Indian J. Genet., 79(4) 719-725 (2019)
DOI: 10.31742/IJGPB.79.4.10

Molecular diversity analysis and DNA fingerprinting of cotton varieties of India

V. Santhy*[@], M. Meshram, H. B. Santosh[@] and K. R. Kranthi[#]

Division of Crop Improvement, ICAR-Central Institute for Cotton Research, Nagpur, Maharashtra

(Received: June 2019; Revised: October 2019; Accepted: November 2019)

Abstract

Genetic diversity was assessed in 48 popular varieties of tetraploid cotton from each cultivated zone of India using 68 SSR markers distributed across linkage groups. The markers produced a total of 144 alleles with an average of 2.19 per locus. The polymorphism information content (PIC) ranged from 0.04 to 0.57 with a mean of 0.23 indicating lesser diversity in the studied material. Jaccard's similarity index based neighbourhood joining cluster analysis grouped the genotypes into three major clusters, each of which was further classified into sub-groups. Inconsistencies were observed between the clusters and known pedigree of the cultivars. A narrow genetic base was also revealed among the cotton cultivars. The SSR markers revealed a genetic similarity of 73% among the varieties studied. The DNA fingerprint developed using a selected set of 14 markers showed a probability of identical match of 2.47×10^{-3} with high goodness of fit (r^2=0.86). The identified markers have great potential in DNA fingerprinting in cotton which in future could be integrated with DUS data descriptors for effective cultivar identification and differentiation.

Key words: Cotton, genetic diversity, fingerprinting, SSR markers, varieties

Introduction

Development of improved cultivars through conventional plant breeding and genetic engineering had played a great role in cotton, an important commercial crop most widely grown across the world for fibre, fuel wood and oil. It occupies an area of 30.5 million hectares and produce 22 million tonnes of seed cotton (ICAC 2016). India ranks first both in cultivated area (11.8 million hectares) as well as production (7.27 million metric tons) with transgenic Bt cotton hybrids dominating commercial cultivation. A large area is covered by the Bt cotton hybrids in the country (Choudhary and Gaur 2015) but recently, it is observed that higher seed cost and lower net returns are forcing the farmers to abandon cotton and cultivate other crops. In this regard, importance of varietal cultivation for better returns is emphasized through technologies like high density planting system (HDPS) and Bt varieties (Venugopalan et al. 2014). With the implementation of Protection of Plant Varieties and Farmers' Rights (PPV&FR) Act, there is a growing interest in detailed characterization of released varieties to prevent their unauthorized exploitation. Concerns can be more serious with transgenics because mere introgression of one gene produces a new cultivar. India has developed as set of detailed morphological descriptors based on UPOV guidelines for all crops including cotton. Phenotyping based on enlisted DUS (Distinctness, Uniformity and Stability) traits is a pre-requisite for varieties to be registered under PPV&FR Act, 2001. Most of these traits are quantitative in inheritance and environmentally influenced. Often, there is a risk of categorising genetically different cultivars as similar or vice-versa owing to subjective assessment (Santhy and Meshram 2015). Taking into account the large number of cotton cultivars eligible to be protected, problems may arise in establishing distinctness only based on morphological descriptors. Molecular variations at DNA level can be a reliable supplement in such cases. DNA markers have been mapped in the genome and described in a searchable database making them ideal to detect vast polymorphisms in a rapid and cost effective manner.

Among the various markers systems presently available, microsatellite or simple sequence repeat (SSR) markers have become more valuable and reliable tool for genetic diversity analysis of crop varieties

*Corresponding author's e-mail: santhy100@gmail.com; [@]Equal contribution
#Present address: Technical Information Section, International Cotton Advisory Committee (ICAC), USA
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

(Rakshit et al. 2010; Ahmed et al. 2013; Abd El-Moghny et al. 2017; Tyagi et al. 2014) owing to their reproducibility, co-dominant inheritance, genome-wide presence, robustness, higher polymorphism and analytical simplicity. Molecular data generated from SSR markers has also been utilized for QTL mapping, DNA fingerprinting and genetic purity testing in cotton (Rakshit et al. 2011; Ahmed et al. 2013; Ashraf et al. 2016).

