Vadodaria *et al.*
*CABI Agriculture and Bioscience* (2024) 5:111
https://doi.org/10.1186/s43170-024-00314-2

CABI Agriculture and Bioscience

RESEARCH

Open Access

# Characterization of desi cotton (*Gossypium arboreum*) germplasm through morpho-biochemical and fiber traits and exploring molecular marker based diversity along with allied (un)cultivated Gossypium species

Hrushik Vadodaria[1,2], Dhramendra Patidar[2*], Sushil Kumar[1*], Dipak A. Patel[1] and Hardik Patel[1]

## Abstract

*Desi* cotton (*Gossypium arboreum*) or the Asian tree cotton occupies the least area amongst all the cultivated *Gossypium* species, but it is of paramount importance as a vital source of high-quality lint and stress resistance for commercially cultivated cotton species. Hence, it is essential to analyze the genetic diversity of this crop for future advancements in cotton breeding. An study comprising of 30 *G. arboreum* accessions was undertaken to estimate the genetic variability using morphological traits, fibre quality parameters and oil quality. Likewise, molecular diversity analysis using 24 SSR markers was carried out for the 30 *G. arboreum* genotypes and 11 additional cotton species. Presence of substantial genetic variability was revealed by analysis of variance (ANOVA) amongst all genotypes for all the traits. Morphological studies indicated that genotype PA 255 had higher seed cotton yield (80.07 g/plant). Examination of fibre and oil quality suggested that genotype Cernuum had the highest fibre fineness (7.27 µg/inch) and oil (17.68%), while genotypes 824 (29.57 mm) and PA 812 (29.30 g/tex) had the maximum UHML and fibre strength, respectively. 30 genotypes were classified into 3 classes based on the seed cotton yield, which is considered a very important trait. Manhattan dissimilarity co-efficient based phenotypic diversity divided the 30 genotypes into four main clusters, with an average dissimilarity value of 0.17, indicating low phenotypic variability. The dendrogram generated using SSR marker data based on Nei's genetic distance grouped 41 genotypes into three clusters, with average distance of 0.47. Based upon the placement of the number of genotypes in same clusters in both matrices, it can be inferred that there is considerable correlation between morphological and molecular analysis as the value for Mantel statistic R was 0.42 (*p* value 0.0003). Fatty acid profiles of 30 *desi* cotton genotypes revealed the presence of nine different fatty acids of which linoleic, palmitic and oleic acid were found to be in highest amount, with an average of 43.22%, 25.13% and 24.97%, respectively. Overall, the study suggested that the variability available in diploid cotton can be exploited through hybridization, mapping population development and polyploidization based pre-breeding.

**Keywords** ANOVA, Asiatic cotton, Fatty acid, SSR, Variability

*Correspondence:
Dhramendra Patidar
dharmendrapbg@gmail.com
Sushil Kumar
sushil254386@yahoo.com; sushil_k@aau.in

## Introduction

Cotton holds a prominent position as one of the most significant commercial crops, playing a vital role in agricultural, economic, and social spheres. It is recognized as a cash crop and a valuable source of fiber, making it highly important for various industries and livelihoods. It is classified within the family Malvaceae and the tribe Gossypieae. Among the eight recognized genera in Gossypieae tribe, cotton falls under the genus Gossypium. Cotton displays extensive genetic diversity with over 50 species. Of these, 46 are diploids and 7 are tetraploids. Worldwide, only four cultivated cotton species are recognized, including *Gossypium arboreum, G. herbaceum, G. hirsutum*, and *G. barbadense*, all of which are grown in India. *G. arboreum* and *G. herbaceum*, known as old world or Asian cottons, are diploid ($2n=2\times=26$) with A1 and A2 genomes, respectively, and are exclusively cultivated in Asia. Tetraploid species, *G. hirsutum* and *G. barbadense*, account for 98% of global cotton production, while diploid species, *G. arboreum* and *G. herbaceum*, contribute only 2% of the total global output (Shim et al. 2018).

*Gossypium* species are classified into primary, secondary, and tertiary gene pools based on their genetic hybridization properties. The primary gene pool comprises of both cultivated (*G. hirsutum* and *G. barbadense*) and wild allotetraploids (*G. tomentosum, G. mustelinum*, and *G. darwinii*). Diploids with A, B, D, and F genomes make up the secondary gene pool, while diploids with C, E, G, and K genomes form the tertiary gene pool (Shim et al. 2018).

The leading cotton-producing nations include China, India, the United States, Brazil, and Pakistan. India stands out with the largest cotton cultivation area, accounting for approximately 40% of the global total. In terms of production, India ranks second, contributing 23% to the world's cotton output. In the 2021–22, cotton cultivation in India covered an area of 12.37 million hectares, yielding 31.1 million bales of 170 kg each. India's productivity stands at 428 kg/ha, which is lower than the global average yield of around 755 kg/ha (Anonymous 2021).

In addition to being a primary fiber crop, cotton holds a significant position as a major oilseed crop worldwide. Cotton is predominantly cultivated worldwide for its fiber, but it also serves as a significant source of edible oil (15–25%). Cottonseed oil, derived from cotton seeds, is renowned as a premium quality vegetable oil. Notably, cottonseed oil is cholesterol-free, enhancing its appeal as a healthy dietary option (Khan et al. 2007). The fatty acid profile of cottonseed shows that it consists of 70% unsaturated fatty acids including 18% monounsaturated (oleic), and 52% polyunsaturated (linoleic and linolenic) and 30% saturated fatty acids (Daniel 2007).

Asian cotton (*G. arboreum* L.), also known as tree/*desi* cotton, is indigenous to the Indian subcontinent. From *G. arboreum*, several races have evolved, including indicum, burmanicum, sinense, sudanense, cernuum, and bengalense. Despite accounting for only 2–3% of global cotton production, diploid species play a crucial role as they serve as a valuable source of essential genes for resistance against biotic and abiotic stresses (Parekh et al. 2018). *Desi* cotton, cultivated extensively in India, is well-suited for rainfed conditions. With the rise of challenges like pink bollworm and the whitefly-Cotton Leaf Curl Virus complex, there is a renewed interest in *desi* cotton (Parekh et al. 2016). Unlike *Bt* hybrids, *desi* cotton demonstrates sturdiness, tolerance to various pests, resistance to abiotic stresses, immunity to the cotton leaf curl virus, and reduced reliance on chemical inputs due to natural selection under environmental stress. Due to its lower production costs, higher net returns, and climate resilience, *desi* cotton is anticipated to play a vital role in ensuring sustainable and increased cotton yields in India (Blaise et al. 2020). The diverse landraces of *desi* cotton hold great significance as genetic resources for the enhancement of tetraploid cotton. These landraces possess adaptive characteristics that can contribute immensely to cotton-breeding programs by introducing valuable genes for early maturity, stress tolerance, and high fiber strength (Xiang 1988; Rahman et al. 2002; Mehetre et al. 2003; Liu et al. 2006).

A thorough understanding of the genetic relationships at morphological and DNA level among *desi* cotton genotypes as well genetic relation with other species is crucial as it facilitates efficient utilization and harnessing of their potential for developing superior cotton cultivars with desirable agronomic traits. By unraveling these genetic relationships, breeders can make informed decisions and implement effective strategies to exploit the diversity present in *desi* cotton, ultimately leading to the development of improved cotton varieties. Research on genetic diversity using microsatellites (simple sequence repeats, SSRs) is relatively limited, especially in India (Sethi et al. 2015; Saravanan et al. 2021; Santosh et al. 2021). Additionally, many of these studies have not conducted comparisons of genetic diversity between cultivated and wild species. Moreover, previous studies have been conducted without considering morphological or quantitative traits. As a result, it is difficult to obtain a clear understanding of the variability within the germplasm. Taking into account all the aforementioned facts, the current study was carried out to investigate the genetic variability of Asian cotton. The study encompassed both morphological traits and molecular markers to comprehensively assess the diversity within *desi* cotton.

## Materials and methods

### Experimental materials and field evaluation

The current study was conducted during the *Kharif* (an Indian term for summer) season of 2021–22 at the Regional Cotton Research Station (RCRS), Anand Agricultural University (AAU), Viramgam. The experiment was executed in a randomized complete block design with three replications, and the plants were spaced at 120 × 30 cm. During the study, standard agricultural practices were implemented. In the present experimental study, a total of thirty diverse Asian cotton genotypes were selected for the evaluation of morphological traits.

### Phenotyping for morphological and fibre quality traits

For each genotype in the study, five competitive plants were randomly selected and labeled within each replication. These plants were used for collecting various characters viz. days to 50% flowering (DFF), days to maturity (DM), plant height (PH; cm), monopodia per plant (MP), sympodia per plant (SM), bolls per plant (BP), boll weight (BW; g), seed cotton yield per plant (SCYP; g), lint yield per plant (LYP; g), ginning out turn (GOT; %), seed index (SI; g), upper half mean length (UHML; mm), fibre strength (FS; g/tex), fibre fineness, (FF; µg/inch), oil content (OC; %). Plot-based recording was conducted specifically for two phenological characters: days to 50% flowering and days to maturity.

Fibre samples (100gm) were submitted to Central Institute for Research on Cotton Technology (CIRCOT) lab, Navsari Agricultural University (NAU), Surat to assess fiber quality. The pooled fibre sample was analyzed for fibre quality traits. The testing was performed under controlled conditions of 65 ± 2% relative humidity and a temperature of 27 ± 2 °C, using the High Volume Instrument (HVI) mode. The determination of total oil content was performed through the extraction of oil using a Soxhlet apparatus and hexane (AOAC, 1965) as the extraction solvent. For the measurement of oil content, a quantity of 0.5 g of finely ground cotton seed powder (from which lint has been removed) was employed.

### Fatty acid profiling

Fatty acid profiling, also referred to as the analysis of fatty acid methyl esters (FAME), serves the purpose of assessing the quality of oil. This analysis involves the identification and quantification of the various fatty acids present within a sample. Oil profiling was determined using the procedure given by Ravi et al. (2013). To separate and detect different fatty acids from the total oil, a gas chromatograph (Thermo Trace 1110) equipped with an FID (Flame Ionization Detector) was utilized. The separation of fatty acids was performed using a SP-2560 column (100 m × 0.25 mm i.d., 0.2 µm). The injector temperature was set at 230 °C. The oven temperature program was as follows: starting at 120 °C (held for 2 min), ramping at 4 °C per minute to 210 °C (held for 3 min), and then further increased at 7 °C per minute to 230 °C (held for 10 min). Nitrogen gas served as the carrier gas at a constant flow rate of 1.5 mL/min. The detector temperature was maintained at 250 °C. The total run time for the analysis was 40 min. For each compound, peak identification was achieved by comparing retention times with standards. The peak area was automatically calculated by integrating the peak with baseline-to-baseline measurements. The concentration of each compound was determined using an equation based on the ratio of peak areas between the respective standard and the sample. The fatty acid profiling in pooled samples was carried out without replication.

