $180^\circ$ . The distance between the genotype and the biplot origin can be used to measure the genotypic performance in biplots. For one or more attributes, the distant genotypes may have the greatest value. The biplot of the PCA separates the GISV-389 and GISV-399 genotypes as remote genotypes that exhibit a strong positive association with SPP and

BW. The biplot of the PCA separates the GISV-391 genotypes as remote genotypes that exhibit a strong positive association with SI, UHML and FS. The key traits identified in PC I and PC II, including BPP and SCYP. Principal component II (PC II) exhibited an eigenvalue of 2.069, explaining 17.24 % of the variability. The germplasm in PC II showed positive effects on PH, SPP, BW and SI were major contributors to the collective variability. These attributes should be prioritized in future cotton improvement programs. Similar findings, highlighting the contribution of these variables to cumulative variability and their potential in advancing cotton breeding, have also been reported (28-30). The genotypes GISV- 322, GISV- 324, GISV- 323, GISV- 313, GISV- 402, GISV- 399, GISV- 386 and GISV- 391 exhibited the highest diversity for various traits, as they were positioned far from the origin Fig. 3D.

### Hierarchical clustering of genotypes and traits

Researchers employ hierarchical cluster analysis to characterize genotypes based on similarities across several traits. The heat map provides a visual representation of how various genotypes perform relative to each other under both optimal growing conditions and heat stress. Based on average, the heat maps divided the wheat genotypes into various clusters and different groups based on the variation available in the studied traits Fig. 4. The distances between genotype pairs were calculated using the Euclidean distance method.

Following this, all genotypes were clustered together using the complete method to create a comprehensive dendrogram. Among 30 cotton lines formed 5 clusters in the depicts that 9, 6, 1, 8 and 5 genotypes were included in Groups -1, -2, -3, -4 and -5 respective Fig. 4. The diagram shows the genotypes performance as a gradient of colors, from red to blue, which correspond to the highest and lowest performance, respectively. This visualization is acquired from clustering analysis conducted. Genotypes in group 1, such as GISV-318, GISV-319,



**Fig. 4.** Heat map study of different 12 characteristics.

GISV-382, GISV-385, GISV-389, GISV-401, GISV-399, GISV-402 and GISV-386 exhibited higher yields and attributing traits and moderate UHML, FF and FS values. However, genotypes within cluster-III, such as GISV-391 displayed good performance for UHML and FS traits, indicating their fiber quality traits.

### Selection of high yielding and good fiber quality genotypes using MGIDI

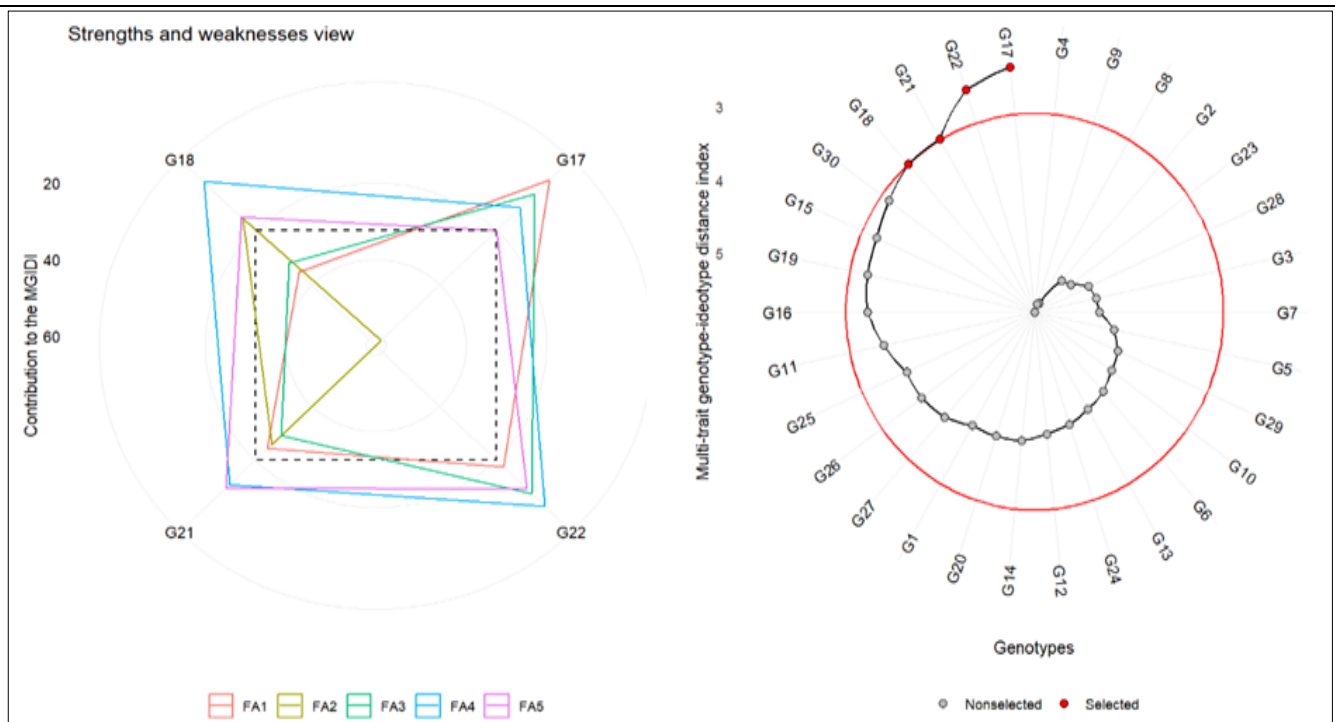
There are various drawbacks to cluster analysis that can make it difficult to pick tolerant genotypes. These include subjective interpretation, difficulties managing missing data, inadequate dimensionality reduction, the inability to consider interaction effects and lack of statistical rigor (31). To address these constraints, it is critical to incorporate additional analytical approaches. Cluster analysis can be integrated with quantitative indicators such as the MGIDI to help identify high yielding with good fiber quality genotypes. MGIDI is an ideal method for genotypic selection due to its ability to address multicollinearity and eliminate the need for assigning weights (9). The method enhances breeding efficiency by reducing the time and resources needed to develop cultivars (32).