Exclusive and detailed SSR profiling of promising Indian cotton varieties developed from ICAR-Central Cotton Research Institute (CICR) as well as State Agricultural Universities (SAU's) have not been attempted so far. To address this knowledge gap, the study was planned to assess the molecular diversity available among the tetraploid cotton varieties of India and to identify a set of SSR markers to develop DNA fingerprint which could in future be integrated with DUS traits recorded morphologically.

Materials and methods

Forty eight public-sector released popular varieties of tetraploid cotton (Table 1) released for all the three cotton growing zones of India were selected for the study. Among these, 46 varieties belonged to upland cotton (*G. hirsutum*) and 2 belonged to Egyptian cotton (*G. barbadense*). Our investigation included exclusively the varieties released and notified by Central Varietal Release and Notification Committee (CVRC) for north, central and southern cotton cultivation zones of India. Apart from their utilization in cotton breeding, some of these varieties are regularly used as reference varieties in routine DUS tests conducted for candidate genotypes seeking variety protection. The breeder seeds of above varieties were procured from authorized breeders and planted in experimental farm of ICAR-CICR, Nagpur (Coordinates 21°01'45.5"N, 79°03'43.7"E) during *kharif* 2014-15 and 2015-16. Each variety was monitored for possible off types at different stages of crop growth and the true to type plants were selected for molecular study. Extraction of genomic DNA, its quantification and quality checking, SSR genotyping, scoring of amplicons and data analysis was carried out as per Abd El-Moghny et al. (2017) and DNA fingerprint was developed as per Choudhury et al. (2001) and Rakshit et al. (2011). The goodness of fit of DNA barcode fingerprint was tested through cophenetic correlation coefficient calculated using 'R' statistical package.

Results and discussion

SSR marker analysis

Assessment of genetic diversity among the cotton varieties is important to know the diverse parents for

Table 1. Name and source of varieties of upland cotton (*Gossypium hirsutum*) employed in the study

Name of the varieties	Source of material
LRA5166[1], Anjali[3], Suraj[5], Surabhi[6], Supriya[7], Sumangala[8], MCU5[13], MCU5VT[14], Suvin*,[47]	ICAR-Central Institute for Cotton Research, Regionl Station, Coimbatore
CNHO12[2], Pratima[4], Arogya[11]	ICAR-Central Institute for Cotton Research Nagpur
Sahana[9], Abhadita[10], RMPBS155[46]	University of Agricultural Sciences, Dharwad
Narasimha[12]	Regional Agricultural Research Station, Nandyal
MCU10[15], MCU12[16]	Tamil Nadu Agricultural University Coimbatore
AKH8828[19], AKH081[20], DHY286[17], PKV Rajat[45]	Punjab Rao Deshmukh Krishi Vidyapeeth, Akola
Khandwa2[21], JK4[24]	Jawaharlal Nehru Krishi Vishwavidyalaya, Khandwa
NH545[22], NH615[23]	Marathawada Agricultural University, Nanded
Gcot10[25], Gcot16[26], Gcot20[27] Gcot18[28], Gcot12[29], Deviraj[43]	Navsari Agricultural University, Surat
Bikeneri Narma[18], Ganganagar Ageti[30], RS875[36], RS2013[37], RS810[38], RST9[39]	Rajasthan Agricultural University, Sriganganagar
F1054[32], F1378[33], F846[34], F1861[35]	Punjab Agricultural University, Faridkot
HS6[40], H1226[41]	Chaudhary Charan Singh Haryana Agricultural University, Hisar
KC3[42]	Tamil Nadu Agricultural University, Kovilpatti
Kanchana[44]	Regional Agricultural Research Station, Lam
JLH168[31] and Sujata*,[48]	Not available