### Molecular diversity using SSR markers

The molecular diversity investigation was undertaken using SSR markers among 30 genotypes of *desi*cotton and 11 other accessions of different *Gossypium* species of cotton. DNA was extracted from newly expanded fresh leaves using Cetyl Trimethyl Ammonium Bromide (CTAB) protocol (Doyle and Doyle 1990). The quality assessment of DNA was performed on a 0.8% agarose gel, while the quantity was measured using a Nanodrop ND-1000 (Thermo, USA). Then, the DNA was diluted to a concentration of 25 ng/uL using nuclease-free water.

To conduct the genetic analysis, SSR (Simple Sequence Repeat) primers developed by Jena et al. (2012) were employed for PCR (Polymerase Chain Reaction). The PCR reaction conditions involved an initial denaturation step at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at a temperature specific to the primers (ΔT °C) for 45 s, extension at 72 °C for 45 s, and a final extension step at 72 °C for 7 min. The PCR was carried out using a SensoQuest Thermocycler (Germany). Electrophoresis and gel documentation were performed following the procedure described by Sagar et al. (2022).

### Data analysis

The data collected for the studied traits were utilized to conduct an analysis of variance (ANOVA) and calculate the critical difference (CD) to identify significantly different genotypes. For SSR marker data, the size of each amplified band for every microsatellite marker was determined by comparing its migration to a molecular weight size marker (100 bp DNA Ladder). A morpho-biochemical and fiber traits-based dissimilarity matrix was constructed using Manhattan coefficients using NTSYSpc 2.0 (Rohlf, 1998). The data obtained from the allele molecular weight analysis was processed using

Power Marker v 3.25 software (Liu and Muse 2005) to calculate major allele frequency, gene diversity, heterozygosity, and polymorphism information content (PIC). The processed data obtained from this analysis was then utilized to calculate the dissimilarity matrix using DARwin 6.0 (Perrier et al. 2003). For constructing the dendrogram, a pairwise dissimilarity matrix was generated using the simple matching coefficient. The Neighbor-Joining method, as proposed by Saitou and Nei (1987) and implemented in DARwin 6.0, was used for this purpose. The method employs the relative neighborhood criterion, the weighted average for dissimilarity updating, and an adjustment to an additive tree distance to build the dendrogram. Mantel statistic based on Spearman's rank correlation was carried out between morphological and SSR markers based distance matrices in R using vegan (Oksanen et al., 2024). Since, fatty acid profiling was carried out with replication hence only mean and range was calculated.

## Results and discussion

Genetic variability refers to the sum of genetic distinctions present among individuals within a population. A diversity study involves quantitatively estimating these genetic differences. Such studies provide valuable insights into the level of genetic diversity within the population. This understanding helps to better comprehend the genetic composition and supports research in fields like biology, evolution, and breeding.

The analysis of variance revealed that the mean square due to genotypes was highly significant for all 15 characters studied, indicating a substantial amount of variability among the genotypes (Table 1). This abundance of variation opens up numerous research opportunities for plant breeders to select elite and superior genotypes for further crop improvement. The wide range of diversity observed among the genotypes offers promising prospects for enhancing crop traits and achieving desirable agricultural outcomes.

### Morphological traits

Early maturing genotypes offer significant advantages, especially in regions with frequent terminal droughts. They can also evade high pest or disease occurrences, adapt to abnormal weather conditions, and free up the field for timely sowing of the subsequent crop. Early flowering in cotton is favored, indicated by days to 50% flowering ranging from 60.67 to 77.00 days (mean 66.81). PA 826 and CNA 1065 were earliest (60.67 days), while Cernuum was late flowered (77.00 days), followed by AKA 9431 (75.00 days) and PAIG 8/1 (74.33 days). In the present study, the germplasm exhibited earlier flowering than observed in Sangwan et al. (2008) with an average of 80 days to 50% flowering. However, the current result is comparable to Ranjan et al. (2014), where the mean days to 50% flowering were 57 days.

The role of plant height in cotton is not well-defined, as it depends on the specific objective. For machine picking,

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

| Trait | Source of variation and Mean squares | | | Mean | Range | S. Em.± | CD at 5% | CV% |
|---|---|---|---|---|---|---|---|---|
| | Replications (Df: 2) | Genotypes (Df: 29) | Error (Df: 58) | | | | | |
| Days to 50% flowering | 1.01 | 76.96** | 3.55 | 66.81 | 60.67–77.00 | 1.09 | 3.08 | 2.82 |
| Days to maturity | 3.38 | 36.75** | 6.03 | 154.34 | 149–160.33 | 1.42 | 4.01 | 1.59 |
| Plant height (cm) | 84.72 | 559.00** | 135.69 | 139.89 | 82.60–158.47 | 6.73 | 19.04 | 8.33 |
| Monopodia per plant | 0.21 | 2.56** | 0.11 | 3.46 | 1.93–6.00 | 0.19 | 0.54 | 9.52 |
| Sympodia per plant | 6.31 | 49.79** | 2.02 | 15.4 | 4.93–21.73 | 0.82 | 2.33 | 9.24 |
| Bolls per plant | 6.26 | 137.15** | 22.55 | 36.76 | 16.13–47.27 | 2.74 | 7.76 | 12.92 |
| Boll weight (g) | 0.03 | 0.63** | 0.04 | 2.06 | 1.70–4.35 | 0.12 | 0.34 | 10.03 |
| Seed cotton yield per plant (g) | 75.93 | 478.80** | 68.8 | 61.2 | 40.68–80.07 | 4.79 | 13.56 | 13.55 |
| Lint yield per plant (g) | 5.45 | 76.49** | 8.02 | 19.29 | 10.73–28.73 | 1.64 | 4.63 | 14.68 |
| Ginning out turn (%) | 0.84 | 36.56** | 2.05 | 31.29 | 23.79–37.83 | 0.83 | 2.34 | 4.57 |
| Seed index (g) | 0.2 | 1.70** | 0.11 | 5.99 | 5.00–7.83 | 0.2 | 0.55 | 5.66 |
| Upper half mean length (mm) | 0.96 | 13.81** | 1.09 | 25.8 | 20.00–29.57 | 0.6 | 1.7 | 4.04 |
| Fibre strength (g/tex) | 0.84 | 12.88** | 1.84 | 25.86 | 20.07–29.30 | 0.78 | 2.21 | 5.24 |
| Fibre fineness (µg/inch) | 0.28 | 1.79** | 0.1 | 3.88 | 3.00–7.27 | 0.18 | 0.51 | 7.98 |
| Oil content (%) | 0.29 | 3.64** | 0.32 | 16.07 | 14.31–17.68 | 0.33 | 0.93 | 3.55 |

** indicate significant at 1% levels of probability

shorter sympodial varieties are preferred. However, under poorly managed conditions, taller genotypes with more monopods may be more suitable. In general, a plant height of around 150 cm is reflected best suitable (Sangwan et al. 2008). Plant height exhibited substantial variation among genotypes [82.60 (Cernuum) to 158.47 cm (PAIG 379)], with a mean of 139.89 cm. Generally, dwarf plants are well-suited for high-density planting, leading to increased yield potential.

Monopodia (vegetative branches) and sympodia (fruiting) branches are crucial in cotton as they determine plant architecture, branching pattern, and fruiting efficiency. Monopodia contribute to the main stem and primary branches, while sympodia form secondary and fruiting branches. Balancing their development and management is vital for maximizing crop yield and overall cotton productivity. Moreover, monopodial genotypes can be sown with wider spacing, reducing the need for costly hybrid seeds and subsequently lowering production costs. Generally, a lower number of monopodia per plant is preferred. Mean data for monopodia ranged from 1.97 (PA 812) to 6 (CINA 310) had the maximum. The overall mean for monopodia per plant was 3.46. The sympodia per plant varied from 4.93 to 21.73 with mean value of 15.40. The genotypes PAIG 379 had the highest sympodia per plant (21.73) while lowest value was recorded for the genotype Cernuum (4.93). Of 30 genotypes, 15 had higher for sympodia per plant than mean. The per plant monopdia and sympodia was higher than Erande et al. (2014), where mean was 1.29 and 15.88, respectively.

Bolls per plant significantly contribute to cotton yield. The 30 genotypes showed a wide variation in bolls per plant (16.13 to 47.27). PA 255 (47.27) had the most, followed by CNA 1065 (43.53). Cernuum (16.13) and AKA 9431 (23.93) had the lowest values. Low bolls per plant for race cernuum have also reported earlier by Mohan et al. (2001). The general mean was 36.76, with 22 genotypes surpassing this mean. The studied genotypes showed more bolls per plant than Ranjan et al. (2014) and Erande et al. (2014) where balls per plants were 34.14 and 14.76, respectively. The seed cotton yield is determined by the balance between flower production and the shedding of flowers or bolls. Prioritizing genotypes with a higher number of effective bolls is essential for maximizing cotton yield.

Genotypes with higher boll weight play a crucial role in contributing to cotton yield as well as easy picking with less waste content. Larger boll weight signifies the potential for more cotton fiber production, resulting in increased overall yield per plant and per unit area. Selecting and breeding for genotypes with higher boll weight is an essential strategy for improving cotton yield and productivity. The range of boll weight varied from 1.70 to 4.35 g with the mean 2.06 g. Genotype Cernuum showed exceptionally high average boll weight (4.35 g). On the other hand, genotypes GSav-1056 (1.70 g) and AKA 9431 (1.71 g) exhibited low boll weight.

The range of mean values for days to maturity varied from 149.00 to 160.33 days and general mean was 154.34 days. Among all the accessions, PA 826 and PA 869 (149.00 days) were the earliest in days to maturity followed by PA 778 (149.33) and PA 741 (150.00 days). PAIG 8/1 was found to be the latest genotype (160.33 days) in terms of days to maturity followed by Cernuum (159.33 days) and AKA 9431 (159.00 days). The trends are similar to that of days to 50% flowering but no exact literature was obtained to support the results of present study for days to maturity.

#### Yield traits

The main objective of cotton breeders is to enhance the development of varieties and hybrids with higher yield potential while ensuring acceptable fiber characteristics. This focus on yield improvement and fiber quality has two main objectives. First, it aims to enhance cotton production to meet the demands of the textile industry. Second, it seeks to ensure agricultural sustainability.The knowledge of variability for yield, a complex polygenic trait, is essential to maximize the crop harvest. In current study, the seed cotton yield per plant exhibited a wide range of variability, ranging from 40.68 to 80.07 g among all genotypes. PA 255, CINA 1067, and PA 842 were among the highest yielding genotypes, while GSav-1056, PAIG 8/1, and AKA 9431 had the lowest yields. Thirteen out of 30 genotypes surpassed the mean yield of 61.2 g per plant. The seed cotton yield per plant was lower than Ranjan et al. (2014) where it was 72 g. The seed cotton yield per plant in the study of Mankar et al. (2021) ranged from 22.00 to 45.50 g with a mean of 33.49.

