https://doi.org/10.17221/12/2023-CJGPB

Genetic diversity among coloured cotton genotypes in relation to their fibre colour and ploidy level based on SSR markers

Revanasiddayya[1], Jayaprakash Mohan Nidagundi[1], Bashasab Fakrudin[2], Prakash Kuchanur[1], Lingappa Neelagiri Yogeesh[1], Shivanand Hanchinal[3], Talagunda Chandrashekar Suma[4], Gururaj Sunkad[5], Bharamappanavara Muralidhara[1], Doddamani Maheshkumar[2], Channabasava[1], Patil Sudha[6], Shivamurthy Rashmi[6], Virupapura Cholaraya Raghavendra[1]*

[1]*Department of Genetics and Plant Breeding, University of Agricultural Sciences, Raichur, Karnataka, India*
[2]*Department of Biotechnology and Crop Improvement, University of Horticultural Sciences, Bagalkot, Karnataka, India*
[3]*Department of Agricultural Entomology, University of Agricultural Sciences, Raichur, Karnataka, India*
[4]*Department of Crop Physiology, University of Agricultural Sciences, Raichur, Karnataka, India*
[5]*Dean (Post Graduate Studies), University of Agricultural Sciences, Raichur, Karnataka, India*
[6]*AICRP on Cotton, MARS, University of Agricultural Sciences, Raichur, Karnataka, India*

**Corresponding author: revanasiddayya3@gmail.com*

Citation: Revanasiddayya, Nidagundi J.M., Fakrudin B., Kuchanur P., Yogeesh L.N., Hanchinal S., Suma T.C., Sunkad G., Muralidhara B., Maheshkumar D., Channabasava, Sudha P., Rashmi S., Raghavendra V.C. (2024): Genetic diversity among coloured cotton genotypes in relation to their fibre colour and ploidy level based on SSR markers. Czech J. Genet. Plant Breed., 60: 12–24.

Abstract: Genetic diversity is referred to as any variation at the phenotypic, DNA or genomic level of an individual, population or species. The appraisal of diversity is important to understand its pattern and evolutionary relationships between germplasms or genotypes, which will aid in sampling the genetic resources in a more systematic manner for conservation and crop improvement. The present study employed 50 simple sequence repeat (SSR) markers linked to the yield and fibre quality/colour traits for estimating the genetic diversity in 33 cotton genotypes of diploid and tetraploid species differing in fibre colour. The diversity analysis was performed in GenAlEx (Ver. 6.41) and Powermarker (Ver. 3.25) while DARwin (Ver. 6.0.21) software was used to establish the phylogenetic relationships following neighbour-joining (NJ) and unweighted pair group method with arithmetic (UPGMA) mean method. Markers generated 186 polymorphic loci as genotypic data with an average of 3.72 alleles and an average polymorphic information content (PIC) value of 0.59 per SSR locus. The NJ and UPGMA grouped 33 genotypes into three major clusters I, II and III consisting of 21 tetraploid *Gossypium hirsutum*, 10 *G. arboreum* coloured and 2 white cotton genotypes, respectively. In the PCA, the first two components (PC1 and PC2) explained 74.69% of the variation and the biplot plotted the 33 genotypes in three groups. The study established the diverse nature of 33 cotton genotypes based on their fibre colour and ploidy level. With confirmation of the prevalent genetic diversity, we suggest that hybridisation can be planned among diverse genotypes to unleash greater variation in the fibre colour or to derive superior cross combinations.

Keywords: *Gossypium hirsutum*; neighbour-joining; PCA; polymorphism level; UPGMA

https://doi.org/10.17221/12/2023-CJGPB

India has been a traditional home for cotton and cotton textiles. The majority of players in the textile industry prefer white fibres because they provide a uniform substrate for dyeing and processing. New market trends have led to the emergence of other niches, such as naturally coloured fibres that required no or less dyeing during textile production, minimising the environmental pollution caused by minimal residual chemical toxicants (Feng et al. 2013; Rathinamoorthy & Parthiban 2017).

Cotton with naturally coloured lint, other than white, is commonly referred to as coloured cotton. Cultivation of coloured cotton was discouraged and almost abandoned in the latter half of the last century owing to their poor yield, fibre quality characteristics and non-uniform colours (Sun et al. 2021). In recent years, coloured cotton is gaining increasing importance because of its eco-friendly nature as it eliminates the dyeing stage in industrial production (Rathinamoorthy & Parthiban 2017). In addition, synthetic dyes have adverse effects on human health (Hijazi et al. 2015; Nagarajan et al. 2022). As the world moves towards organic fabrics and products, naturally coloured cotton is going to be the next big thing in the market (Nagarajan et al. 2022).

Nearly all cotton fibres produced in the world are white; however, lint and fibre of diploid and tetraploid cottons occur in several lint colours such as brown, pink, yellow, tannin and green (Günaydin et al. 2019). Brown is the most common colour and shades vary from light brown to intense mahogany red in the four cultivated species as well as many of the wild species. First, coloured cotton was discovered to be a mutant of the white cotton *Gossypium hirsutum* and *G. barbadense* species (Gong et al. 2018). So far, varieties of coloured cotton have mostly been developed by selection and recurrent crossing approaches from the germplasm. Approximately, 45 diploid ($2n = 2x = 26$) and five allotetraploid ($2n = 4x = 52$) species make up the vast genetic resources of the *Gossypium* genus, which are scattered globally over five continents (Brubaker et al. 1999). The poor popularity among cotton growers has demanded focused breeding efforts for improving the coloured cotton genotypes to make coloured cotton commercially viable (Basavaradder & Maralappanavar 2014). Developing more vibrant, diverse fibre colours, high yielding, quality coloured cotton varieties need the combined usage of conventional and biotechnological crop improvement methods.

The manifestation of heterosis in coloured cotton relies upon on the accessible portion of genetic diversity (Malik et al. 2014). Baloch et al. (2015) reaffirmed that the genetic diversity of the parental lines determines the choice to produce heterotic hybrids. Since it provides details on the allelic variation that can be used to produce novel, desirable gene combinations, the genetic diversity and comprehension of the relationships between the genotypes are important in cotton breeding. The genetic distance or similarity, which indicates either genetic differences or similarities, is typically used to quantify diversity. Traditionally, morphological markers have been used for this, but more recently, the introduction of molecular markers like simple sequence repeats (SSRs) has increased the efficiency of genetic diversity analyses (Tyagi et al. 2015; Kusuma et al. 2018). Due to the environmental influence and stage or tissue-specific gene expression, morphological traits may only represent a portion of the total genetic variation, which may not accurately reflect the genetic divergence between genotypes/species (Sundar et al. 2014). Molecular markers are considered an efficient and powerful tool for the assessment of genetic relationships. SSR markers could be an ideal means for the identification of the genetic diversity and relationship of cotton resources at the genomic level (Seyoum et al. 2018) owing to their high reproducibility and high veracity (Wu et al. 2020). In recent times, the focus of studies on the diversity of coloured cotton has changed from phenotypic, cellular, and biochemical levels to DNA levels. Regarding the genetic diversity of coloured cotton, only a few reports are available (Ma et al. 2003; Guo et al. 2004; Zhang et al. 2004; Sun et al. 2009; de Rocha et al. 2016). Earlier efforts of Sun et al. (2009) and de Rocha et al. (2016) indicated the existence of low and high genetic diversity levels in coloured cotton genotypes emphasising the need for identifying and evaluating coloured-cotton germplasm systematically to hasten up coloured cotton genetic improvement. Owing to a narrow genetic base and lack of an adequate number of coloured cotton germplasms (Sun et al. 2009), an initial assessment of the genetic diversity guides a breeder in choosing diverse parents and helps in predicting the degree of variation and extent of heterosis (Zhu et al. 2014). In this background, the present study was aimed at assessing the genetic diversity at the molecular level by employing SSR markers in colour cotton genotypes derived from diploid and tetraploid species to understand the genetic diversity levels among them, which will help in devising appropriate crop improvement programmes either to create

https://doi.org/10.17221/12/2023-CJGPB

variability or exploit heterosis if diverse genotypes are identified.