Superscript designate the serial number of varieties; * = belongs to Egyptian cotton (*G. barbadense*)

cotton improvement. All the 48 cotton varieties were profiled for DNA polymorphism using 250 SSR markers sourced from cotton marker database. Previous reports on genetic diversity analysis among Indian cotton varieties of *G. hirsutum* are largely using ISSR, RAPD and SSR markers (Rana et al. 2005; Salunkhe et al. 2009; Chaudhary et al. 2010; Rakshit et al. 2010)). In the present study the popular released varieties from each cultivated zone of India were included. Among the 250 SSR markers, only 68 markers were polymorphic (Table 2) with the percent polymorphism of 27.20%. The banding pattern of polymorphic marker TMB1181 has been provided in Fig. 1. Low level of SSR polymorphism in *G. hirsutum* was reported earlier by several researchers (Ehsan et al. 2013; Abbas et al. 2015; Bertini et al. 2006; Rakshit et al. 2010). The 68 markers detected a total of 144 alleles with an average of 2.19 alleles per locus. Majority of the polymorphic markers (83.82%) detected 2 alleles per locus. Islam et al. (2012) and Ambreen et al. (2013) also observed similar results in cotton. Two markers (BNL 4096 and BNL 3280) revealed highest number of alleles (4) and nine markers revealed 3 alleles (Table 2). The higher number of alleles detected implies greater allelic diversity in those loci. The markers which detected more number of alleles incidentally had higher PIC (r^2=0.53**) and gene diversity (r^2=0.48**) values.

The PIC values ranged from 0.02 (NAU2691) to 0.57 (TMB2295) with a mean of 0.23. Twenty three markers detected 2-4 alleles per loci and showed medium to high PIC values (0.35-0.50) indicating their usefulness for discriminating varieties. PIC values depends on many factors such as breeding behaviour of the species, genetic diversity in the collection, size of the collection, sensitivity of genotyping method and location of primers in the genome used for study (Kalivas et al. 2011; Singh et al. 2013). Low allele detection and low PIC values was observed even when high resolution metaphor agarose/PAGE gels was employed for separating the amplicons (Bertini et al. 2006; Yu et al. 2012). When the study material was diverse such as wild accessions and hybrids, the PIC values can be more than 0.80 as observed by Ahmed et al. (2013) and Zhang et al. (2013). Lower PIC value observed in the present study may be the result of the lesser diversity available in the study material which consisted of only released varieties. SSR markers being co-dominant in inheritance are known to detect heterozygosity. Heterozygosity is a measure of genetic variation within a population. Average heterozygosity detected by the markers ranged from 0.00 (35.29% markers) to 1.00 (DPL168, BNL686 and GH434 *i.e.,* 5.88% of markers) with an average 0.28 which is relatively higher for an often cross pollinated crop like cotton. As many as 44 (64.70%) markers showed heterozygosity in at least one variety. The heterozygosity of $\geq$15% was detected by 27 markers (39.71%; Supplementary Fig. 1) while remaining markers (25%) detected lower level of heterozygosity. Earlier, Rakshit et al. (2010) observed 14 out of 38 of markers (36.8%) that detected

Table 2. Marker details and parameters of molecular diversity analysis

Marker	Genomic location	Major allele frequency	Allele No.	Gene diversity	Hetero-zygosity	PIC
BNL1531	AD_Chr08	0.53	2	0.50	0.72	0.37
BNL2275	N/A	0.88	2	0.22	0.00	0.19
BNL2496	AD_Chr03	0.91	2	0.17	0.00	0.15
BNL2634	AD_Chr12	0.59	2	0.48	0.50	0.37
BNL2709	AD_Chr12	0.59	2	0.48	0.20	0.37
BNL3103	AD_Chr01	0.96	2	0.08	0.00	0.08
BNL3259	AD_Chr13	0.79	2	0.33	0.00	0.28
BNL3280@	AD_Chr18&20	0.71	4	0.44	0.58	0.39
BNL3408	AD_Chr03	0.65	2	0.36	0.18	0.29
BNL3792	AD_Chr02	0.98	2	0.04	0.00	0.04
BNL3902	AD_Chr13	0.92	2	0.15	0.00	0.14
BNL3971@	AD_Chr05	0.98	2	0.04	0.04	0.04
BNL3992@	AD_Chr06	0.47	3	0.56	0.94	0.46
BNL 409	AD_Chr13	0.97	2	0.06	0.07	0.06
BNL4096	AD_Chr09	0.80	4	0.33	0.40	0.29
BNL686@	AD_Chr06	0.50	3	0.52	1.00	0.40
BNL786	AD_Chr02	0.96	2	0.08	0.00	0.08
BNL830	AD_Chr15	0.68	2	0.43	0.00	0.34
BNL2448	AD_Chr09	0.97	2	0.06	0.06	0.06
BNL2734	AD_Chr16	0.91	2	0.16	0.00	0.15
BNL3008	AD_Chr01	0.79	2	0.33	0.02	0.27
BNL3028@	N/A	0.47	3	0.59	0.98	0.50
BNL3371	AD_Chr17	0.89	2	0.19	0.17	0.17