The lint yield per plant in cotton holds immense importance as it directly influences the overall productivity and profitability of cotton cultivation. Maximizing lint yield per plant ensures higher fiber production, meeting the demands of the textile industry and generating better economic returns for farmers. Selecting and breeding cotton varieties with improved lint yield is a key strategy for enhancing cotton production and meeting global cotton demands. Lint yield per plant ranged from 10.73 to 28.73 g. PA 842 (28.73 g) ranked first, followed by PA 826 (28.20 g) and PA 255 (26.29 g). Conversely, GSav-1056 (10.73 g) and AKA 9431 (11.31 g) had the lowest yields. The overall mean for lint was 19.29 g. A wide range of variability for lint index (3.21–6.37 g) was also recorded by Mankar et al. (2021). Erande et al. (2014; 10.90 g) observed lower lint yield.

Ginning out turn (GOT) is a crucial factor in the cotton industry as it directly impacts the amount of lint obtained after ginning. A high GOT is desirable for cotton growers as it increases the market value of lint and improves profitability. The ginning out turn (GOT) displayed a range from 23.79% to 37.83%. PA 826 recorded the highest GOT (37.83%), while PA 741 had the lowest (7.80%). The overall mean for GOT was 31.29%. The GOT in previous studies of Erande et al., (2014) and Iqbal et al. (2015), GOT ranged from 33.61% to 39.87% (mean: 37.71%) and 18.75% to 36.94% (mean: 28.95%), respectively. These variations highlight the importance of GOT assessment in cotton, as it directly influences lint yield and overall cotton quality.

Seed index is important in determining yield and oil content. The mean seed index of the 30 genotype was 5.99 g, ranging from 5.00 g (PA 255) to 7.83 g (AKA 9431). These results are in agreement with Sagar et al. (2022; 6.09 g) and Saravanan et al. (2021; 5.29 g). The larger-seed have reduced surface area-to-seed mass ratio, which leads to lower seed lint. Thus, seed index is a contrasting trait in cotton, and a lower seed index is considered favorable.

#### Fibre quality parameters

Cotton Upper Half Mean Length (UHML) is a critical measure in the textile industry to evaluate the quality of cotton fibers. Longer UHML values are desirable, as they indicate the presence of more consistent and uniform fibers, which are ideal for producing high-quality textiles. Cotton with UHML below 24 mm may have reduced spinning performance. Cotton-based industries favor longer fiber lint. In current study, UHML ranged from 20.00 to 29.57 mm. Genotype 824 recorded the highest UHML (29.57 mm), while Cernuum had the lowest (20 mm). The overall UHML mean was 25.80 mm. Earlier Mohan et al. (2001) reported 2.5% span length of 17.0–21.8 mm. A UHML in a range of 12.91% -23.81% was recorded by Iqbal et al. (2015) in *G. arboreum* of Pakistan origin. However, in a previous study conducted by Wang et al. (2013), a higher mean value of 35.35 mm for UHML was recorded for barbadense cotton. Variability in UHML values can be attributed to different cotton species and environmental conditions.

Cotton varieties with superior tensile strength are preferred by the textile industry for producing durable and high-performance textiles. Earlier, fiber strength in a range of 17.50–20.50 g/tex with a mean of 19.27 g/tex was reported by Erande et al. (2014) in *G. arboretum* of Indian origin. Similarly, Mankar et al. (2021) fiber strength was between 18.40 and 28.60 g/tex (mean: 25.54 g/tex). In current study the fibre strength varied from 20.07 (Cernuum) to 29.30 (PA 812) g/tex. The highest fibre strength was found in accession PA 812 (29.30 g/tex). The overall mean (25.99 g/tex) was similar to earlier studies.

Micronaire is a critical measure of fiber fineness, with a lower value being desirable. It significantly impacts the texture of the fiber, distinguishing between soft and coarse, and silky and harsh attributes, which can be affected by environmental factors. Maximum micronaire value was observed in genotype Cernuum (7.27 µg/inch) which had exceptionally high fibre fineness as compared to other genotypes under study. Minimum micronaire value was observed in genotype PA 741 (3 µg/inch). This genotype considered as a fine fibre. The overall mean for fibre fineness was 3.88 µg/inch. With a mean of 5.47 µg/inch, Iqbal et al. (2015) reported micronaire value between 4.37 and 6.79 µg/inch. The mean value of current study was significantly lower than Erande et al. (2014) and Mankar et al. (2021).

#### Oil content and fatty acid profiling

Despite cotton being primarily cultivated for fiber production, the whole cottonseed serves as the important source of vegetable oil (Liu et al. 2002). The oil extracted from cottonseed is rich in nutritional value and finds applications in the food industry as edible oil. According to Shahrajabian et al. (2020), cotton ranks as the sixth major source of edible oil worldwide. Its high tocopherol content makes cottonseed oil valuable for human consumption and various food applications. With a mean of mean 16.07%, oil in current study ranged from 14.31 (PA 873) to 17.68% (Cernuum). Earlier, Jogender et al. (2023) recorded oil content in a range of 14.20–15.60% with a mean of 14.87% (Table 1). Kulkarni et al. (2009) also reported a similar range of values, with an oil content of 20.4%.

Cottonseed oil is renowned as a cholesterol-free "heart oil" and is highly regarded in the world of edible oils. Its unique taste and cooking qualities are attributed to specific ratios of saturated and unsaturated fatty acids. The relative content of various fatty acids influences the quality and palatability of cottonseed oil. Cotton seed oil consists of various saturated and unsaturated fatty acids in variable proportions.

The major fatty acids detected in the oil of 30 *desi* cotton genotypes were linoleic, palmitic and oleic, with an average of 43.22%, 25.13% and 24.97%, respectively (Table 2). The fatty acids detected in minor quantities were α-linolenic, arachidic and myristic with an average of 0.17%, 0.30% and 0.42%, respectively. Myristic (14:0), palmitic (16:0), palmitoleic (16:1), oleic (18:1), linoleic (18:2) and α-linolenic (18:3) were detected in all the 30 genotypes, while the fatty acids namely arachidic (20:0),

**Table 2** Range and detection of various fatty acids in *desi* cotton

| Fatty Acid | Range (%) | | Fatty acid detected in number of genotypes |
|---|---|---|---|
| | Min | Max | |
| Myristic acid (14:0) | 0.29 | 0.69 | 30 |
| Palmitic acid (16:0) | 21.73 | 26.82 | 30 |
| Palmitoleic acid (16:1) | 0.75 | 1.54 | 30 |
| Stearic acid (18:0) | 2.87 | 4.43 | 17 |
| Oleic acid (18:1 Cis-9) | 15.72 | 31.48 | 30 |
| *Cis*-vaccenic acid (18:1 Cis-7) | 1.75 | 4.8 | 12 |
| Linoleic acid (18:2) | 36.78 | 54.05 | 30 |
| α-Linolenic acid (18:3) | 0.08 | 1.16 | 30 |
| Arachidic acid (20:0) | 0.1 | 0.5 | 29 |

stearic (18:0) and *cis*-vaccenic (18:1) were detected in 29, 17 and 12 genotypes, respectively.

Linoleic acid was the highest among all fatty acids and was detected in the range of 36.78% – 54.05% (J. Tapti). Palmitic acid was the second highest fatty acid after linoleic acid and was recorded in the range of 21.73–26.82% (PAIG 27). Oleic acid varied in the range of 15.72%—31.48% (PAIG 379) (Table 3). Previously, Garg et al. (2003) obtained similar results where linoleic acid, palmitic acid and oleic acid had the ranges of 38.43%—55.14%, 24.08% – 36.33% and 15.95% – 33.70% respectively in *desi* cotton.

**Table 3** Oil profiling of different *desi* cotton genotypes (ND: not detected)

| Genotypes | Myristic acid (C 14:0) (%) | Palmitic acid (C 16:0) (%) | Palmitoleic acid (C 16:1) (%) | Stearic acid (C 18:0) (%) | Oleic acid (C 18:1 Cis-9) (%) | *cis*-vaccenic (C 18:1 Cis-7) (%) | Linoleic acid C 18:2 (%) | α-inolenic acid C 18:3 (%) | Arachidic acid C 20:0 (%) |
|---|---|---|---|---|---|---|---|---|---|
| J. Tapti | 0.69 | 25.18 | 1.03 | 3.14 | 15.72 | ND | 54.05 | 0.1 | 0.1 |
| CINA 310 | 0.43 | 22.63 | 0.92 | 4.43 | 22.14 | ND | 49 | 0.11 | 0.32 |
| AKA 9431 | 0.58 | 25.08 | 0.91 | ND | 23.05 | ND | 43.84 | 0.19 | 0.29 |
| PAIG 8/1 | 0.67 | 21.73 | 0.89 | 2.05 | 17.2 | ND | 43.98 | 1.16 | ND |
| PAIG 27 | 0.47 | 26.82 | 1.19 | ND | 22.38 | 4.26 | 44.37 | 0.17 | 0.35 |
| DLSA 17 | 0.46 | 22.7 | 0.75 | ND | 26.27 | 4.8 | 44.66 | 0.12 | 0.24 |
| DLSA 24 | 0.38 | 25.69 | 1.38 | 2.87 | 25.74 | ND | 43.59 | 0.11 | 0.25 |
| PA 255 | 0.69 | 26.28 | 1.06 | 2.93 | 17.55 | ND | 51.08 | 0.14 | 0.26 |
| 824 | 0.34 | 25.87 | 1.32 | 3.6 | 27.81 | ND | 40.61 | 0.13 | 0.31 |
| Cernuum | 0.4 | 26.07 | 1.43 | ND | 28.26 | 3.52 | 39.27 | 0.19 | 0.5 |
| AKA 7 | 0.33 | 25.65 | 1.17 | 3.5 | 27.59 | ND | 41.31 | 0.12 | 0.33 |
| Gsav-1056 | 0.38 | 23.83 | 1.47 | ND | 30.97 | 3.82 | 39.08 | 0.13 | 0.34 |
| SRT GMS-1 | 0.36 | 25.95 | 1.35 | ND | 28.17 | 3.76 | 39.96 | 0.13 | 0.32 |
| PA 741 | 0.54 | 24.56 | 0.86 | 3.99 | 21.11 | ND | 48.47 | 0.14 | 0.33 |
| PA 778 | 0.38 | 25.92 | 1.36 | ND | 28.05 | 3.99 | 39.82 | 0.12 | 0.36 |
| PAIG 373 | 0.39 | 26.08 | 1.3 | ND | 28.19 | 3.87 | 39.73 | 0.14 | 0.3 |
| PA 796 | 0.31 | 25.65 | 1.17 | 3.35 | 24.69 | ND | 44.45 | 0.08 | 0.29 |
| PA 812 | 0.29 | 26.24 | 1.2 | 3.33 | 25.52 | ND | 43.02 | 0.1 | 0.28 |
| PA 833 | 0.35 | 26.17 | 1.33 | 3.08 | 24.7 | ND | 44.06 | 0.11 | 0.2 |
| PA 842 | 0.32 | 25.03 | 1.25 | 2.98 | 24.2 | ND | 45.91 | 0.13 | 0.19 |
| CNA 1054 | 0.36 | 25.38 | 1.14 | ND | 26.55 | 3.62 | 42.56 | 0.12 | 0.28 |
| PAIG 377 | 0.33 | 25.37 | 1.3 | 3.41 | 27.29 | ND | 41.91 | 0.13 | 0.27 |
| PA 826 | 0.3 | 23.42 | 1.2 | 3.03 | 23.8 | ND | 47.82 | 0.13 | 0.31 |
| CNA 1065 | 0.38 | 25.11 | 1.16 | 3.63 | 26.11 | ND | 43.21 | 0.11 | 0.29 |
| PA 873 | 0.33 | 25.11 | 1.25 | 3.15 | 25.03 | ND | 44.8 | 0.12 | 0.21 |
| PAIG 379 | 0.45 | 25.14 | 1.54 | ND | 31.48 | 4.11 | 36.78 | 0.16 | 0.34 |
| PA 869 | 0.3 | 24.5 | 1.22 | 3.32 | 25.27 | ND | 44.91 | 0.14 | 0.33 |
| CINA 1068 | 0.45 | 26.12 | 1.45 | ND | 31.46 | 1.75 | 38.26 | 0.17 | 0.34 |
| CINA 1069 | 0.45 | 25.46 | 1.47 | ND | 31.07 | 2.62 | 38.44 | 0.17 | 0.33 |
| CINA 1067 | 0.42 | 25.08 | 1.47 | ND | 30.8 | 4.23 | 37.51 | 0.15 | 0.35 |
| Average | 0.42 | 25.13 | 1.22 | 3.28 | 24.97 | 3.69 | 43.22 | 0.17 | 0.3 |