This index relies on factor analysis, an efficient technique for modeling detected genotypes using undetected latent components, while exploiting the common variance among related genotypes.

To identify high yielding genotypes, the MGIDI was calculated using all attributes. Table 2 shows the findings of the factor analysis. The traits were clustered into five factors, FA1 - BPP and SCYP; FA2 - PH, SYP and BW; FA3 - DF, MPP and GOT; FA4 - SI, UHML and FS and FA5 - FF with the average communality equal to 0.77. Broad-sense heritability ( $h^2$ ) was spread from 0.52 (SPP) to 0.97 (UHML) for different traits Table 2. These high heritability estimates indicate that these attributes are strong candidates for cotton improvement through selective breeding. Table 2 displays the predicted average genotypic values and the selection gains. DF and GOT exhibited negative gains, while SCYP demonstrated a positive gain of 3.59 %. The genotypes GISV-389, GISV-395, GISV-394 and GISV-391 were selected Fig. 5A. Fig. 5B illustrates the relative strengths and weaknesses of the examined genotypes, as determined by total of five factors (FA1, FA2, FA3, FA4 & FA5) each factor's contribution to the MGIDI score for each genotype.

**Table 2.** Estimated Factor analysis, broad sense heritability and genetic gain for the physio-quality characteristics in the MGIDI

SOV	FA1	FA2	FA3	FA4	FA5	Factor	$h^2$ (bs)	SG (%)
Days to 50 % flowering	0.08	0.00	0.61	0.07	-0.19	FA3	60.6	-1.52
Plant height (cm)	0.04	-0.87	-0.02	0.04	-0.07	FA2	69.3	4.95
Monopodia per plant	-0.09	-0.1	-0.8	0.07	-0.02	FA3	90.8	6.81
Sympodia per plant	-0.39	-0.6	0.56	0.06	-0.09	FA2	52.7	0.76
Bolls per plant	-0.94	0.09	0.02	0.19	-0.08	FA1	84.5	1.54
Boll weight (g)	-0.2	-0.64	-0.37	-0.09	0.1	FA2	83.7	1.68
Seed index (g)	0.16	-0.47	0.13	-0.59	0.31	FA4	54.3	1.01
Ginning out turn (%)	0.22	-0.31	-0.53	0.46	-0.39	FA3	82.7	-0.138
Upper half mean length (mm)	0.04	0.04	-0.01	-0.91	-0.12	FA4	97.9	5.92
Fibre strength (g/tex)	0.28	0.00	-0.06	-0.87	-0.16	FA4	96.1	3.12
Fibre fineness (mv)	0.12	-0.01	-0.18	0.18	0.86	FA5	84.6	2.05
Seed cotton yield per plant (g)	-0.92	-0.25	-0.19	0.12	-0.03	FA1	88.0	3.59



**Fig. 5.** The MGIDI analysis of lines ordering is presented in ascending order (A, B). The genotypes with the highest rankings and selection are highlighted in red. The central red circle indicates the cut-off point, determined by the selection pressure (A). The percentage contribution of each factor in the generated MGIDI index illustrates the strengths and weaknesses of each lines (B). The closer a factor's indices are to the ideotype, the lower the fraction of explanation, indicating proximity to the outer boundary.