MATERIAL AND METHODS

Experimental material. A total of 33 cotton genotypes comprising 21 tetraploid brown coloured cotton (*G. hirsutum*, AADD genome, $2n = 4x = 52$) (TBCC), 10 diploid brown coloured cotton (*G. arboreum*, AA genome, $2n = 2x = 26$) (DBCC) and 2 tetraploid white coloured cotton (*G. hirsutum*, AADD genome, $2n = 4x = 52$) (TWCC) genotypes were used (Table S1 in the Electronic Supplementary Material (ESM)). Eight of the 21 tetraploid genotypes are dark brown (DB), seven are medium brown (MB), and six are light brown (LB) and there were three DB, three MB, and four LB colour genotypes among the ten *G. arboreum* genotypes.

Experimental methodology and agronomic practices. The experiment was laid out in a randomised complete block design with two replications in deep black soil at a cotton block, the University of Agricultural Sciences, Raichur, Karnataka. Each row was 6 m in length and a spacing of 90 cm between the rows and 30 cm between the plants was maintained. The seeds were hand dibbled. The first and last rows in each main plot were sown as border rows to minimise the border effect. Fertilisers at the recommended doses were applied and other cultural practices were carried out at regular intervals as per the recommended package of practices during the course of the experiment to raise a good and healthy crop. A fertiliser dose of 80 : 40 : 40 N, P_2O_5, K_2O kg per hectare was applied. The application of N was split into two equal instalments, *i.e.*, basal and top dressing. Plant protection measures were taken up at the appropriate time to control pests and diseases.

Sample collection and laboratory conditions. The young leaves were collected in the early hours and stored in the mini cooler (–20 °C) after labelling them. The collected samples were then moved to a –80 °C storage unit at the Biotechnology Laboratory, Department of Biotechnology and Crop Improvement, College of Horticulture, UHS Campus, GKVK Post, Bengaluru to carry out the molecular work. The cetyltrimethylammonium bromide (CTAB) DNA extraction method given by Doyle and Doyle (1987) was used with some modifications for the extraction of the genomic DNA.

The amount of DNA in each sample was quantified by taking readings at 260 and 280 nm in a NanoDrop Spectrophotometer (UV Technologies, USA). The DNA was diluted to prepare the working solutions. The polymerase chain reaction (PCR) amplification of genomic DNA samples of the cotton genotypes was performed in a thermal cycler (Master Cycler gradient 5331-Eppendorfversion 2.30.31-09, Germany) using 50 pairs of SSR markers were reported to be highly polymorphic and linked to various quantitative traits and fibre quality traits (Table S2 in the ESM) in the Cotton Microsatellite Database https://www.cottongen.org/. A 10 µL PCR mixture consisted 1 µL of genomic DNA (50 ng µL), 3.0 µL of *Taq* 2× Master Mix, 0.5 µL of forward primer (5 pmol/µL), 0.5 µL of reverse primer (5 pmol/µL) and 5 µL of nuclease free water. Each PCR cycle was performed with initial denaturation for 3 min at 94 °C, denaturation for 1 min at 94 °C, primer annealing for 1 min at 55 °C, primer extension for 1 min at 72 °C, and a final extension for 10 min at 72 °C with a hold for an extended time at 4 °C. The separation and visualisation of the SSR products were conducted on 3% agarose gel. The agarose gel (3%) was prepared by dissolving 3 g of agarose in 100 mL of 1× TAE (tris HCL, acetic acid and EDTA) buffer. Slabs were cast in a horizontal gel frame (Hoefer HE99X 18 × 30 cm Amersham Bioscience Pvt. Ltd. USA), the products were visualised by incorporating 5 µL (10 mg/mL) of ethidium bromide and viewed in a gel documentation system (Syngene, Pvt. Ltd. USA). The amplified products generated from the SSR PCR reaction were resolved on agarose gel. The amplicons which distinguish the cotton genotypes were observed and scored. Scoring was undertaken using G:Box F3 GENESys software which scores based on the molecular weight of each band with reference to the 100 bp DNA ladder for the SSR products.

Genetic diversity analysis. The level of the polymorphism among the samples was assessed for the 50 SSR markers. The genotypic data were analysed for the mean number of alleles per locus (Na), the number of effective alleles per locus (Ne), Shannon's Information Index (*I*), the observed heterozygosity (Ho) and expected heterozygosity (He) using GenAlEx (Ver. 6.41) (Peakall & Smouse 2006). The polymorphic information content (PIC) was calculated using the software package Powermarker (Ver. 3.25) according to Liu and Muse (2005).

Phylogenetic analysis. A phylogenetic tree based on the neighbour-joining (NJ) method, dissimilarity coefficients, and a dendrogram based on the unweighted pair group method with arithmetic mean (UPGMA) was constructed to analyse the genetic

https://doi.org/10.17221/12/2023-CJGPB

relationship among the individuals using DARwin (Ver. 6.0.21) software. The principal component analysis (PCA) was carried out in R software (Ver. 4.1.2) using the Factoextra statistical package and the results of the PCA were seen through graphical biplot (Jollife 2014).

RESULTS AND DISCUSSION

In the present study, the PIC value displayed by each marker varied from a minimum of 0.11 in three primer pairs (CGR-5167, HAU-0590 and SHIN-1400) to a maximum of 0.89 in NAU-5433 (Table 1) which is comparable with Kuang et al. (2022) with a range of values of 0.18 to 0.90 based on 71 SSR primers revealing the efficiency of the markers to differentiate the genotypes (Ammad et al. 2015; Kencharaddi et al. 2018). An average PIC value of 0.59 per SSR locus was noticed indicating the highly informative nature of the markers which correlated with the findings of Ghuge et al. (2018), Kuang et al. (2022) and Isong et al. (2019). Thirty-four markers had a PIC value higher than 0.50 inferring the potential of the markers chosen (Ahmad et al. 2015) while the obverse was true for sixteen markers. To contemplate, Sun et al. (2009) noticed the PIC of the markers from 0.35 to 0.88 with an average of 0.76 in 40 brown and 21 green coloured genotypes using 66 SSR markers. The banding profile indicating the polymorphism level of the SSR markers is depicted in Figure 1.