### Germplasm clustering using morpho-biochemical and fiber traits

The Manhattan dissimilarity coefficient based clustering partitioned 30 genotypes into four clusters at a cut-off value of 0.17 (Fig. 1, Table 4) indicating susbtantial phenotypic variability. The dissimilarity between genotypes ranged from 0.054 (PAIG 373 and PA812) to 0.54 (PA873 and Cernuum). Cluster 1 and 2 had 8 genotypes each while cluster 3 and 4 had 13 and 1 genotype(s), respectively. Cluster 3 was characterized with high sympodia per plant, bolls per plant, seed cotton yield per plant, lint yield per plant and ginning out turn. Cluster 4 with single genotype was superior for oil content, seed index and boll weight. Of 4 CINA series genotypes, 3 grouped together in cluster 3, while PAIG series genotypes were scattered in 3 different clusters. This suggested that PAIG genotypes are more diverse at morphological traits that CINA genotypes. The cluster mean values of different traits are presented in Table 4. The SCYP was maximum (60.16 g) in cluster 2 which was also had higher bools per plant and Sympodia per plant. The HSD test further confirmed that the cluster means are significantly different for most of the traits studied (Fig. 2). Among all genotypes, cernnum was highly distinct as values for all studied traits for this genotype was very far from mean values of clusters. Genotypes with superior fiber quality traits can be directly released as varieties. Alternatively, they can be used as donor parents in hybridization after multi-year evaluation of their yield performance across

**Table 4** Variability for mean values of 15 traits in four clusters identified by clustering analysis of 30 genotypes

| Trait | Cluster and genotypes per cluster | | | |
|---|---|---|---|---|
| | I (8) | II (8) | III (13) | IV (1) |
| Days to 50% flowering | 72.04 | 65.13 | 63.85 | 77.00 |
| Days to maturity | 157.79 | 153.21 | 152.54 | 159.33 |
| Plant height (cm) | 143.42 | 139.74 | 142.22 | 82.60 |
| Monopodia per plant | 4.37 | 3.52 | 2.97 | 2.13 |
| Sympodia per plant | 11.48 | 15.69 | 18.44 | 4.93 |
| Bolls per plant | 31.12 | 37.36 | 41.46 | 16.13 |
| Boll weight (g) | 1.84 | 1.96 | 2.09 | 4.35 |
| Seed cotton yield per plant (g) | 47.93 | 60.16 | 71.26 | 45.07 |
| Lint yield per plant (g) | 14.62 | 16.74 | 24.08 | 14.75 |
| Ginning out turn (%) | 30.31 | 27.85 | 33.89 | 32.65 |
| Seed index (g) | 6.74 | 6.04 | 5.43 | 6.83 |
| Upper half mean length (mm) | 24.46 | 28.32 | 25.51 | 20.00 |
| Fibre strength (g/tex) | 24.31 | 27.61 | 26.49 | 20.07 |
| Fibre fineness (µg/inch) | 3.80 | 3.91 | 3.88 | 4.27 |
| Oil content (%) | 16.27 | 16.10 | 15.86 | 16.78 |

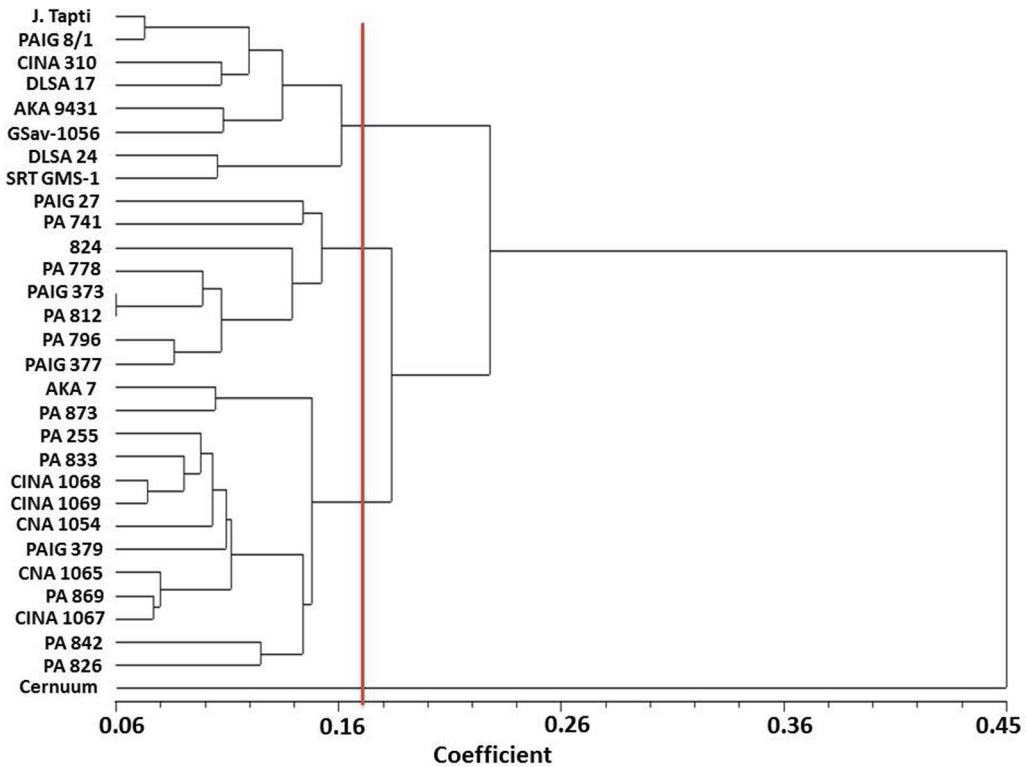


**Fig. 1** Dendrogram showing relationship among 30 *Gossypium arboreum* genotypes generated using Manhattan distance

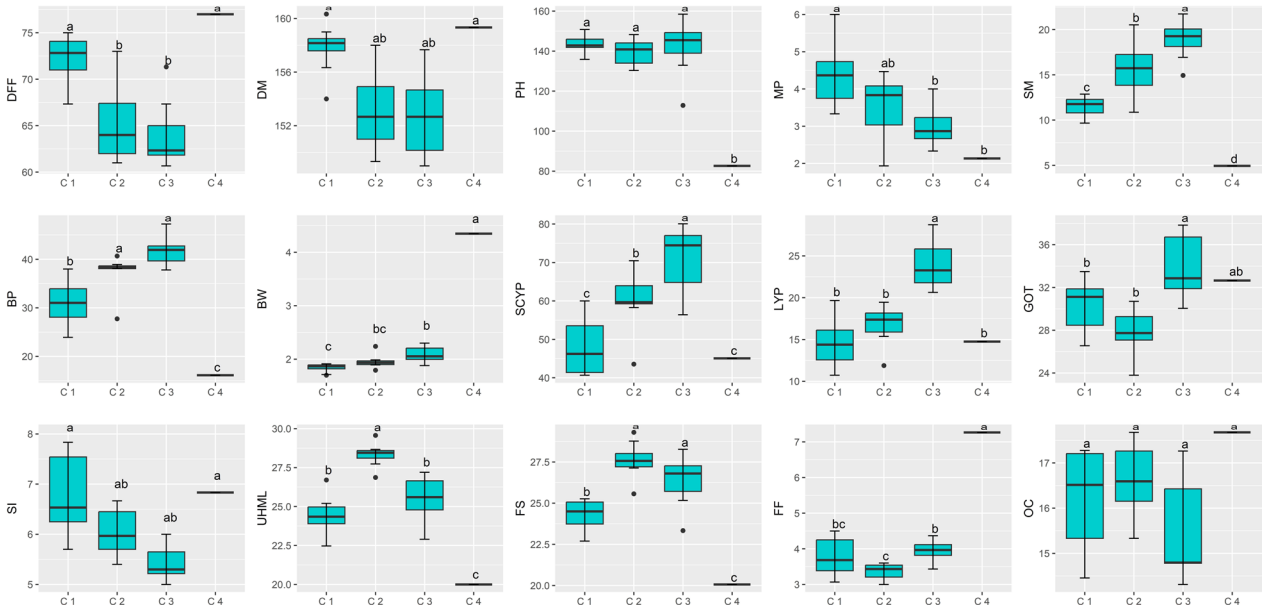


**Fig. 2** Cluster mean comparison of various traits with HSD Test and whisker plot visualization. DFF: Days to 50% flowering; DM: Days to maturity; PH: Plant height (cm); MP: Monopodia per plant; SM: Sympodia per plant; BP: Bolls per plant; BW: Boll weight (g); SCYP: Seed cotton yield per plant (g); LYP: Lint yield per plant (g); GOT: Ginning out turn (%); SI: Seed index (g); UHML: Upper half mean length (mm); Fibre strength (g/tex); FS: Fibre fineness (µg/inch); OC: Oil content (%);