MGIDI aids as a valued graphical tool that highlights the strengths and weaknesses of genotypes, offering insights into how they perform in traits that require enhancement. A strength-weakness analysis revealed that FA1 had the greatest influence on G17, FA2 on G18 and FA3 on G22. FA4 contributed most significantly to G18, while FA5 had the highest impact on G21. For instance, FA1's strong influence on G17 signified its superior performance in BPP and SCYP. Meanwhile, FA2 and FA4 for G18 were linked to PH, SPP, BW, SI, UHML and FS. A similar methodology was utilized to evaluate the performance of 13 strawberry lines (103). A system using MGIDI to identify promising guar lines with high gum and seed yield was developed in a different study (33). Likewise, MGIDI was proposed as a powerful tool for enhancing selection methods in breeding climate-resilient maize hybrids, assessing their performance under varying moisture and drought conditions (34). The use of MGIDI in quinoa, focusing on different plant spacing strategies, was also demonstrated (35). In our study, MGIDI is applied to upland cotton, providing a comprehensive framework for identifying genotypes with both high yield and superior quality traits, which are well-suited for hybrid development.

## Conclusion

The MGIDI index, applied to both morpho-fibre quality traits, was effective in identifying the top-performing cotton genotypes. Among the various traits, bolls per plant, sympodia per plant and boll weight emerged as the most reliable indicators for selecting optimal genotypes. Seven genotypes, GISV-389, GISV-395, GISV-394, GISV-391, GISV-398, GISV-399 and GISV-400 were consistently identified as the best performers for morpho-quality traits. These genotypes hold significant potential as valuable resources for breeding programs aimed at developing high-yielding cotton varieties and they can serve as parental lines for the creation of superior hybrids.

## Acknowledgements

The authors acknowledge Main Cotton Research Station, Navsari Agricultural University, Surat, for providing experimental land and other necessary facilities for conducting the experiment. We appreciate the facilities, resources and research assistance provided by the Department of Genetics and Plant Breeding, N. M. College of Agriculture, NMCA, NAU, Navsari.

## Authors' contributions

CPC was the primary drafter of the manuscript, contributed to writing, review, figure and editing and performed data analysis. NT was responsible for data curation, data analysis and investigation. DHP, VBP, BK and KM designed the overall study, established the methodology and provided project administration and supervision. DS, MCS and PM assisted in drafting the original manuscript and contributed to its review and editing. All authors read and approved the final manuscript.