A total of 186 polymorphic loci were generated as the genotypic data with an average of 3.72 alleles per SSR locus validating the sample size of the study (Table 1) while a higher mean of the alleles shows the diversity or variation present on the chromosomes as the SSR loci are distributed throughout the genome. Earlier, Ali et al. (2019) noticed an average of 4.3 alleles per SSR locus while Kuang et al. (2022) and Isong et al. (2019) noted 142 and 155 alleles with an average of 2.01 and 2.51 alleles per SSR locus, respectively. The number of different alleles (Na) varied from 1 to 7 alleles in marker CGR-6378 (least diverse) and HAU-1430 (most diverse), respectively. Such a pattern was earlier reported by Sun et al. (2009) and Kencharaddi et al. (2018) with the number of different alleles varying from 1 to 13 with an average

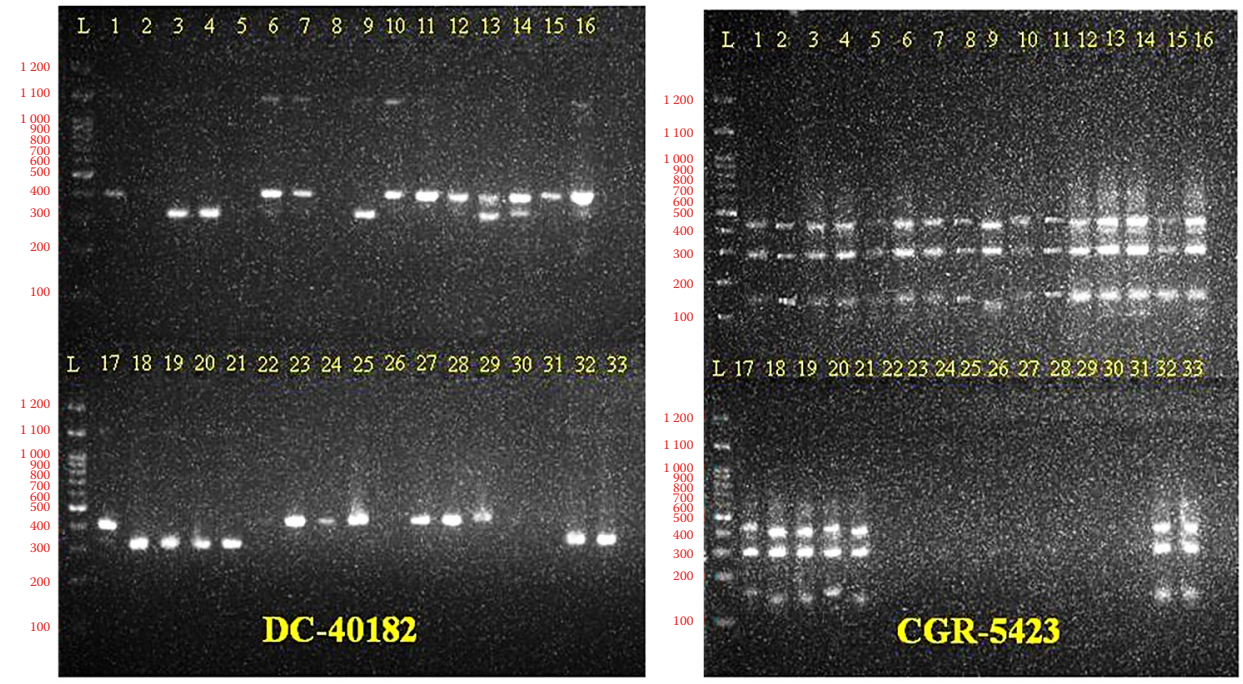

Figure 1. Simple sequence repeats banding profile of the coloured cotton genotypes generated for the primers, *viz.*, DC-40182 and CGR-5423

L – ladder, 1–21: 21 genotypes of TBCC, 22–31: 10 genotypes of DBCC and 31–32: 2 genotypes of TWCC; TBCC – tetraploid brown coloured cotton genotypes; DBCC – diploid brown coloured cotton genotypes; TWCC – tetraploid white coloured cotton genotypes

https://doi.org/10.17221/12/2023-CJGPB

Table 1. Locus-wise information on the polymorphism parameters across the coloured cotton genotypes

No.	SSR marker	PIC	Na	Ne	*I*	Ho	He
1	BNL-3590	0.78	6	4.00	1.51	1.00	0.75
2	BNL-1395	0.64	3	2.00	0.82	0.00	0.50
3	BNL-1604	0.62	3	2.00	0.82	0.00	0.50
4	NAU-3308	0.59	4	2.83	1.17	0.00	0.65
5	JESPR-204	0.71	5	2.38	1.02	1.00	0.58
6	CIR-221	0.42	3	2.00	0.82	0.00	0.50
7	BNL-1672	0.43	3	2.00	0.82	0.00	0.50
8	NAU-2658	0.43	3	2.00	0.82	0.00	0.50
9	DC-40182	0.64	3	2.00	0.82	0.00	0.50
10	CIR-307	0.76	4	2.38	0.99	1.00	0.58
11	HAU-1417	0.73	3	2.00	0.82	0.00	0.50
12	JESPR-153	0.64	2	1.19	0.30	0.00	0.16
13	TMB-1618	0.81	2	2.00	0.69	1.00	0.50
14	DC-40182	0.43	3	2.00	0.82	0.00	0.50
15	BNL-2572	0.67	4	2.58	1.07	0.00	0.61
16	BNL-1440	0.64	4	3.31	1.25	0.00	0.70
17	JESPR-274	0.42	3	2.00	0.82	0.00	0.50
18	CGR-5282	0.43	3	2.00	0.82	0.00	0.50
19	CGR-6378	0.42	1	1.00	0.00	0.00	0.00
20	DC-40052	0.43	3	2.00	0.82	0.00	0.50
21	DC-20076	0.60	4	2.92	1.19	0.00	0.66
22	CGR-5867	0.39	3	2.00	0.82	0.00	0.50
23	CGR-5399	0.71	6	4.00	1.51	1.00	0.75
24	DPL-0555	0.43	3	2.00	0.82	0.00	0.50
25	CGR-5167	0.11	3	1.13	0.27	0.00	0.12
26	CGR-5258	0.71	6	2.39	1.05	1.00	0.58
27	SHIN-1343	0.43	3	2.00	0.82	0.00	0.50
28	CGR-5423	0.65	4	2.38	0.99	1.00	0.58
29	DPL-0441	0.43	3	2.00	0.82	0.00	0.50
30	CGR-5541	0.63	3	2.00	0.82	0.00	0.50
31	HAU-0591	0.71	4	3.02	1.23	0.00	0.67
32	HAU-0590	0.11	3	1.13	0.27	0.00	0.12
33	HAU-1430	0.58	7	2.74	1.30	0.00	0.64
34	BNL-3790	0.64	5	3.37	1.30	0.00	0.70
35	SHIN-1400	0.11	2	1.13	0.23	0.00	0.11
36	HAU-1321	0.63	4	3.23	1.24	0.00	0.69
37	NAU-2277	0.62	4	3.12	1.22	0.00	0.68
38	NAU-2508	0.43	3	2.00	0.82	0.00	0.50
39	NAU-2437	0.72	5	3.96	1.46	0.00	0.75
40	NAU-0998	0.59	4	2.83	1.17	0.00	0.65
41	BNL-3479	0.75	4	3.21	1.25	0.00	0.69
42	BNL-3452	0.63	4	3.23	1.24	0.00	0.69
43	BNL-2544	0.82	5	4.24	1.50	0.00	0.76
44	CGR-5565	0.77	4	2.87	1.21	0.06	0.65