BNL3649	AD_Chr04	0.96	2	0.08	0.08	0.08
BNL3903	AD_Chr19&25	0.81	2	0.30	0.00	0.26
BNL4028	AD_Chr09	0.97	2	0.06	0.06	0.06
BNL4029	AD_Chr13&18	0.77	2	0.35	0.00	0.29
CIR009	AD_Chr02	0.92	2	0.15	0.00	0.14
CM 43	AD_Chr08&24	0.83	2	0.28	0.00	0.24
DPL0026	AD_Chr11	0.98	2	0.04	0.04	0.04
DPL0071[@]	AD_Chr10	0.52	3	0.51	0.96	0.39
DPL0094	AD_Chr09	0.52	2	0.50	0.96	0.37
DPL0135[@]	AD_chr. 20	0.94	2	0.12	0.13	0.11
DPL0209[@]	AD_Chr03	0.60	3	0.53	0.79	0.44
DPL168[@]	AD_Chr08	0.50	3	0.51	1.00	0.39
DPL196	AD_Chr06	0.98	2	0.04	0.00	0.04
DPL398	AD_Chr13	0.93	2	0.14	0.15	0.13
DPL468	AD_Chr10	0.63	2	0.47	0.00	0.36
DPL522	AD_Chr07	0.79	2	0.33	0.00	0.28
GH434[@]	AD_Chr07	0.50	3	0.52	1.00	0.41
GH486	AD_Chr06	0.62	2	0.47	0.21	0.36
HAU0058	N/A	0.89	2	0.20	0.19	0.18
JESPR0297	N/A	0.75	2	0.38	0.00	0.30
JESPR050	AD_Chr12	0.94	2	0.12	0.13	0.11
JESPR114	AD_Chr06	0.96	2	0.08	0.00	0.08
MUCS0164	AD_Chr01&15	0.75	2	0.38	0.00	0.30
MUCS400	AD_Chr09	0.88	2	0.22	0.00	0.19
NAU1141	AD_Chr13	0.89	2	0.20	0.23	0.18
NAU1167[@]	AD_Chr13	0.67	2	0.44	0.00	0.35
NAU2083	AD_Chr01	0.93	2	0.14	0.15	0.13
NAU2140	AD_Chr12	0.85	2	0.25	0.29	0.22
NAU2152	AD_Chr07	0.98	2	0.04	0.00	0.04
NAU2265	AD_Chr03	0.88	2	0.22	0.00	0.19
NAU2277	AD_Chr02	0.53	2	0.50	0.94	0.37
NAU2691	AD_Chr11	0.99	2	0.02	0.02	0.02
NAU2980	AD_Chr18	0.92	2	0.15	0.00	0.14
NAU3207	AD_Chr04	0.97	2	0.06	0.06	0.06
NAU3365	AD_Chr06	0.6	2	0.48	0.79	0.36
NAU5013	AD_Chr03	0.94	2	0.12	0.13	0.11
NAU997	AD_Chr05	0.92	2	0.14	0.15	0.13
TMB0799	AD_Chr03	0.54	2	0.5	0.92	0.37
TMB1181	AD_Chr01&15	0.57	2	0.49	0.08	0.37
TMB1288[@]	AD_Chr10	0.67	2	0.44	0.67	0.35
TMB1295	AD_Chr09	0.96	2	0.08	0.08	0.08
TMB1484	AD_Chr01	0.58	2	0.49	0.83	0.37
TMB1578[@]	AD_Chr05	0.48	3	0.56	0.94	0.46
TMB2295[@]	AD_Chr18	0.40	3	0.64	0.79	0.57
TMB2303	AD_Chr07	0.95	2	0.10	0.10	0.09
Mean	0.78	2.19	0.28	0.28	0.23	
Min	0.40	2.00	0.02	0.00	0.02	
Max	**0.99**	**4.00**	**0.64**	**1.00**	**0.57**	