## Compliance with ethical standards

**Conflict of interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Ethical issues:** None

## References

- Patel J, Lubbers E, Kothari N, Koebernick J, Chee P. Genetic and genomic analysis of cottonseed oil. *Oil Crop Genomics*. 2021;9(4):53–74. [https://doi.org/10.1007/978-3-030-70420-9\\_3](https://doi.org/10.1007/978-3-030-70420-9_3)
- Salimath SS, Romsdahl TB, Konda AR, Zhang W, Cahoon EB, Dowd MK. Production of tocotrienols in seeds of cotton (*Gossypium hirsutum* L.) enhances oxidative stability and offers nutraceutical potential. *Plant Biotechnol J*. 2021;19:1268–82. <https://doi.org/10.1111/pbi.13557>
- Patel CT. Hybrid-4: A new hope towards self-sufficiency in cotton in India. *Cotton Dev*. 1971;1(2):1–6.
- Khan NU, Hassan G, Kumbar MB, Parveen A, Ahmed S. Gene action of seed traits and oil content in upland cotton (*Gossypium hirsutum* L.). *SABRAO J Breed Genet*. 2007;39(1):17–29. <https://www.researchgate.net/publication/292759563>
- Latif A, Bilal M, Hussain SB, Ahmad F. Estimation of genetic divergence, association, direct and indirect effects of yield with other attributes in cotton (*Gossypium hirsutum* L.) using biplot correlation and path coefficient analysis. *Trop Plant Res*. 2015;2(2):120–6. <https://www.tropicalplantresearch.com/archives/2015/vol2issue2/9.pdf>
- Ali MA, Khan IA. Assessment of genetic variation and inheritance mode in some metric traits of cotton (*Gossypium hirsutum* L.). *J Agric Soc Sci*. 2007;4(3):112–6. <https://www.researchgate.net/publication/252597944>
- Liaqat S, Ahmed RI, Ahmad S, Bilal M, Karim A, Qayyum A, et al. Evaluation of diverse germplasm of cotton (*Gossypium hirsutum* L.) for yield and fibre attributes using a multivariate analysis approach. *J Agric Sci Rev*. 2015;4(5):146–50.
- Ahmad RI, Liaqat S, Noor E, Qayyum A, Irshad A, Hussain A, et al. A multivariate analysis approach for evaluating diverse germplasm of cotton (*Gossypium hirsutum* L.) for seed and fibre characteristics. *Basic Res J Agric Sci Rev*. 2015;4(6):164–71.
- Olivoto T, Lúcio AD. metan: An R package for multi-environment trial analysis. *Methods Ecol Evol*. 2020;11:783–9. <http://doi.org/10.1111/2041-210X.13384>
- Olivoto T, Nardino M. MGIDI: Toward an effective multivariate selection in biological experiments. *Bioinformatics*. 2021;37(10):1383–9. <https://doi.org/10.1093/bioinformatics/btaa981>
- Dhivya R, Amalabalu P, Pushpa R, Kavithamani D. Variability, heritability and genetic advance in upland cotton (*Gossypium hirsutum* L.). *Afr J Plant Sci*. 2014;8(1):1–5. <https://doi.org/10.5897/AJPS2013.1099>
- Dahiphale KD, Deshmukh JD, Jadhav AB, Bagade AB. Genetic variability, correlation and path coefficient analysis for yield and its attributing traits in cotton (*Gossypium hirsutum* L.). *Int J Trop Agric*. 2015;33(1):15–22. <https://www.cabidigitallibrary.org/doi/full/10.5555/20153325880>
- Chaudhari MN, Faldu GO, Ramani HR. Genetic variability, correlation and path coefficient analysis in cotton (*Gossypium hirsutum* L.). *Adv Biomed Res*. 2017;8(6):226–33.
- Pour-Aboughadareh A, Ahmadi J, Mehrabi AA, Etminan A, Moghaddam M. Insight into the genetic variability analysis and relationships among some *Aegilops* and *Triticum* species, as genome