https://doi.org/10.17221/12/2023-CJGPB

Table 1 to be continued

No.	SSR marker	PIC	Na	Ne	*I*	Ho	He
Specific coloured cotton markers							
45	NAU-3735	0.59	3	2.69	1.04	0.00	0.63
46	NAU-1043	0.76	6	4.67	1.64	0.00	0.79
47	NAU-2862	0.81	3	1.87	0.82	0.03	0.47
48	NAU-5433	0.89	6	4.62	1.61	1.00	0.78
49	NAU-5434	0.76	4	2.31	1.06	0.00	0.57
50	NAU-2968	0.84	3	1.88	0.82	0.30	0.47
Range	Min	0.11	1.00	1.00	0.00	0.00	0.00
	Max	0.89	7.00	4.67	1.64	1.00	0.79
Mean		0.59	3.72	2.49	0.97	0.17	0.54
Total		186.00	124.61	48.56	8.39	27.23	
SE		0.172	0.125	0.051	0.052	0.025	

PIC – polymorphic information content; Na – No. of different alleles; Ne – No. of effective alleles; *I* – Shannon's information index; Ho – observed heterozygosity; He – expected heterozygosity; SE – standard error

of 5.2 and 3.37 alleles per locus, respectively. In the current study, the observed high number of alleles might be due to the fact that the genotypes used included tetraploid and diploid species of brown and white coloured genotypes contributing to their diverse nature. The diversity of the genotypes under study is reflected in the number of alleles amplified by each marker (Isong et al. 2019). Furthermore, Ali et al. (2019) noted the lowest polymorphism and smaller number of alleles concluding that there was a narrow genetic base of the studied genotypes.

The number of effective alleles (Ne) that provides a good measure of genetic variation in accordance with the population size was the maximum for the locus NAU-1043 (4.67) followed by NAU-5433 (4.62), BNL-3590 (4.00) and BNL-2544 (4.24) while the lowest was recorded for CGR-6378 (1.00) (Table 1). An Ne ranging from 1.2 to 10.4 was reported earlier by Kuang et al. (2022) concluding that the observed high number of effective alleles amplified by each marker not only corresponded to the diversity in the studied genotypes, but it also highly correlated with the marker type, the applied fragment separation technique, and the resolution (Kuang et al. 2022).

The grand mean of the observed heterozygosity (Ho) and expected heterozygosity (He) was 0.17 and 0.54, respectively. The Ho was found to be zero for 41 SSR loci and 1.00 for eight markers (Table 1) while the mean He values ranged from 0.00 (CGR-6378) to 0.79 (NAU-1043) reflecting the high degree of genetic diversity and the possibility of the very high selection potential for the loci studied. Earlier, Mishra et al. (2013) observed a mean heterozygosity of 0.60 upon using 35 EST- SSR markers in 24 genotypes involving tetraploid and diploid genotypes while Ali et al. (2019) noticed an Ho up to 0.74 with a mean Ho of 0.57 based on 22 SSR markers, who mentioned that the higher Ho is due to the differences in the ploidy level of the genotypes included in the study.

The overall mean of Shannon's Information Index (*I*) was 0.97 with the lowest at 0.00 (CGR-6378) and the largest at 1.64 (NAU-1043). The higher *I* value suggests the more diverse nature of the genotypes (Table 1). The observed high *I* value might be due to the same number of alleles or a few common alleles or an even contribution by all the alleles (Konopinski 2020) in the coloured cotton populations of the tetraploid and diploid species. In contrast, Noormohammadi et al. (2013) reported lower *I* values based on the ISSR markers in the diploid and tetraploid genotypes of cotton.

Phylogenetic analysis

Dissimilarity coefficients. Dissimilarity coefficients (DCs) give an account about the dissimilar nature of genotypes. In the present study, these coefficients were compared within and between the tetraploid brown coloured cotton (TBCC), diploid brown coloured cotton (DBCC) and tetraploid white coloured cotton (TWCC) genotypes. When the DC values were estimated within the TBCC, DBCC and TWCC genotypes, the mean DC values were found to be 0.29, 0.05 and 0.27, respectively, indicating

https://doi.org/10.17221/12/2023-CJGPB

Table 2. A comparison of the dissimilarity coefficients among the genotypes obtained based on the UPGMA method using SSR markers

Genotypes		TBCC	DBCC	TWCC	Overall (TBCC + DBCC + TWCC)
TBCC	Min	0.00	0.00	0.00	0.00
	Max	0.57	0.95	1.00	1.00
	Mean	0.29	0.60	0.40	0.64
DBCC	Min		0.00	0.00	
	Max		0.12	1.00	
	Mean		0.05	0.80	
TWCC	Min			0	
	Max			0.82	
	Mean			0.27	

TBCC – tetraploid brown coloured cotton genotypes; DBCC – diploid brown coloured cotton genotypes; TWCC – tetraploid white coloured cotton genotypes

a narrow diversity and genetic base (Table 2). However, when compared between each other, the mean DC values increased inferring the genetically diverse nature of the genotypes. For instance, when the TBCC and TWCC were compared against the DBCC, the mean DC value was 0.60 and 0.80 respectively, suggesting that the genotypes belonging to the tetraploid and diploid species distinguishable from each other owing to their ploidy level. Whereas, the TBCC and TWCC genotypes have a moderate level of diversity with a DC value of 0.40, even though the studied genotypes are genetically closer to each other, a fair amount of genetic diversity might exist among them (Kuang et al. 2022). Since these two belongs to the tetraploid type, the observed genetic diversity can be attributed to their fibre colour. The mean DC value was recorded as 0.29 among the TBCC genotypes, the spectrum of the values increased to 0.60 upon including 10 DBCC genotypes (Table 2 and Figure 2) which reflected the differences in their ploidy level. Furthermore, the addition of two TWCC genotypes inflated the genetic diversity with a DC value of 0.64 indicating the diversity pattern varies upon including genotypes with different fibre colours (Figure 2). This change in pattern genetic diversity can be attributed to the fibre colour and ploidy level of the genotype which was earlier noticed by de Rocha et al. (2016) and Noormohammadi et al. (2013). As noticed in the present study, a very low genetic diversity among the Bt cotton genotypes of same species was reported by Ashraf et al. (2016) based on the EST-SSR, ISSR and morphological markers. A narrow genetic background among 61 coloured cotton lines was reported by Sun et al. (2009) as a genetic similarity value of more than 0.80 was recorded. Hence, while choosing distant parental plants for