[@]Markers were used for development of DNA fingerprint of cotton varieties (Fig. 2)

heterozygosity in cotton. Estimated mean heterozygosity among the genotypes was found to be higher compared to those observed by Rakshit et al. (2010) and Tyagi et al. (2014). The higher heterozygosity for few markers might be due to the amplification of similar sequences in different genomic regions mainly due to duplications during evolution of tetraploid cotton. Cotton being a complex allotetraploid, extensive heterogeneity and inherent residual heterozygosity is bound to exist (Zhang et al. 2013). Cryptic genetic variations at DNA level also contribute to higher heterozygosity (Rakshit et al. 2011) apart from limited self pollinations and rare pollen contamination. The gene diversity is the probability of two randomly chosen alleles being different from a population. The gene diversity ranged from 0.02 (NAU2691) to 0.64 (TMB2295). The mean gene diversity value of 0.28 was observed among 48 tetraploid cultivars.

Cluster analysis

The neighbourhood joining cluster analysis was carried out for 48 varieties which put them into 11 distinct clusters (Fig. 3). Cluster I was the largest cluster with 9 varieties followed by cluster IX with 7 varieties. Cluster VII was the smallest cluster with single variety, Narasimha The *G. barbadense* varieties *viz.,* Suvin and Sujatha were placed distinctly from all the *G. hirsutum* varieties. Clustering based on SSR molecular profile of the varieties by and large, did not match with the pedigree or source of the varieties. The varieties having LRA5166 (cluster IX) in the pedigree *viz.,* Suraj (cluster I), Anjali (cluster VII) and Narasimha (cluster VIII) found grouped in separate clusters. Varieties which had MCU5 (cluster III) in their parentage *viz.,* Surabhi and Supriya were also grouped separately in different clusters. These disagreements between varietal pedigree and molecular clustering may arise due to use of limited number of SSR markers. Large numbers of SSRs are needed to capture the complete diversity available in a large, complex, allotetraploid cotton genome. In a similar study conducted in date palm (Phoenix dectylifera) Sharifi et al. (2018) failed to observe clustering of genotypes as per their place of collection. The isogenic lines namely MCU5 and MCU5VT were also found clustered separately. It might

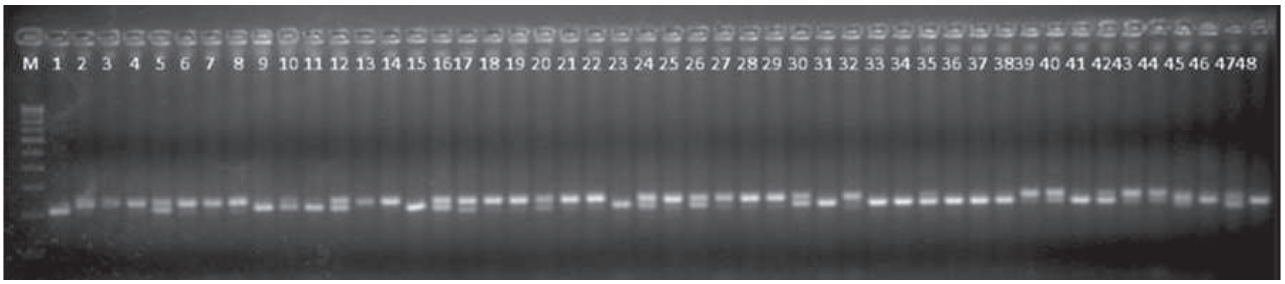


Fig. 1. Molecular profile of cotton varieties as revealed by SSR marker TMB1181. (M - 100bp DNA ladder; Number 1 to 48 corresponds to varieties as listed in Table 1)