**Table 5** Top 10 genotypes for different fiber quality and oil content traits with their agronomic performance

| Trait | Name of genotypes [Bolls per plant, Seed cotton yield per plant (g)] | (Range, mean) | |
|---|---|---|---|
| | | Bolls per plant | Seed cotton yield per plant (g) |
| Lint yield per plant (g) | PA 842 (41.93, 78.40), PA 826 (43, 74.68), PA 255 (47.27, 80.07), CINA 1067 (41.2, 79.67), PA 869 (40.33, 77.87), CNA 1065 (43.53, 76.15), CINA 1069 (40.27, 67.13), CINA 1068 (42.27, 74.47), PAIG 379 (39.53, 70.53), PA 833 (42.47, 73) | 39.53–47.27, 42.18 | 67.13–80.07, 75 |
| Ginning out turn (%) | PA 826 (43, 74.68), AKA 7 (39.67, 58.92), PA 873 (37.8, 56.39), PA 842 (41.93, 78.4), CNA 1054 (39.67, 59.1), CINA 1069 (40.27, 67.13), DLSA 24 (37.33, 54.32), PA 255 (47.27, 80.07), SRT GMS-1 (38, 59.99), Cernuum (16.13, 45.07) | 16.13–47.27, 38.11 | 45.07–80.07, 63.41 |
| Seed index (g) | AKA 9431 (23.93, 41.4), DLSA 17 (32.6, 46.97), CINA 310 (32.8, 53.3), Cernuum (16.13, 45.07), 824 (38.07, 58.28), PAIG 8/1 (28.73, 41.3), DLSA 24 (37.33, 54.32), PA 741 (38.07, 67.17), PAIG 27 (27.73, 43.56), J. Tapti (29.47, 45.47) | 16.13–38.07, 30.49 | 41.30–67.17, 49.68 |
| Upper half mean length (mm) | 824 (38.07, 58.28), PA 778 (38.33, 62.9), PAIG 377 (40.67, 70.47), PAIG 373 (38.53, 59.65), PA 812 (38.53, 59.67), PA 796 (38.93, 59.57), PA 741 (38.07, 67.17), PA 873 (37.8, 56.39), PA 833 (42.47, 73), PAIG 27 (27.73, 43.56) | 27.73–42.47, 37.91 | 43.56–73, 61.06 |
| Fibre strength (g/tex) | PA 812 (38.53, 59.67), PA 778 (38.33, 62.9), PA 873 (37.8, 56.39), PAIG 379 (39.53, 70.53), CINA 1069 (40.27, 67.13), PAIG 373 (38.53, 59.65), PA 796 (38.93, 59.57), 824 (38.07, 58.28), PA 833 (42.47, 73), CINA 1068 (42.27, 74.47) | 37.8–42.47, 39.47 | 56.39–74.47, 64.16 |
| Fibre fineness | Cernuum (16.13, 45.07), PAIG 8/1 (28.73, 41.3), DLSA 17 (32.6, 46.97), CINA 1065 (43.53, 76.15), PA 869 (40.33, 77.87), CINA 1067 (41.2, 79.67), DLSA 24 (37.33, 54.32), CINA 1069 (40.27, 67.13), AKA 7 (39.67, 58.92), PA 873 (37.8, 56.39) | 16.13–43.53, 35.76 | 41.30–79.67, 60.38 |
| Oil content (%) | Cernuum (16.13, 45.07), PAIG 377 (40.67, 70.47), DLSA 17 (32.6, 46.97), AKA 9431 (23.93, 41.4), PA 741 (38.07, 67.17), PA 796 (38.93, 59.57), CINA 1065 (43.53, 76.15), CINA 310 (32.8, 53.3), PAIG 379 (39.53, 70.53), CINA 1067 (41.2, 79.67) | 16.13- 43.53, 34.74 | 41.40–79.67, 61.03 |

various locations. In this study, the top ten fiber quality rich genotypes had bolls per plant 16.13 to 47.27 and had 41.30 to 80.07 g of seed cotton yield per plant (Table 5).

#### Genetic diversity using SSR markers and Mantel test

The present investigation was undertaken to study the genetic diversity using SSR markers among 30 genotypes of *desi* cotton and 11 other accessions which consisted of different species of cotton other than *G. arboretum.* Initially, 55 SSR primers were tested on five DNA samples to identify reproducible, scorable, and polymorphic primers. Among them, 50 primers (89%) successfully amplified. Out of the 50 primers, 26 (52%) were monomorphic, and the remaining 24 (48%) were found to be polymorphic (Table 6). A representation of SSR markers amplification in cotton genotypes is shown in Fig. 3.

The molecular weight of amplicons ranged from 89 bp (NBRI_gJ018) to 280 bp (NBRI_gL029), reflecting significant differences in the number of repeats among different alleles. Among 41 cotton accessions, a total of 134 alleles were detected, with an average of 5.58 alleles per locus (range: 2 to 11).

This average number of alleles per locus is consistent with findings by Azmat and Khan (2010) with 5.76, Celik (2022) with 5.82, and Sagar et al. (2022) with 5.31 alleles per locus in various Gossypium species. However, Guang and Xiong-Ming (2006) reported a lower number of alleles per locus (3.6), and Guo et al. (2006) found even fewer alleles (2.13) using different sets of SSR markers on genotypes with varying degrees of variability.

The major allelic frequency ranged from 0.49 (NBRI_gH040) to 0.95 (NBRI_gE013-1), with a mean value of 0.70, which is consistent with the findings of Saravanan et al. (2021). The gene diversity (He) showed the highest level for NBRI_gH040 (0.67) and the lowest for NBRI_gE013-1 (0.09), with an overall mean gene diversity of 0.46 (Table 6). Santosh et al. (2021) reported similar results with an average gene diversity of 0.37 (ranging from 0.04 to 0.66), while Saravanan et al. (2021) reported gene diversity values ranging from 0.10 to 0.53 with an average of 0.36. Heterozygosity displayed significant variation, ranging from 0.00 (NBRI_gPC_10) to 0.17 (NBRI_gE005-1), with an average of 0.07. As cotton is often cross-pollinated, gene flow between populations is limited, potentially resulting in reduced genetic diversity and increased homozygosity, preventing the combination of different gene pools. This process can have implications for the overall genetic variability and adaptability of cotton crops.

The PIC values of SSR markers in the present study ranged from 0.09 (NBRI_gE013-1) to 0.62 (NBRI_gH040), with an average of 0.42. The markers were categorized based on their informative power: NBRI_gE013-1 and NBRI_gM175 were uninformative (PIC < 0.3), while NBRI_gH040 was highly informative (PIC > 0.3). The remaining 21 markers fell within the range of moderately informative markers (PIC: 0.3 to 0.6). The average PIC value (0.42) in the current study was higher compared to other studies. Sethi et al. (2015) reported an average PIC value of 0.21 (0.02–0.80) in *desi* cotton, Saravanan et al. (2021) reported an average PIC value of 0.29 (0.09–0.45), and Santosh et al. (2021) reported an average PIC value of 0.31 (0.04–0.6). On the other hand, Guo et al. (2006) obtained a higher average PIC value of 0.89 (0.52–0.98) than the present study. The discrepancies in PIC values can be attributed to the use of different genotypes and markers in the respective studies. The high PIC values and a large number of alleles per marker could also be influenced by the nature of the materials studied (Saiyad and Kumar 2018). The diverse germplasm used in the current study may have contributed to the observed variations in PIC values, highlighting the significance of genotype selection for genetic diversity studies.

The results from the molecular studies have provided substantial evidence that the SSR markers successfully amplified not only the DNA of *desi* cotton genotypes but also that of all other cotton species considered in the study. Table 7 displays the amplicon sizes of the *desi* cotton genotypes and other cotton species, along with the number of alleles obtained. The inclusion of diverse cotton species has significantly contributed to a comprehensive understanding of the genetic diversity among cultivated and wild cotton species.

The reliability of the SSR markers used for molecular analysis is further validated by their successful amplification in all cotton species. This verification confirms the robustness and suitability of the chosen primers for accurately assessing genetic diversity in the various cotton species under investigation. The study's comprehensive approach and the use of well-performing SSR markers have enhanced our insights into the genetic relationships and variations within the cotton genus, making it a valuable contribution to the field of cotton genetics and breeding research.

The 41 cotton accessions were classified into three primary clusters: A, B, and C, comprising 23, 17, and 1 genotype, respectively (Fig. 4). Cluster A exclusively consisted of *desi* cotton genotypes, further subdivided into two subclusters, A1 (22 genotypes) and A2 (1 genotype). Subcluster A1 was divided into two groups, one containing genotype Cernuum and the other consisting of 21 additional genotypes. The solitary genotype in A2 was PA 255, which demonstrated superior morphological performance in terms of seed cotton yield per plant. Cluster B comprised 7 *desi* cotton genotypes and 10 other cotton species, split into two subclusters, B1 (16 genotypes)