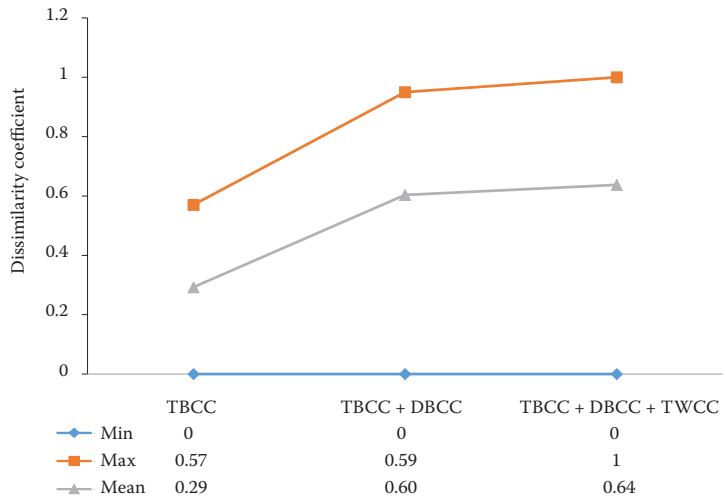

Figure 2. Change in the spectrum of the genetic diversity among cotton genotypes upon including genotypes with differences in the ploidy and fibre colour
TBCC – tetraploid brown coloured cotton genotypes; DBCC – diploid brown coloured cotton genotypes; TWCC – tetraploid white coloured cotton genotypes

https://doi.org/10.17221/12/2023-CJGPB

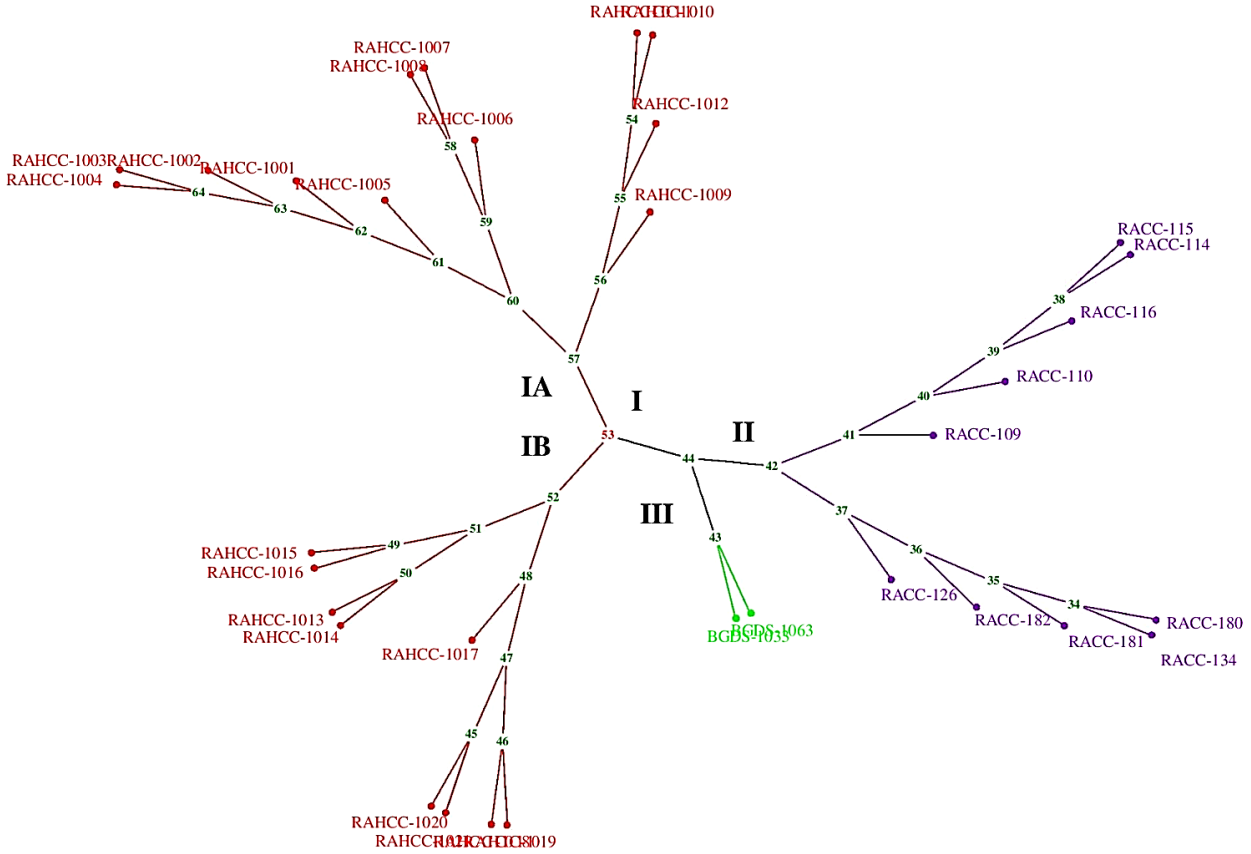

Figure 3. Phylogenetic tree constructed based on the neighbour-joining method using simple sequence repeats markers RAHCC-1001 to RAHCC-1021 – TBCC; RAC-109, 110, 114, 115, 116, 126, 134, 180,181,182 – DBCC; BGDS-1033 and BGDS-1063 – TWCC; TBCC – tetraploid brown coloured cotton genotypes; DBCC – diploid brown coloured cotton genotypes; TWCC – tetraploid white coloured cotton genotypes

future hybridisation, the genetic dissimilarity found between these genotypes might be seen as a useful guide (Noormohammadi et al. 2016).

Phylogenetic tree based on the NJ and UPGMA method. Interestingly, the phylogenetic relationship obtained based on the NJ analysis (Figure 3) is parallel with that of the UPGMA method (Figure 4). The phylogenetic analysis separated 33 genotypes into three major clusters, cluster I, II and III. The major cluster I was larger in size with 21 TBCC and it had two subclusters 1A and 1B. Subcluster 1A had 12 genotypes which involve eight dark brown and four medium brown coloured genotypes while subcluster 1B had nine genotypes (three medium and six light brown). Clearly, ten DBCC formed a separate cluster, *viz.*, cluster II, while cluster III was smaller in size with only two TWCC genotypes. The reason attributed to this could be the fact that ten genotypes belong to the *G. arboreum* species had a diploid ploidy level while the others were tetraploid. According to Noormohammadi et al. (2016), markers are able to distinguish between diploids and tetraploids with adequate precision and can also distinguish the studied genotypes from one another. Based on an NJ analysis, Ali et al. (2019) observed two clusters for 28 cotton genotypes based on 22 SSR markers, while Cardoso et al. (2021) noticed six groups for 12 coloured cotton genotypes based on the UPGMA method using Mahalanobis distance. Based on UPGMA involving tetraploid and diploid species, eight groups were observed based on the SSR, ISSR and RAPD analysis by Ghuge et al. (2018). Akin to this study, diversity among 11 coloured and 4 white linted genotypes was presented in three clusters based on the UPGMA method by Punitha and Raveendran (2008) using 32 RAPD markers while, Khan et al. (2010) used 11 coloured and 5 white cotton genotypes of tetraploid and diploid species which formed four clusters and noticed separate clusters for *G. hirsutum* and *G. arboreum* who stated that the diploid status