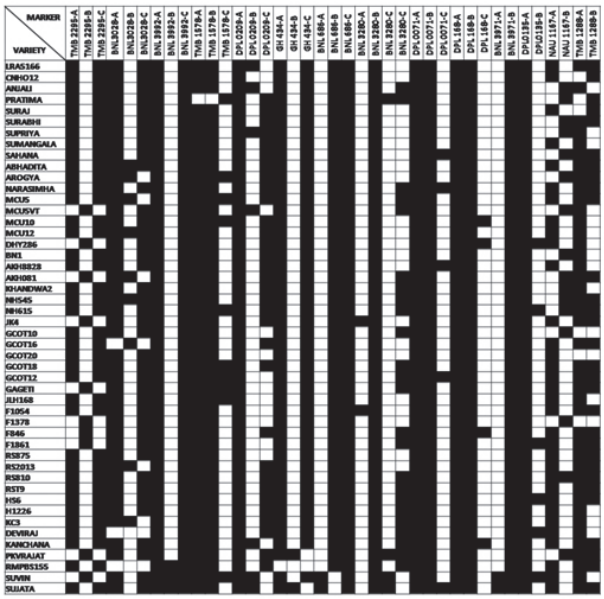


Fig. 2. Diagrammatic presentation of DNA fingerprints of 48 cotton varieties using 14 most informative markers (*denoted with @ in Table 2*)

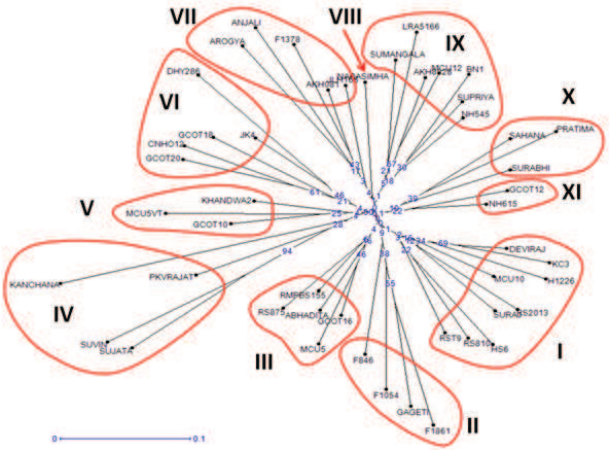


Fig. 3. Clustering of cotton varieties of India based on genetic dissimilarity

be due to the DNA polymorphisms available in the non-genic region of the genome which again supports the potential of these SSR markers in cultivar identification and differentiation. The varieties developed by ICAR-CICR *viz.,* Suraj, Surabhi, Supriya, LRA5166, Anjali, Arogya, Kanchana, CHNO12 etc were found grouped in different clusters indicating existence of sufficient variability among them. This might be due to the utilization of diverse germplasm in breeding programmes across different centres of CICR and access to broader germplasm base. It was observed that varieties with wide adaptability (e.g. LRA 5166) have been repeatedly utilized in crossing with local breeding strains to evolve newer varieties. Recurrent use of limited genotypes results in narrow genetic base (Zhang et al. 2013).

An average of 27% dissimilarity was observed among 48 varieties under study indicating relatively higher similarity among the cultivars. The highest similarity of 90% was observed between RAMPBS 155 and G. Cot 16, followed by 89% between Deviraj and KC3 (Supplementary Table S1). The highest dissimilarity of 46% was observed Kanchana and F 1378, followed by 45% between Suvin and Kanchana. Variety Kanchana showed high dissimilarity co-efficient with most of the *G. hirsutum* varieties. Both the *G. barbadense* varieties *viz.,* Suvin and Sujatha showed consistently high dissimilarity (0.30-0.45) with all *G. hirsutum* varieties. A dissimilarity of 0.23 was observed between two morphologically similar varieties of *G. barbadense* (Suvin and Sujatha) indicating the potential of identified SSR markers in cultivar differentiation. Although, *G. barbadense* and *G. hirsutum* may share a common progenitor, the two species substantially differ (Liu et al. 2015). The low average dissimilarity index observed in the present study indicates a considerable degree of relatedness between the cultivars. A low genetic distance (0.19) was also observed by Tyagi et al. (2014) among *G. hirsutum* germplasm accessions of US. In a study on diversity of Bt cotton genotypes of Pakistan, Ashraf et al. (2016) observed high level of genetic similarity among the genotypes which was attributed to the monoculture

of a small number of successful varieties and their recurrent use in Bt cotton breeding programme.