**Table 6** Allele information, diversity index and PIC value of SSR primers

| Locus name | 5′-3′ Primer sequence (Forward/reverse) | Amplicon size (bp) | Number of alleles | Major allele frequency | $H_e$ | $H_1$ | PIC |
|---|---|---|---|---|---|---|---|
| NBRI_gA112 | CCTATCTTCAAGCAATACAGC/ TTAAGCCACGTTCTTTACATGAAC | 175–199 | 9 | 0.66 | 0.54 | 0.07 | 0.52 |
| NBRI_gB008 | AAATTCACGATCACCATCTTC/ CTCCCTTGCCTTCCTAGC | 128–145 | 6 | 0.71 | 0.47 | 0.05 | 0.44 |
| NBRI_gD111 | TATGGCCTATGGCACTATTAA/ ACTCGTTGCTAGTTGAATCATG | 145–176 | 8 | 0.68 | 0.51 | 0.05 | 0.48 |
| NBRI_gE005-1 | ATGCCATGTGCCTGATAAGA/ CCTCAACATGGATTCTTGGTC | 117–130 | 6 | 0.72 | 0.46 | 0.17 | 0.44 |
| NBRI_gE013-1 | CCAAGTTCTTTCTACTTGCTGTTG/ TCACATGCGTGAACACTTTG | 185–200 | 4 | 0.95 | 0.09 | 0.05 | 0.09 |
| NBRI_gF012 | GAGTTGAGGGTCTGGGATTG/ ACCGGGTTACCTGACTAGCAA | 118–135 | 4 | 0.72 | 0.44 | 0.02 | 0.40 |
| NBRI_gF025 | TCTTAAGCTGTGAAGTCGATGG/ TGAAGAACATGATGCTGAAGG | 139–169 | 6 | 0.65 | 0.54 | 0.07 | 0.51 |
| NBRI_gH040 | CGGAACTCAAACACGAGTCA/ CTCGCTTCTCATCACTTTTGG | 132–154 | 6 | 0.49 | 0.67 | 0.17 | 0.62 |
| NBRI_gJ018 | CAACTTTCCGAGCTGGATTC/ AATTCCCCCACAAGAAAACC | 89–104 | 2 | 0.67 | 0.44 | 0.02 | 0.34 |
| NBRI_gJ036 | CCTTTTGTGGGGTGGAAATA/ TTGAAGGAATTTAAAACAAGAGGAA | 154–167 | 3 | 0.71 | 0.43 | 0.02 | 0.37 |
| NBRI_gL006 | TTGGATGAAATCGAAAAA/ GGGTGGAAAAGTGGAGGAAT | 152–165 | 5 | 0.74 | 0.41 | 0.12 | 0.37 |
| NBRI_gL029 | CAAAAGCCTTTCTTGACACTGA/ GGCGACTCCACATTTTTGTT | 257–280 | 8 | 0.59 | 0.60 | 0.10 | 0.56 |
| NBRI_gPC_10 | ACATTTGTGGACTTCATTCAC/ GATCGTGCGTAGTATGAAAAT | 131–150 | 3 | 0.80 | 0.33 | 0.00 | 0.31 |
| NBRI_gPC_51-1 | GGAACCTGTTTTTCATCTGTT/ GCATCAAATTAAACAGCCATA | 129–142 | 4 | 0.80 | 0.33 | 0.00 | 0.31 |
| NBRI_gPC_52-1 | GACGGTGACGAGCAAAGT/ GGTTCACTGCAGAAAGTAAGA | 154–161 | 3 | 0.85 | 0.26 | 0.12 | 0.24 |
| NBRI_gPC_56 | AATAGCCGAAGACTCTGTTTT/ TACTAGTTCCATTGGTTCAGC | 167–180 | 8 | 0.57 | 0.60 | 0.05 | 0.55 |
| NBRI_gPC_67 | AATAAATGGTGAGGAAGGAAA/ CCATAATGGAAACCCAAAATA | 127–143 | 6 | 0.65 | 0.54 | 0.12 | 0.51 |
| NBRI_gPC _72-1 | CCTTTACCTCTTATCTAACCCTA/ GAAGAGTTGGATTTTGTGTGA | 125–171 | 8 | 0.63 | 0.56 | 0.05 | 0.53 |
| NBRI_gPD_27-1 | ACATCAACATCAGAGGAAGTG/ GGACGGAAAATTAAAAACTGT | 169–191 | 5 | 0.57 | 0.60 | 0.15 | 0.56 |
| NBRI_gPD_61 | GAAATCGTAGGTGATTTCCTT/ AAAGCACACGAGATTTGTTTA | 144–156 | 4 | 0.63 | 0.53 | 0.02 | 0.48 |
| NBRI_gPD_8 | TTATAAAGATTGGAGCCTGGT/ GGACATAGCTTTTCCCATTAT | 143–153 | 4 | 0.82 | 0.32 | 0.12 | 0.30 |
| NBRI_gPD_88-1 | ATTCCCAAAAGGAAATGAGTA/ AAGAAATTTAGCCTCATTCGT | 149–164 | 8 | 0.70 | 0.50 | 0.10 | 0.48 |
| NBRI_gM6 | CGTTAATTAATAGCCATTTCG/ ATACAACGAGAACCCTCTCTC | 145–187 | 11 | 0.63 | 0.57 | 0.10 | 0.55 |
| NBRI_gM175 | GCAGAGTCTTGAATGAGAGAA/ TCCTGTATTTCAATTCCAAAG | 155–161 | 3 | 0.91 | 0.16 | 0.02 | 0.15 |
| Total | | | 134 | – | – | – | – |
| Average | | | 5.58 | 0.71 | 0.46 | 0.07 | 0.42 |

and B2 (1 genotype). Cluster C included a single accession, namely *G. davidsonii*. The clustering analysis has effectively grouped the cotton accessions based on their genetic similarities, facilitating a better understanding of the relationships among different cotton genotypes and species. The clustering behavior of cernuum was same in

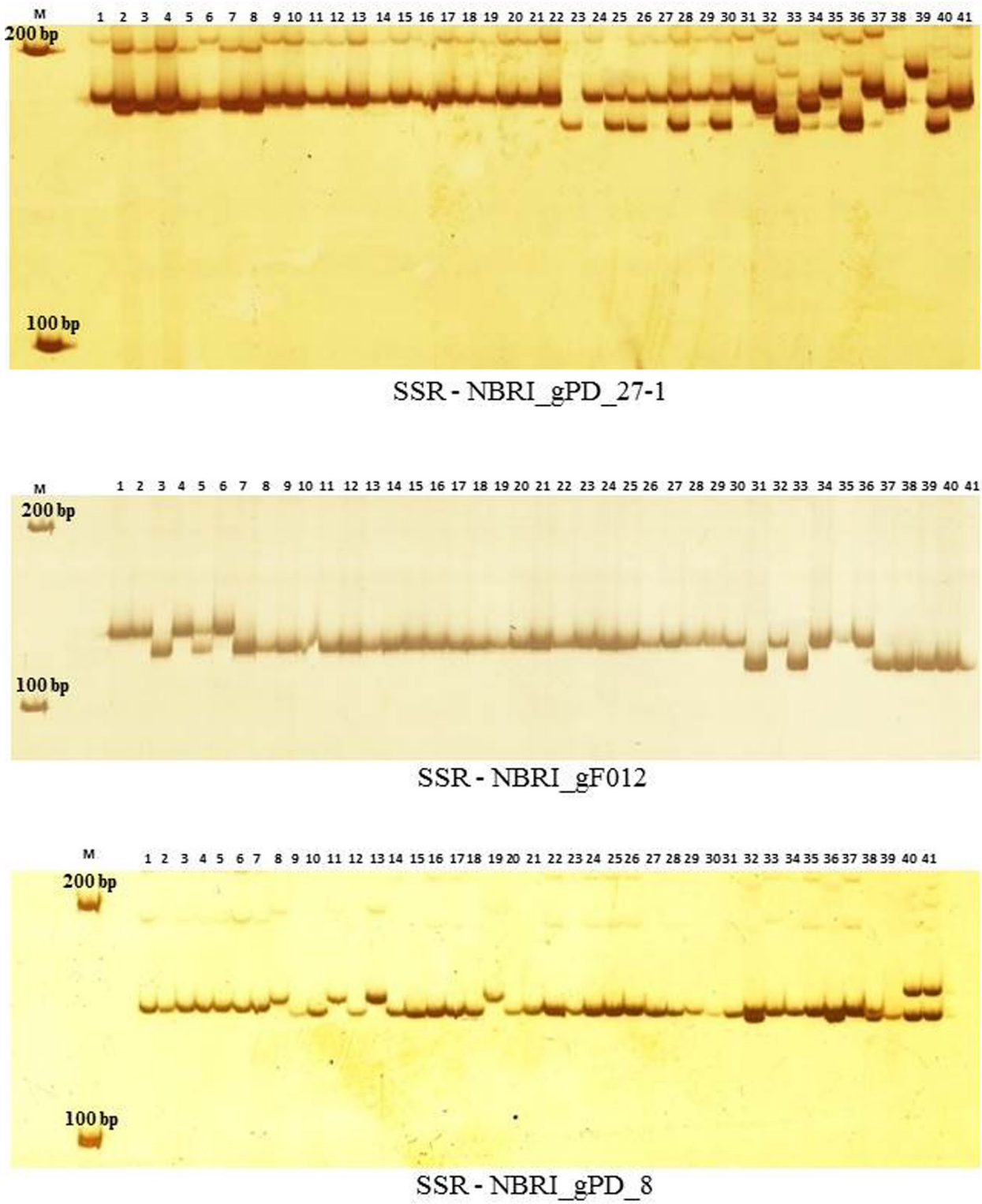


**Fig. 3** Amplification profiles of SSR markers; M: ladder, Lane 1 to 41: *Gossypium* genotypes

**Table 7** Amplicon size (base pairs) of SSR markers in 12 different species of cotton (value in parenthesis: total bands in each species)