https://doi.org/10.17221/12/2023-CJGPB

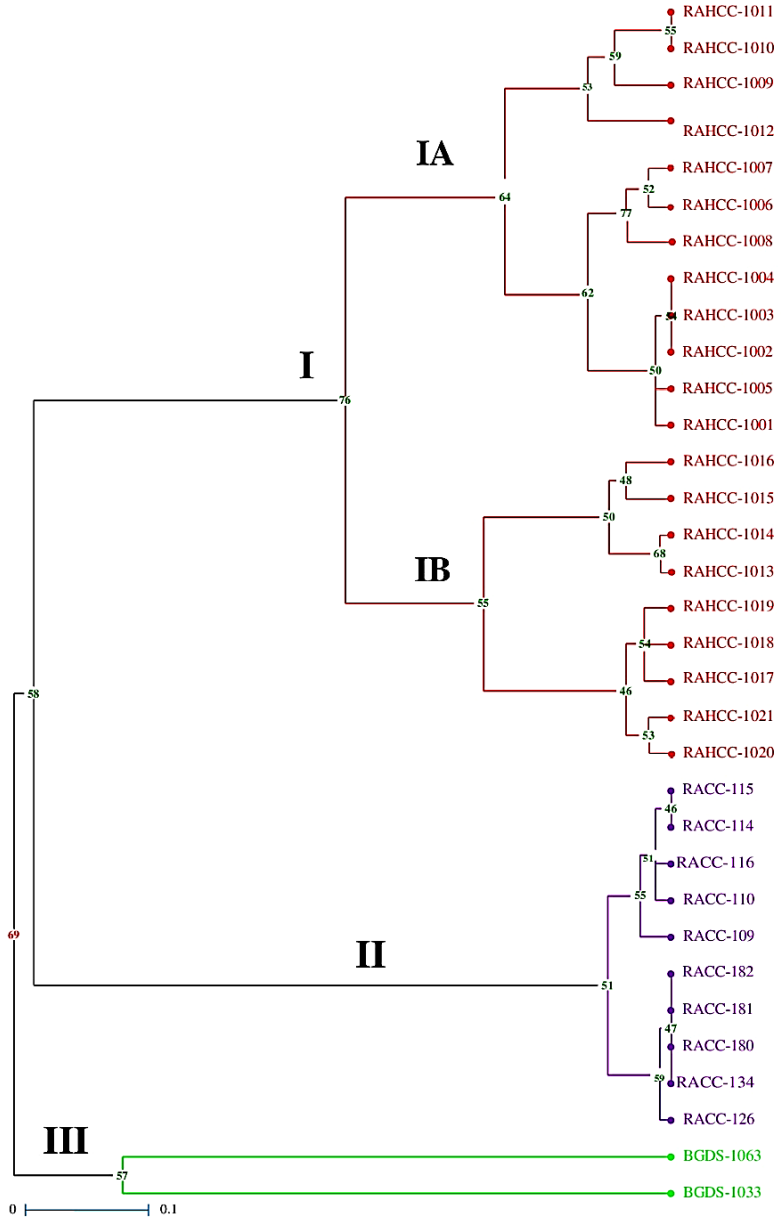

Figure 4. Dendrogram of 33 coloured cotton genotypes depicting the genetic relationships constructed based on the unweighted pair group method with arithmetic using simple sequence repeats markers

RAHCC-1001 to RAHCC-1021 – TBCC; RAC-109, 110, 114, 115, 116, 126, 134, 180,181,182 – DBCC; BGDS-1033 and BGDS-1063 – TWCC; TBCC – tetraploid brown coloured cotton genotypes; DBCC – diploid brown coloured cotton genotypes; TWCC – tetraploid white coloured cotton genotypes

of both the genotypes made them have a clear-cut differentiation from the tetraploid ones.

According to the phylogenetic tree, the genetic diversity within each group was lower which might be due to the isolation of the superior genotype from the same population of an upland cotton, but the diversity between groups is high which suggests the hybridisation between the genotypes would broaden the genetic base in cotton genotypes thereby improved cultivars can be obtained.

Principal component analysis

To better understand the relationship between the cotton genotypes, a principal component analysis (PCA) was carried out using genotypic data. The first two principal components, *viz.*, PC1 (54.80) and PC2 (19.90) together explained 74.69% of the variation (Table 3). These results were in agreement with the findings of Shah et al. (2018), Kumari and Gunasikaran (2019) and Sarwar et al. (2021). Based on a PCA, Rathinavel (2018) summarised the diversity present among the 101 genotypes in eight components which explained 83.11 percent of the variation. The biplot analysis revealed that two TWCC genotypes were plotted in the first quadrant (Group 1) (Figure 5). Twenty-one TBCC (G1 to G21) genotypes shared two quadrants, *i.e.,* the 2nd and 3rd quadrants (Group 2) while ten DBCC genotypes (G22 to G31) were plotted in the

https://doi.org/10.17221/12/2023-CJGPB

Table 3. Variation explained by the individual principal components along with their Eigen values

Principal components	Eigen values	Variance percentage	Cumulative variance percentage
PC1	54.80	54.80	54.80
PC2	19.90	19.90	74.69
PC3	11.75	11.75	86.44
PC4	6.51	6.51	92.95
PC5	4.06	4.06	97.01
PC6	0.88	0.88	97.89
PC7	0.63	0.63	98.52
PC8	0.42	0.42	98.94
PC9	0.34	0.34	99.28
PC10	0.19	0.19	99.47
PC11	0.18	0.18	99.65
PC12	0.13	0.13	99.78
PC13	0.08	0.08	99.85
PC14	0.06	0.06	99.92
PC15	0.02	0.02	99.94
PC16	0.02	0.02	99.96
PC17	0.02	0.02	99.98
PC18	0.01	0.01	99.99
PC19	0.01	0.01	100.00

4th quadrant (Group 3). Similarly, Javaid et al. (2017) found that genotypes belonging to group II and III were diverse based on a PCA using 11 SSR primers. Furthermore, Noormohammadi et al. (2016) noticed the separate plotting of diploid and tetraploid genotypes using REMAP markers. Genotypes were very distant from the origin and distinguished themselves from other genotypes indicating the diverse nature of 33 genotypes. In this way, the information produced by the genetic dissimilarity matrix, NJ analysis and UPGMA method was validated by the PCA plot. Hybridisation between the genotypes belonging to group 2 and group 3 can be planned as they are genetically distant from each other.

The present investigation revealed low to moderate genetic variability within the genotypes of same species studied, but also revealed a relatively high genetic diversity between the genotypes belonging to different species. Upland cotton (*G. hirsutum*) is known to have relatively low levels of genetic diversity. The genetic variability available, especially in the diploid species, which are the putative donors of the A and D genomes for the commercially important allotetraploid cottons (AADD), could be used as the genetic resources for broadening the genetic base and genetic improvement in cotton.