Few markers, inspite of showing low PIC, were found to be very informative since they were able to distinguish some selected varieties/species. The marker BNL3971 clearly differentiated varieties of *G. barbadense* (Suvin and Sujatha) from all the *G. hirsutum* varieties. The marker BNL3971 is located on chromosome 02, very near to the marker MUCS620 which is known to be closely associated with fibre length (Frelichowski et al. 2006). It would be useful in separating species at lint level there by providing a means for identifying lint admixtures.

DNA fingerprinting of Indian cotton varieties

One of the most important applications of molecular fingerprinting is to identify a marker or set of markers which can differentiate a particular genotype from the remaining genotypes. To decide an appropriate DNA fingerprint for the cotton varieties, the ability of fingerprint to discriminate the given set of genotypes was assessed through probability of identical match by chance (Choudhury et al. 2001). Analysis of all the 68 polymorphic markers revealed an average dissimilarity co-efficient of 0.27 among 48 cotton varieties. The DNA fingerprint thus developed showed the probability of identical match of 8.90×10^{-13} indicating potential of these SSR markers in cultivar identification and differentiation. When a selected set of 19 markers (17 having PIC $\geq$0.37 and two markers *viz.,* BNL3971 and DPL0135 having unique genotype/ specific bands) were analyzed, the probability of identical match was found to be 4.70×10^{-4}. Cophenetic correlation coefficient (r^2) was calculated to assess the credibility of dendrogram and DNA fingerprint developed using pair-wise distances. The co-efficient was found to be 0.80 when all the 68 markers were analyzed together while, it got improved to 0.84 when selected set of 19 markers were employed indicating better goodness of fit of later. Earlier, Raskhit et al. 2010 differentiated a set of 47 upland cotton accessions using 38 SSR markers with probability of identical match of 3.98×10^{-8} with a goodness of fit of 0.73. Though, utility of DNA fingerprints in registration and protection of plant varieties is debated (Bhat 2008), the ability of SSR markers in cultivar differentiation is widely accepted (Kumar et al. 2008; Rakshit et al. 2010; 2011). Presently, owing to reduction in cost of genome sequencing, the sequence based genotyping/ fingerprinting is getting popular (Kumar et al. 2018). In this study, when analysis was performed with a minimal set of 14 markers, the fingerprint (Fig. 2) was able to differentiate all 48 varieties with probability of identical match of 2.47×10^{-3} with the highest level of confidence (r^2= 0.86). The polymorphic markers identified in the present study will be of immense utility in cotton improvement. The DNA fingerprint developed for popular cotton varieties of India will be helpful for unambiguous identification of cotton varieties and their protection against unauthorized exploitation.

Authors' contribution

Conceptualization of research (VS, HBS, KRK); Designing of the experiments (VS, HBS, KRK); Contribution of experimental materials (VS, HBS); Execution of field/lab experiments and data collection (VS, HBS, MM); Analysis of data and interpretation (HBS, VS); Preparation of manuscript (VS, HBS, MM).

Declaration

The authors declare no conflict of interest.

Acknowledgement

The financial support provided by Indian Council of Agricultural Research (ICAR) for the study is gratefully acknowledged.

References

Abbas A., Iqbal M. A., Rahman M. and Paterson A. H. 2015. Estimating genetic diversity among selected cotton genotypes and the identification of DNA markers associated with resistance to cotton leaf curl disease. Turk. J. Bot., **39**: 1033-1041.

Abd El-Moghny A. M., Santosh H. B., Raghavendra K. P., Sheeba J. A., Singh S. B. and Kranthi K. R. 2017. Microsatellite Marker based Genetic diversity analysis among cotton (*Gossypium hirsutum*) accessions differing for their response to drought stress. J. Plant Biochem. Biotechnol., **26**(3): 366-370.