| Markers | *G. arboreum* | *G. herbaceum* | *G. barbadense* | *G. stocksii* | *G. davidsonii* | *G. robinsoni* | *G. capitis-viridis* | *G. sturtianum* | *G. barbadense* var. *braziliense* | *G. thurberi* | *G. raimondii* | *G. hirsutum* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NBRI_gA112 | 178–193 (31) | 183 | 183 | 175–183 | 183 | 183 | 180 | 180 | 180–194 (2) | 199 | 180 | 178 |
| NBRI_gB008 | 137–145 (31) | 137 | 128 | 137 | 130–137 (2) | 137 | 137 | 137 | 128 | 140 | 134 | 134 |
| NBRI_gD111 | 153–176 (31) | 147–165 (2) | 156 | 156 | 153 | 153 | 156 | 153 | 153 | 153 | 145 | 145 |
| NBRI_gE005-1 | 120–130 (34) | 128 | 117 | 120 | 120 | 125 | 120 | 125 | 120 | 120–130 (2) | 120–130 (2) | 120–130 (2) |
| NBRI_gE013-1 | 185–189 (30) | 189 | 189 | 189–200 (2) | 189 | 189 | 189 | 189 | 189 | 189–195 (2) | 189 | 189 |
| NBRI_gF012 | 125–135 (31) | 118 | 125 | 118 | 125 | 125 | 125 | 118 | 118 | 118 | 118 | 118 |
| NBRI_gF025 | 151–159 (31) | 159–166 (2) | 169 | 169 | 159 | 169 | 159 | 159–167 (2) | 139 | 139 | 139 | 139 |
| NBRI_gH040 | 132–154 (35) | 154 | 140 | 140 | 154 | 145 | 140 | 140 | 145 | 145 | 140 | 140 |
| NBRI_gJ018 | 89–104 (30) | 89 | 104 | 89 | 104 | 104 | 104 | 89–104 (2) | 89 | 89 | 89 | 89 |
| NBRI_gJ036 | 158–167 (30) | 158 | 158 | 167 | 167 | 158 | 158 | 167 | 158 | 167 | 154 | 154–167 (2) |
| NBRI_gL006 | 160–165 (30) | 156–165 (2) | 160–165 (2) | 165 | 158–165 (2) | 152–158 (2) | 165 | 165 | 152–158 (2) | 165 | 165 | 165 |
| NBRI_gL029 | 260–280 (32) | 275 | 275 | 275 | 264 | 264 | 264 | 260 | 264 | 264 | 257 | 257 |
| NBRI_gPC_10 | 144–150 (30) | 144 | 131 | 144 | 131 | 144 | 144 | 144 | 131 | 144 | 144 | 144 |
| NBRI_gPC_51-1 | 135–139 (30) | 135 | 135 | 129 | 142 | 135 | 129 | 135 | 139 | 135 | 135 | 139 |
| NBRI_gPC_52-1 | 157 (30) | 157 | 154–161 (2) | 157 | 154–161(2) | 157 | 157 | 157 | 154–161 (2) | 154 | 154–161 (2) | 154–161(2) |
| NBRI_gPC_56 | 167–175 (31) | 175 | 171 | 168 | 172–180 (2) | 174 | 167 | 175 | 171 | 175 | 171 | 171 |
| NBRI_gPC_67 | 130–143 (30) | 130 | 127–139 (2) | 127 | 130–135 (2) | 130 | 127–139 (2) | 130 | 127–139 (2) | 132 | 127 | 127–132 (2) |
| NBRI_gPC_72-1 | 130–171 (30) | 125–138 (2) | 125 | 125 | 130 | 125 | 130 | 125–132 (2) | 130 | 125 | 125 | 125 |
| NBRI_gPD_27-1 | 169–179 (34) | 179 | 175 | 169 | 169–175 (2) | 182 | 169 | 182 | 179 | 191 | 169–179 (2) | 179 |
| NBRI_gPD_61 | 144–152 (31) | 156 | 156 | 156 | 151 | 156 | 151 | 156 | 156 | 151 | 156 | 156 |
| NBRI_gPD_8 | 148–152 (30) | 148 | 143–148 (2) | 148 | 148 | 148 | 143–148 (2) | 148 | 143–148 (2) | 148 | 143–153 (2) | 143–153 (2) |
| NBRI_gPD_88-1 | 149–158 (32) | 164 | 149 | 161 | 149 | 161 | 149 | 149–163 (2) | 149–163 (2) | 159 | 159 | 159 |
| NBRI_gM6 | 145–182 (31) | 179 | 179 | 173 | 173 | 179 | 173 | 181 | 175 | 168 | 179–187 (2) | 179–187 (2) |
| NBRI_gM175 | 155–159 (30) | 155 | 161 | 155 | 155 | 155–161 | 155 | 155 | 155 | 155 | 155 | 155 |

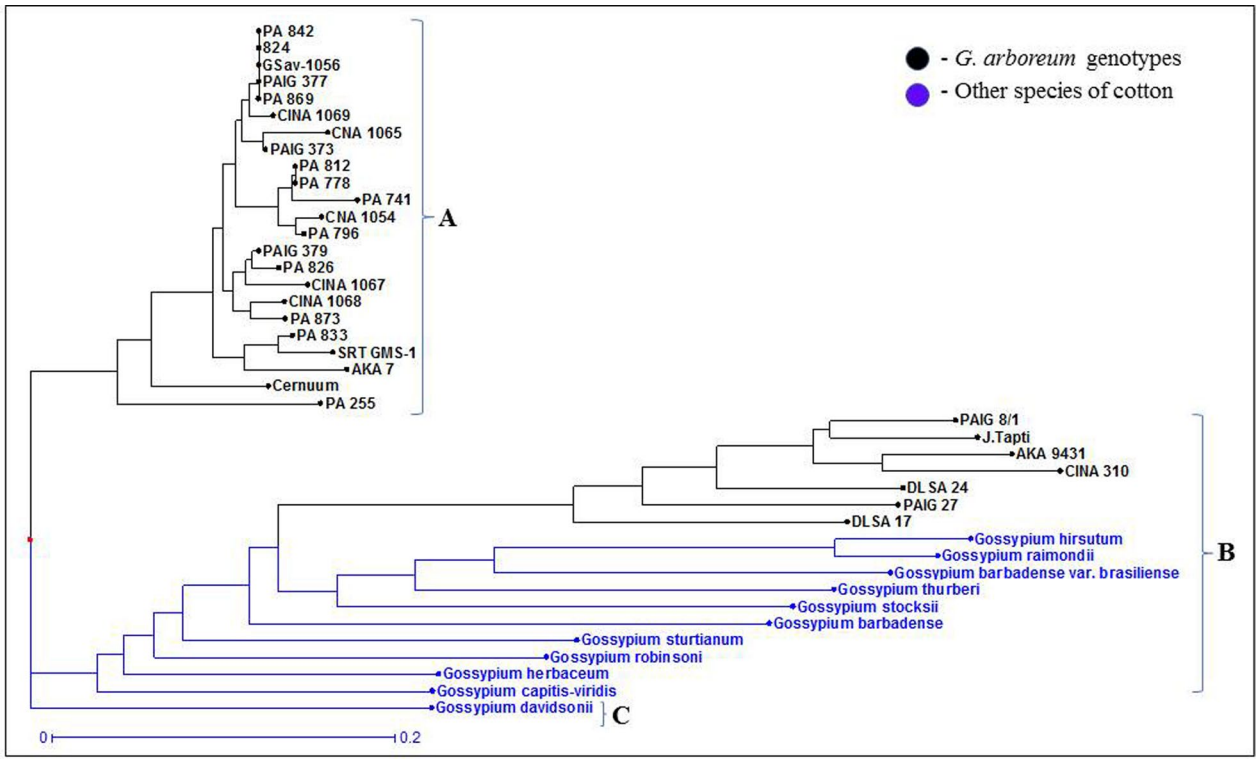


**Fig. 4** Dendrogram showing relationship among 41 *Gossypium* genotypes generated by DARwin using SSR marker data

both clustering analysis as it was detected highly diverse and fall away from rest of the genotypes. In both clustering analysis there was some harmony between the grouping pattern for example CINA and DLSA genotypes grouped together in different clusters.

The pairwise comparison values of Nei's (1973) genetic distance among the 41 cotton genotypes displayed a wide range, varying from 0 to 0.875. The highest genetic distance of 0.875 was observed between PA 741 and CINA 310, indicating a substantial genomic difference between these two genotypes. Similarly, genetic distances of 0.854 were observed between *G. davidsonii* and *J. Tapti, G. davidsonii* and CINA 310, *G. thurberi* and CINA 310, and *G. hirsutum* and PAIG 8/1, highlighting their distinct genetic backgrounds.

These highly divergent genotypes can be valuable resources for developing bi-parental mapping populations and enhancing cotton improvement programs to broaden the genetic diversity of various cotton genotypes. On the other end of the spectrum, the lowest genetic distance of 0 (zero) was observed between PA 869 and 824, indicating their close genetic relationship.

The average genetic distance among the 41 cotton genotypes was calculated to be 0.47, reflecting the overall genetic variability within the studied cotton population. The knowledge of genetic distances between these genotypes facilitates the identification of potential parental combinations for breeding strategies aimed at enhancing cotton crop diversity and improving desirable traits.

The Mantel test analysis showed that a moderate correlation (0.42 at $p$ value 0.0003) between SSR marker-based and quantitative trait-based distance matrices. This suggested that genetic differences captured by molecular markers moderately aligned with differences in morpho-biochemical traits. In case when association is not strong indicated that genetic variation influence the quantitative trait variation some extent but other factors like environment and crop management may also a play role in expression of studied quantitaive traits (Saiyad and Kumar 2018). Though, in current study, correlation was moderate but grouping of few genotypes in dendrogram was same like cernnum was highly different from rest of the *G. arboreum* genotypes in morphological trait and marker based grouping. Like wise, not only CINA series genotypes were grouped closely but also PA 812 and PA 878 were in same group. PAIG 8/1 and J. Tapti also showed a consistent relationship in both dendrograms, aligning in the same group based on both morphological traits and SSR marker-based grouping. Molecular markers are typically considered selectively neutral, meaning they do not directly influence traits or undergo natural selection. As a result, they may not always accurately

represent the diversity in functional or adaptive traits (Kumar et al. 2013). But in current study, the result indicated a significant level of agreement between the two methods of classification.

### Way forward for utilization of diploid cotton in tetraploid improvement

Globally, cotton breeding is focused either in genetic improvement of tetraploids (*G. hirsutum* or *G. barbadense*) cotton due to their high yield potential and better spinning capacity (Khadi et al. 2003). But tetraploids fiber face challenges of less fiber strength compared to dillpods. Moreover, use of same or similar elite parents in cotton breeding narrowed the genetic variability (Yang et al. 2020). In such case diploid cotton may be used to transfer valuable traits into elite cultivars of tetrataploid cotton to improve fiber quality. One of the key challenges in deploying diploid cotton for G. hirsutum improvement is the difference in ploidy levels, which obscure the interspecific breeding efforts. Polyploidization is the lone tool that leads the genetic diversity of diploids to polyploids but it is frequently difficult to get fertile hybrids through crossing between tetraploid and diploid species as triploid will be sterile (Khidirov et al. 2023). The fertility of hybrids can be recovered by polyploidization, resulting in hexaploids (Montes et al. 2017). Recently, Khidirov et al. (2023) have developed interspecific hybrids of *G. herbaceum* L. (A1-genome) and *G. mustelinum* Miers ex Watt (AD4-genome) species followed by polyplodization of the sterile triploid. The synthetic polyploids were fertile and were also confirmed through genomic and cytogenetic analysis. The successful example of Khidirov et al. (2023) suggested that superior genotypes identified in current study can be used through polyploidization breeding for broadening the tetraploid cotton diversity as well as development of new genetic resources through pre-breeding.

## Conclusion

The experiment effectively acknowledged elite genotypes, such as PA 255, CINA 1067, and PA 842, with more seed cotton yield per plant and associated traits, making them key candidates for future cotton breeding programs designed at improving yield. The exceptional genotype, Cernuum, deals valuable genetic diversity, underlining its prospective role in targeted trait improvement and the management of genetic resources. The amalgamation of SSR markers through diverse Gossypium species has providing a widespread knowledge of species diversity, supporting the reliability of SSR markers. Besides, the consistency between morphological and molecular diversity results highlights the reliability of SSR markers in reflecting phenotypic diversity. The unique genotypes, particularly those like Cernuum, should be conserved and utilized to maintain and broaden the genetic base of cotton, ensuring long-term sustainability and resilience. Including a broader range of cotton species in molecular studies will further enhance our understanding of genetic diversity and improve the robustness of marker-assisted selection. The highly diverse genotypes can be used to create bi-/multi-mapping population followed by QTL mapping to swift the molecular breeding in *desi* cotton.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43170-024-00314-2.

Additional file 1. Suppl. Figure 1 Morphological diversity among the desi cotton genotypes

**Acknowledgements**
Authors acknowledge Anand Agricultural University, Gujarat, India for providing facilities to conduct experiments.