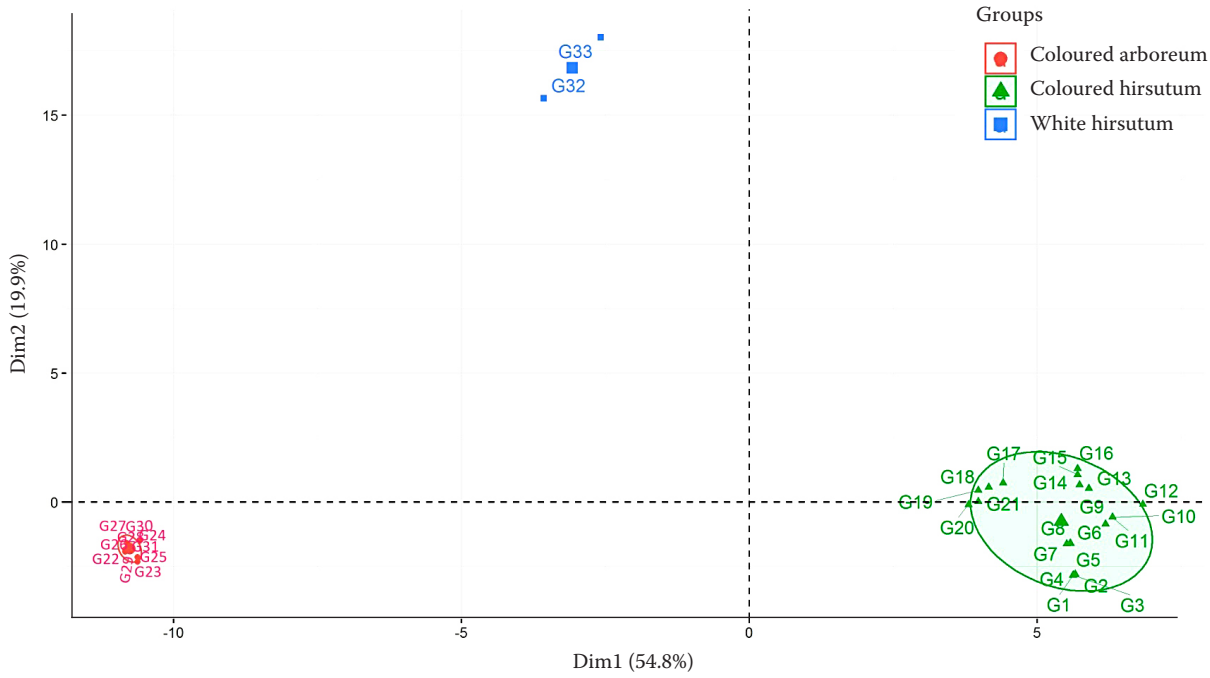

Figure 5. Principal component's biplot of the coloured cotton genotypes based on the simple sequence repeats marker data G1–G21 – TBCC; G22–G31 – DBCC; G32–G33 – TWCC genotypes; TBCC – tetraploid brown coloured cotton genotypes; DBCC – diploid brown coloured cotton genotypes; TWCC – tetraploid white coloured cotton genotypes

https://doi.org/10.17221/12/2023-CJGPB

CONCLUSION

This study concludes the prevalence of a wide spectrum of genetic diversity for the fibre colour/quality/yield traits at a molecular level. The genetic relationships indicated the grouping of genotypes according to their ploidy level and fibre colour. The diverse clusters observed offer a plethora of opportunities to employ new cross combinations between the tetraploid (RAHCC-1001 or RAH00-1006), diploid (RAC-109 or RAC-114) and white cotton genotypes to generate genetic variation for addressing the challenges of the low yield, fibre quality and colour intensity in coloured cotton genotypes.

Acknowledgements. Dr. B. Fakrudin, Professor and Head at the Department of Biotechnology and Crop Improvement is acknowledged for allowing the use of the molecular laboratory in the College of Horticulture, Bengaluru to support this molecular research.

REFERENCES

Ahmad F., Hanafi M.M., Hakim M.A., Rafii M.Y., Arolu I.W., AkmarAbdullah S.N. (2015): Genetic divergence and heritability of 42 coloured upland rice genotypes (*Oryza sativa*), as revealed by microsatellites marker and agromorphological traits. PLoS ONE, 10: 1–18.

Ali I., Khan N.U., Gul S., Khan S.U., Bibi Z., Aslam K., Shabir G., Haq H.A., Khan S.A., Hussain I., Ahmed S., Din A. (2019): Genetic diversity and population structure analysis in upland cotton germplasm. International Journal of Agriculture & Biology, 22: 669–676.

Ammad A., Muhammad A.I., Mehboob-Ur R., Andrew H.P. (2015): Estimating genetic diversity among selected cotton genotypes and the identification of DNA markers associated with resistance to cotton leaf curl disease. Turkish Journal of Botany, 39: 1033–1041.

Ashraf J., Malik W., Iqbal M.Z., Khan A.A., Qayyum A., Noor E., Abid M.A., Cheema H.M.N., Ahmad M.Q. (2016): Comparative analysis of genetic diversity among *Bt* cotton genotypes using EST-SSR, ISSR and morphological markers. Journal of Agricultural Science and Technology, 18: 517–531.

Baloch M.J., Sial P., Qurat-ul-Ain B., Arain T., Arain M.A. (2015): Assessment of heterotic effects in F_1 hybrids of cotton (*Gossypium hirsutum* L.). Pakistan Journal of Agriculture, Agricultural Engineering and Veterinary Science, 31: 193–202.

Basavaradder A.B., Maralappanavar M.S. (2014): Evaluation of eco-friendly naturally coloured *Gossypium hirsutum* L. cotton genotypes. International Journal of Plant Sciences, 9: 414–419.

Brubaker C.L., Bourland F.M., Wendel J.F. (1999): The origin and domestication of cotton. In: Smith C.W., Cothren J.T. (eds.): Cotton: Origin, History, Technology, and Production. New York, John Wiley and Sons: 3–31.

Cardoso A.B.O., Miranda M.C.C., Pimentel I.M., Leite M.A.S., Garcia M.F., Júnior V.N.B., Venâncio A.S., Cruz V.M.P., Santos A.P.X., de Sousa L.B. (2021): Genetic diversity among coloured cotton genotypes to obtain potential parent plants. Revista Cultura Agronomica, 30: 204–215.

de Rocha G.M.G., Cavalcanti J.J.V., de Carvalho L.P., dos Santos R.C., de Lima L.M. (2016): Genetic divergence of coloured cotton based on intersimple sequence repeat (ISSR) markers. African Journal of Agricultural Research, 11: 2663–2668.

Doyle J.J., Doyle J.L. (1987): A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochemical Bulletin, 19: 11–15.

Feng H., Tian X., Liu Y., Li Y., Zhang X., Jones B.J., Sun Y., Sun J. (2013): Analysis of flavonoids and the flavonoid structural genes in brown fiber of upland cotton. PLoS ONE, 8: e58820.

Ghuge S.B., Mehetre S.S., Chimote B.D., Pawar, Naik R.M. (2018): Molecular characterization of cotton genotypes using SSR, ISSR and RAPD markers in relation to fiber quality traits. Journal of Cotton Research and Development, 32: 1–12.

Gong W., Du X., Jia Y., Pan Z. (2018): Colour cotton and its utilization in China. In: Cotton Fiber: Physics, Chemistry and Biology. Los Angeles, Springer: 117–132

Günaydin G.K., Avinc O., Palamutcu S., Yavas A., Soydan A.S. (2019): Naturally coloured organic cotton and naturally coloured cotton fiber production. In: Gardetti M.A., Muthu S.S. (eds.): Organic Cotton, Textile Science and Clothing Technology. Singapore, Springer Nature.

Guo J.Y., Wang Y.Q., Wu M.G., Zhang L.M., Liu F.J., Li W.B., Liu H.F., Zhao T.P., Sun Y.R. (2004): Genetic diversity analysis of brown cotton and green cotton. Yi Chuan, 26: 63–68.

Hijazi A., Mcheik A., Rammal H., Rammal W., Annan H., Toufaily J., Hamieh T. (2015): Biosorption of methylene blue from waste water using Lebanese *Cymbopogon citratus* (citronnelle). European Scientific Journal, 11: 293–308.