- progenitors of bread wheat, using SCoT markers. *Plant Biosyst.* 2018;152:694–703. <https://doi.org/10.1080/11263504.2017.1320311>
15. Gnanasekaran M, Thiyaagu K, Gunasekaran M. Studies on genetic variability, correlation and path analysis in upland cotton. *Electron J Plant Breed.* 2020;11(3):981–6. <https://www.ejplantbreeding.org/index.php/EJPB/article/view/3396>
  16. Nawaz B, Sattar S, Malik TA. Genetic analysis of yield components and fibre quality parameters in upland cotton. *Int J Adv Multidiscip Res Stud.* 2019;9:13–9.
  17. Amanu AW. Phenotypic and genotypic correlation coefficients and path coefficient analysis studies of upland cotton (*Gossypium hirsutum* L.). *Int J Plant Breed.* 2020;7(1):667–76.
  18. Chapepa B, Mubvekeri W, Mare M, Kutwayo D. Correlation and path coefficient analysis of polygenic traits of upland cotton genotypes grown in Zimbabwe. *Cogent Food Agric.* 2020;6(1):1–11. <https://doi.org/10.1080/23311932.2020.1823594>
  19. Rehman A, Mustafa N, Du X, Azhar MT. Heritability and correlation analysis of morphological and yield traits in genetically modified cotton. *J Cotton Res.* 2020;3(1):1–9. <https://link.springer.com/article/10.1186/s42397-020-00067-z>
  20. Pooja OS, Pundir SR, Rath M, Mandhan S. Association studies with direct and indirect effects among different morpho-bio traits in *Gossypium arboreum*. *Electron J Plant Breed.* 2021;12(1):86–90. <https://ejplantbreeding.org/index.php/EJPB/article/view/3514>
  21. Hampannavar MR, Patil BR, Katageri IS, Kumar BA, Janagoudar BS. Genetic variability and correlation analysis for agronomic and fibre quality traits in intraspecific cotton (*G. hirsutum* × *G. hirsutum*) recombinant inbred lines (RILs). *Int J Microbiol Appl Sci.* 2020;9(1):493–503. <https://doi.org/10.20546/ijcmas.2020.901.054>
  22. Nawaz B, Sattar S, Malik TA. Genetic analysis of yield components and fibre quality parameters in upland cotton. *Int J Adv Multidiscip Res Stud.* 2019;9:13–9.
  23. Rai P, Sangwan O. Correlation studies in elite lines of upland cotton (*Gossypium hirsutum* L.). *Electron J Plant Breed.* 2020;11(1):298–300. <https://www.ejplantbreeding.org/index.php/EJPB/article/view/3260>
  24. Reddy KB, Reddy VC, Ahmed ML, Naidu TCM, Srinivasarao V. Correlation and path coefficient analysis in upland cotton (*Gossypium hirsutum* L.). *Int J Pure Appl Biosci.* 2015;3(3):70–80.
  25. Jyoti J, Valu MG, Odedara GN. Correlation, path coefficient and D2 analysis study of seed cotton yield and fibre quality traits in American cotton (*Gossypium hirsutum* L.). *J Pharmacogn Phytochem.* 2021;10(3):222–30.
  26. Khan NU, Abro HK, Kumbhar MB, Hassan GM, Khan. Exploitation of heterosis can combat Cotton Leaf Curl Virus (CLCV) incidence in cotton (*G. hirsutum* L.). *Pak Cottons.* 1999;43(3&4):21–33.
  27. Khan NU. Genetic analysis, combining ability and heterotic studies for yield, its components, fibre and oil quality traits in upland cotton (*G. hirsutum* L.) [PhD thesis]. Sindh Agric Univ Tandojam, Pakistan; 2003.
  28. Saeed F, Farooq J, Mahmood A, Riaz M, Hussain T, Majeed A. Assessment of genetic diversity for cotton leaf curl virus (CLCuD), fibre quality and some morphological traits using different statistical procedures in *Gossypium hirsutum* L. *Aust J Crop Sci.* 2014;8(3):442–7. [https://www.cropj.com/farooq\\_8\\_3\\_2014\\_442\\_447.pdf](https://www.cropj.com/farooq_8_3_2014_442_447.pdf)
  29. Kaleri AA, Rajput SY, Kaleri GA, Marri JA. Analysis of genetic diversity in genetically modified and non-modified cotton (*Gossypium hirsutum* L.) genotypes. *J Agric Vet Sci.* 2015;8(12):70–6.
  30. Latif A, Bilal M, Hussain SB, Ahmad F. Estimation of genetic divergence, association, direct and indirect effects of yield with other attributes in cotton (*Gossypium hirsutum* L.) using biplot correlation and path coefficient analysis. *Trop Plant Res.* 2015;2(2):120–6. <https://www.tropicalplantresearch.com/archives/2015/vol2issue2/9.pdf>
  31. Rupji M, Dwivedi B, Kowalski J. NOJAH: Not Just Another Heatmap for genome-wide cluster analysis. *PLoS One.* 2019;14(3):e0204542. <https://doi.org/10.1371/journal.pone.0204542>
  32. Olivoto T, Diel MI, Schmidt D, Lúcio AD. MGDI: A powerful tool to analyze plant multivariate data. *Plant Methods.* 2022;18:121. <https://link.springer.com/article/10.1186/s13007-022-00952-5>
  33. Benakanahalli NK. A framework for identification of stable genotypes based on MTSI and MGDII indexes: An example in guar (*Cyamopsis tetragonoloba* L.). *Agronomy.* 2021;11(6):1221. <https://doi.org/10.3390/agronomy11061221>
  34. Singamsetti A, Zaidi PH, Seetharam K, Vinayan MT, Olivoto T, Mahato A, et al. Genetic gains in tropical maize hybrids across moisture regimes with multi-trait-based index selection. *Front Plant Sci.* 2023;14:1147424. <https://doi.org/10.3389/fpls.2023.1147424>
  35. Ahmed SR, Ali Z, Ijaz I. Multi-trait selection of quinoa ideotypes at different levels of cutting and spacing. *Sustainability.* 2023;15(14):11446. <https://doi.org/10.3390/su151411446>

#### Additional information

**Peer review:** Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

**Reprints & permissions information** is available at [https://horizonpublishing.com/journals/index.php/PST/open\\_access\\_policy](https://horizonpublishing.com/journals/index.php/PST/open_access_policy)

**Publisher's Note:** Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Indexing:** Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc. See [https://horizonpublishing.com/journals/index.php/PST/indexing\\_abstracting](https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting)

**Copyright:** © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

**Publisher information:** Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.