**Author contributions**
Conceived and designed the experiments: Dhramendra Patidar and Sushil Kumar. Performed the experiments: Hrushik Vadodaria, Dhramendra Patidar, Hardik Patel. Analyzed and interpreted the data: Hrushik Vadodaria, Sushil Kumar, Dhramendra Patidar. Contributed reagents, materials, analysis tools or data: Dhramendra Patidar, Dipak A Patel, Sushil Kumar. Wrote the paper: Hrushik Vadodaria, Sushil Kumar.

**Funding**
This research received no funding.

**Availability of data and materials**
The original contributions presented in this study are included in the article.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare that they have no competing interests.

**Author details**
[1]Department of Agricultural Biotechnology, Anand Agricultural University, Anand 388 110, India. [2]Regional Cotton Research Station, Anand Agricultural University, Viramgam 382 150, India.

Received: 6 March 2024 Accepted: 28 October 2024
Published online: 19 November 2024

**References**

Anonymous. National cotton scenario: Cotton Corporation of India. 2021. https://www.cotcorp.org.in/. Accessed 20 Dec 2022.

Azmat MA, Khan AA. Assessment of genetic diversity among the varieties of *Gossypium arboreum* and *Gossypium hirsutum* through random amplification of polymorphic DNA (RAPD) and simple sequence repeat (SSR) markers. Pak J Bot. 2010;42(5):3173–81.

Blaise D, Kranthi KR, Ravindran CD, Thalal K. High plant density can improve the productivity of rainfed Asiatic cotton (*Gossypium arboreum* L.). Arch Agron Soil Sci. 2020;67(5):607–19.
Celik S. Genetic diversity analysis of some upland cotton (*Gossypium hirsutum* L.) genotypes using SSR markers. Turk J Nat Sci. 2022;11(1):80–9.
Daniel DR. The chemical and functional properties of cottonseed oil as a deep-fat frying medium. Ph.D Thesis, T & M University Texas, USA. 2007.
Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1990;12:13–5.
Erande CS, Kalpande HV, Deosarkar DB, Chavan SK, Patil VS, Deshmukh JD, Puttawar MR. Genetic variability, correlation and path analysis among different traits in desi cotton (*Gossypium arboreum* L.). Afr J Agric Res. 2014;9(29):2278–86.
Garg, H. R., Kapoor, A., Rathore, P., Pathak, D., Gill, J. S., & Raheja, R. K. (2003) Evaluation of *Gossypium* genotypes for oil content and component fatty acids. World Cotton Research Conference 3 (Cape Town - South Africa).
Guo WZ, Zhou BL, Yang LM, Wang W, Zhang TZ. Genetic diversity of landraces in *Gossypium arboreum* L. race sinense assessed with simple sequence repeat markers. J Integr Plant Biol. 2006;48(9):1008–17.
Iqbal Z, Sattar MN, Shafiq M. CRISPR/Cas9: a tool to circumscribe cotton leaf curl disease. Front Plant Sci. 2015;7:475.
Jena SN, Srivastava A, Rai KM, Ranjan A, Singh SK, Nisar T, Sawant SV. Development and characterization of genomic and expressed SSRs for levant cotton (*Gossypium herbaceum* L.). Theor Appl Genetics. 2012;124(3):565–76.
Jogender OS, Kumar D, Shreya AD. Assessment of genetic variability in promising Asiatic cotton (*Gossypium arboreum* L.) genotypes under rainfed conditions. Pharma Innov J. 2023;12(2):1642–6.
Khadi, B.M., Kulkarni,V.N., Ansingkar,A.S., Singh, V.V. & Gotmore, V. (2003). Gene pool concept for diploid gene pool concept for diploid cultivated cotton. In: World cotton research conference-3, Cape Town, South Africa, pp 214–225
Khan NU, Hassan G, Kumbhar MB, Parveen A, Aiman U, Ahmad W, Shah SA, Ahmad S. Gene action of seed traits and its oil content in upland cotton (*Gossyium hirsutum* L.). SABRAO J Breed Genetics. 2007;39:17–29.
Khidirov MT, Ernazarova DK, Rafieva FU, Ernazarova ZA, Toshpulatov AK, Umarov RF, Kholova MD, Oripova BB, Kudratova MK, Gapparov BM, Khidirova MM, Komilov DJ, Turaev OS, Udall JA, Yu JZ, Kushanov FN. Genomic and cytogenetic analysis of synthetic polyploids between diploid and tetraploid cotton (*Gossypium*) species. Plants (Basel). 2023;12(24):4184.
Kulkarni, V. N., Basavaraj M. K, Manjula S. M., Deshapande L. A., & Narayanan S. S. (2009). The worldwide gene pools of *G. arboreum* L. and *G. herbaceum* L. and their improvement. *Genetics and Genomics of Cotton*, 69–97.
Kumar, S., Sharma, R., Kumar, V, Vyas G. K. & Abhishek Rathore (2013) Combining molecular-marker and chemical analysis of *Capparis decidua* (Capparaceae) in the Thar Desert of Western Rajasthan (India). Rev. Biol. Trop. (Int. J. Trop. Biol.), 61 (1): 311–320
Liu K, Muse SV. Power marker: Integrated analysis environment for genetic marker. Bioinformatics. 2005;21:2128–9.
Liu QL, Singh SP, Green AG. High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. Plant Physiol. 2002;129:1732–43.
Liu D, Guo X, Lin Z, Nie Y, Zhang X. Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. Genetic Resourc Crop Evol. 2006;53(6):1145–52.
Mankar AD, Deosarkar DB, Chinchane VN. Studies of genetic parameters for yield, yield contributing and fiber quality characters in desi cotton (*Gossypium arboreum* L.). Pharma Innov J. 2021;10(8):42–4.
Mehetre SS, Aher AR, Gawande VL, Patil VR, Mokate AS. Induced polyploidy in Gossypium: a tool to overcome interspecific incompatibility of cultivated tetraploid and diploid cottons. Curr Sci. 2003;84(12):1510–2.
Mohan P, Singh VV, Mukewar PM. Morphological and fibre characteristics of *Gossypium arboreum* L. germplasm lines immune to grey mildew (*Ramularia areola* ATK.) disease of cotton. Indian J Plant Genetic Resourc. 2001;14(1):81–4.
Montes E, Coriton O, Eber F, Huteau V, Lacape J, Reinhardt C, Marais D, Hofs J, Chevre A, Pannetier C. Assessment of gene flow between *Gossypium hirsutum* and *G. herbaceum*: evidence of unreduced gametes in the diploid progenitor. G3 Genes Genomes Genetics. 2017;7:2185–93.
Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA. 1973;70(12):3321–3.
Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al (2024) vegan: Community Ecology Package. R Package Version 2.6-8. https://CRAN.R-project.org/package=vegan.
Parekh MJ, Kumar S, Zala HN, Fougat RS, Patel CB, Bosamia TC, Kulkarni KS. Development and validation of novel fiber relevant dbEST–SSR markers and their utility in revealing genetic diversity in diploid cotton (*Gossypium herbaceum* and *G. arboreum*). Ind Crops Prod. 2016;83:620–9.
Parekh MJ, Kumar S, Fougat RS, Zala HN, Pandit R. Transcriptomic profiling of developing fiber in levant cotton (*Gossypium herbaceum* L.). Funct Integr Genomics. 2018;18:211–23.
Perrier X, Flori A, Bonnot F. Methods for data analysis. In: Hamon P, Seguin M, Perrier X, Glaszmann JC, editors. Genetic diversity of cultivated tropical plants. Montpellier: Science Publishers Inc and CIRAD; 2003. p. 31–63.
Rahman M, Hussain D, Zafar Y. Estimation of genetic divergence among elite cotton cultivars–genotypes by DNA fingerprinting technology. Crop Sci. 2002;42(6):2137–44.
Ranjan R, Sangwan RS, Siwach SS, Sangwan O, Sah M, Singh N. Studies on genetic divergence in *Gossypium arboreum* L. J Plant Breed Genetics. 2014;2(2):57–61.
Ravi R, Prakash M, Bhat KK. Characterization of aroma active compounds of cumin (*Cuminum cyminum* L.) by GC-MS, e-Nose, and sensory techniques. Int J Food Prop. 2013;16(5):1048–58.
Rohlf FJ. 1998. NTSYS-pc Numerical taxonomy and multivariate analysis system.Version 2.02e. User Guide. Applied Biostatistics, Setauket, New York, USA.
Sagar MR, Kumar S, Patidar D, Sakure AA. Morphological, physico-biochemical and marker-based diversity of desi cotton (*Gossypium herbaceum* L.) germplasm. J King Saud Univ Sci. 2022;34(8):102336.
Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.
Saiyad M, Kumar S. Evaluation of maize genotypes for fodder quality traits and SSR diversity. J Plant Biochem Biotechnol. 2018;27:78–89.
Sangwan RS, Siwach SS, Kumar R. Evaluation of genetic stock of desi cotton (*Gossypium arboreum*). Indian J Plant Genetic Resourc. 2008;21(1):68–71.
Santosh HB, Meshram M, Santhy V, Waghmare VN. Microsatellite marker-based diversity analysis and DNA fingerprinting of Asiatic cotton (*Gossypium arboreum*) varieties of India. J Plant Biochem Biotechnol. 2021;31(2):421–8.
Saravanan M, Misra RC, Mahajan SS, Patil DV, Waghmare VN. Morphological and molecular characterization of desi cotton (*Gossypium herbaceum* L.) landraces collected from different states of India. Electr J Plant Breed. 2021;12(1):142–50.
Sethi K, Siwach P, Verma SK. Assessing genetic diversity among six populations of *Gossypium arboreum* L. using microsatellites markers. Physiol Mol Biol Plants. 2015;21(4):531–9.
Shahrajabian MH, Sun W, Cheng Q. Chemical components and pharmacological benefits of basil (*Ocimum basilicum*): a review. Int J Food Prop. 2020;23(1):1961–70.
Shim J, Mangat PK, Angeles-Shim RB. Natural variation in wild Gossypium species as a tool to broaden the genetic base of cultivated cotton. J Plant Sci Curr Res. 2018;2(1):005–13.
Wang J, Park JS, Wei Y, Rajurkar M, Cotton JL, Fan Q, Mao J. TRIB2 acts downstream of Wnt/TCF in liver cancer cells to regulate YAP and C/EBPα function. Mol Cell. 2013;51(2):211–25.
Xiang XL. Study and utilization on Asiatic cotton (*G. arboreum*) in China. Sci Agric Sin. 1988;21:94.
Yang Z, Qanmber G, Wang Z, Yang Z, Li F. Gossypium genomics: trends, scope, and utilization for cotton improvement. Trends Plant Sci. 2020;25(5):488–500.

## Publisher's Note