Isong A., Balu A., Salihu B.Z., Isong C., Bamishaiye E. (2019): Assessment of genetic diversity using morphological and molecular markers in some cotton genotypes. Badeggi Journal of Agricultural Research and Environment, 1: 8–21.

Javaid A., Awan F.S., Azhar F.M., Khan I.A. (2017): Assessment of allelic diversity among drought-resistant cotton

https://doi.org/10.17221/12/2023-CJGPB

genotypes using microsatellite markers. Genetics and Molecular Research, 16: 1–9.

Jolliffe I. (2014): Principal component analysis. In: Lovric M. (ed.): International Encyclopedia of Statistical Science. Berlin, Heidelberg, Springer: 1094–1096.

Kencharaddi H.G., Hanchinal R.R., Patil S.S., Manjula S.M., Ravi Kulkarni., Usharani C.V., Soregoan C.D. (2018): Molecular characterization of heterotic groups of cotton through SSR markers. International Journal of Current Microbiology and Applied Sciences, 7: 426–434.

Khan A.I., Awan F.A., Sadia B., Rana R.M., Khan I.A. (2010): Genetic diversity studies among coloured cotton genotypes by using RAPD markers. Pakistan Journal of Botany, 42: 71–77.

Konopiński M.K. (2020): Shannon diversity index: A call to replace the original Shannon's formula with unbiased estimator in the population genetics studies. PeerJ, 29: e9391.

Kuang Z., Xiao C., Ilyas M.K., Ibrar D., Khan S., Guo L., Wang W., Wang B., Huang H., Li Y., Li Y., Zheng J., Saleem S., Tahir A., Ghafoor A., Chen H. (2022): Use of SSR markers for the exploration of genetic diversity and DNA finger-printing in early-maturing upland cotton (*Gossypium hirsutum* L.) for future breeding program. Agronomy, 12: 1513.

Kumari V.N., Gunasekaran P. (2019): Evaluation of genetic diversity in cotton (*Gossypium barbadense* L.) germplasm for yield and fibre attributes by principle component analysis. International Journal of Current Microbiology and Applied Science, 8: 2614–2621.

Kusuma J., Ahsan M.Z., Setiawan W., Abdullah K., Tahir M. (2018): SSR-based diversity of domesticated and local cotton (*Gossypium* spp.) populations collected in Indonesia. International Journal of Agricultural and Biology, 20: 2019–2024.

Liu K., Muse S.V. (2005): PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics, 21: 2128–2129.

Ma X., Du X.M., Sun J.L. (2003): SSR fingerprinting analysis on 18 coloured cotton lines. Journal of Plant Genetic Resources, 4: 305–310.

Malik W., Ashraf J., Iqbal M.Z., Khan A.A., Qayyum A., Abid M.A., Noor E., Ahmad M.Q., Abbasi G.H. (2014): Molecular markers and cotton genetic improvement: current status and future prospects. The Scientific World Journal, 20: 1–15.

Mishra K.K., Fougat R.S., Ballani A. (2013): Validation of fibre quality linked SSR markers derived from allotetraploid (*Gossypium hirsutum*) in diploid (*Gossypium arboreum*). International Journal of Scientific Research in Knowledge, 1: 349–357.

Nagarajan S., Purushothaman Y., Selvavinayagam M., Govindharaj P., Musthafa A. (2022): Studies on coloured cotton: Biochemical and genetic aspects. In: Abdurakhmonov I.Y. (ed.): Cotton. IntechOpen.

Noormohammadia Z., Taghavia E., Foroutanb M., Sheidaib M., Alishah O. (2013): Structure analysis of genetic diversity in tetraploid and diploid cotton genotypes. International Journal of Plant, Animal and Environmental Sciences, 3: 79–90.

Noormohammadi Z., Ibrahim-Khalili N., Somayeh G.B., Sheidai M., Alishah O. (2016): Genetic screening of diploid and tetraploid cotton cultivars based on retrotransposon microsatellite amplified polymorphism markers (REMAP). Anales de Biología, 38: 123–132.

Peakall R.O.D., Smouse P.E. (2006): GENALEX 6: Genetic analysis in excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6: 288–295.

Punitha D., Raveendran T.S. (2008): DNA fingerprinting studies in coloured cotton genotypes. Plant Breeding, 123: 101–103.

Rathinavel K. (2018): Principal component analysis with quantitative traits in extant cotton varieties (*Gossypium hirsutum* L.) and parental lines for diversity. Current Agriculture Research Journal, 6: 54–64.

Rathinamoorthy R., Parthiban M. (2017): Coloured cotton: Novel eco-friendly textile material for the future. In: Handbook of Ecomaterials. New York, Springer: 1–21.

Sarwar G.A., Nazir M., Rizwan E., Shahzadi, Mahmood A. (2021): Genetic diversity among cotton genotypes for earliness, yield and fiber quality traits using correlation, principal component and cluster analyses. Sarhad Journal of Agriculture, 37: 307–314.

Seyoum M., Du X., He S.P., Jia Y.H., Pan Z., Sun J.L. (2018): Analysis of genetic diversity and population structure in upland cotton (*Gossypium hirsutum* L.) germplasm using simple sequence repeats. Journal of Genetics, 97: 513–522.

Shah A.S., Khan J.S., Ullah K., Sayal O.U. (2018): Genetic diversity in cotton germplasm using multivariate analysis. Sarhad Journal of Agriculture, 34: 130–135.

Sun D.L., Sun J.L., Jia Y.H., Ma Z.Y., Du X.M. (2009): Genetic diversity of coloured cotton analyzed by simple sequence repeat markers. International Journal of Plant Sciences, 170: 76–82.

Sun J., Sun Y., Zhu Q.H. (2021): Breeding next-generation naturally coloured cotton. Trends in Plant Science, 26: 539–542.

Sundar R., Patil S.S., Ranganath H.M., Srivalli P., Manjula S.M., Pranesh K.J. (2014): Genetic diversity analysis in cotton (*Gossypium hirsutum* L.) based on morphological traits and microsatellite markers. International Journal of Basic and Applied Biology, 1: 19–22.

https://doi.org/10.17221/12/2023-CJGPB

Tyagi S., Verma S.K., Nehra K., Bajya D.R. (2015): Molecular marker based genetic diversity analysis in cotton using RAPD and SSR markers. Journal of Cotton Research and Development, 29: 232–236.

Wu Y., Huang L., Zhou D., Fu X., Li C., Wei S., Peng J., Kuang M. (2020): Development and application of perfect SSR markers in cotton. Journal of Cotton Research, 3: 1–8.

Zhang M.D., Zhan X.J., Zhang X.L. (2004): Polymorphic analysis on natural colour fiber cotton by RAPD. Journal of Huazhong Agricultural University, 23: 606–609.

Zhu Q.H., Spriggs A., Taylor J.M., Llewellyn D., Wilson I. (2014): Transcriptome and complexityreduced, DNA-based identification of intraspecies single nucleotide polymorphisms in the polyploid *Gossypium hirsutum* L. G3 – Genes Genomes Genetics, 4: 1893–1905.

Received: February 2, 2023
Accepted: May 3, 2023
Published online: July 12, 2023