Ahmed M., Guo H., Huang C., Zhang X. and Lin Z. 2013. Selection of core SSR markers for fingerprinting upland cotton cultivars and hybrids. Aus. J. Crop Sci., **7**(12): 1912-1920.

Ambreen I., Ali S., Ijaz U., Smiullah and Tayyaba S. 2013. Molecular characterization of cotton using simple sequence repeat (SSR) markers and application of genetic analysis. Int. J. Genet. Mol. Biol., **5**(4): 49-53.

Ashraf J., Malik W. U., Iqbal M. Z., Khan A., Qayyum A., Noor E., Abid M. A., Cheema H. M. N. and Ahmed M. Q. 2016. Comparative analysis of genetic diversity among bt cotton genotypes using EST-SSR, ISSR and morphological markers. J. Agr. Sci. Tech., **18**: 517-531.

Bertini C. H. C. M., Schuster I., Sediyama T., Barrosi E. G. D. and Moreira M. A. 2006. Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet. Mol. Biol., **29**: 321-329.

Bhat K. V. 2008. DNA Fingerprinting and cultivar identification In: Advances in data analytical techniques. (Eds: Prasad R., Gupta V. K., Bhar L. M. and Bhatia V. K.), New Delhi. pp: VI101-109.

Choudhary B. and Gaur K. 2015. Biotech cotton in India, 2002 to 2014. ISAAA Series of Biotech Crop Profiles. ISAAA: Ithaca, USA.

Choudhury P. R., Kohli S., Srinivasan S., Mohapatra T. and Sharma R. P. 2001. Identification and classification of aromatic rice based on DNA fingerprinting. Euphytica, **118**: 243.

Chaudhary L., Sindhu A., Kumar M., Kumar R. and Saini M. 2010. Estimation of genetic divergence among some cotton varieties by RAPD analysis. J. Plant Breed. Crop Sci., **2**(3): 039-043.

Ehsan B., Haque A., Younas M., Shaheen T., Huma T., Sattar S., Idrees S. and Iqbal Z. 2013. Assessment of genomic diversity of cotton (*Gossypium hirsutum*) genotypes using simple sequence repeats markers through genetic analysis software. Int. J. Agric. Biol., **15**: 968-972.

Frelichowski J. E. Jr., Palmer M. B., Main D., Tomkins J. P., Cantrell R. G., Stelly D. M., Yu J., Kohel R. J. and Ulloa M. 2016. Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol. Genet. Genomics, **275**(5): 479-491.

International Cotton Advisory Committee (ICAC) 2016. 74th Plenary Committee. From farm to fabric: the many faces of cotton, Mumbai, India.

Islam M. N., Rezwan M. M., Rahman M., Mirza H., Naimul I. S. M. and Lutfur R. 2012. DNA fingerprinting and genotyping of cotton varieties using SSR Markers. Not. Bot. Horti. Agrobo., **40**(2): 261-265.

Kalivas A., Xanthopoulos F., Kehagia O. and Tsaftaris A. S. 2011. Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers. Genet. Mol. Res., **10**: 208-217.

Kumar B., Rakshit S., Singh R. D., Gadag R. N., Nath R., Paul A. K. and Wasialam 2008. Diversity analysis of early maturing elite Indian maize (*Zea mays* L.) inbred lines using Simple Sequence Repeat. J. Plant Biochem. Biotechnol., **17:** 133.

Kumar S., Vats S., Barman P., Tyagi N., Kumari R., Bangar P., Tiwari B., Sachdeva S., Gaikwad Ambika B. and Bhat K. V. 2018. Whole genome SNP identification and validation in *Cucumis melo* L. cultivars using genome resequencing approach. Indian J. Genet., **78**(4): 478-486. DOI: 10.31742/IJGPB.78.4.10.

Liu X., Zhao B., Zhao X. and Chen Y. 2015. Gossypium barbadensegenome sequence provides insight into the evolution of extra long staple fiber and specialized metabolites. Sci. Rep., **5**: 14139.

Rakshit A., Rakshit S., Santhy V., Gotmare V. P., Mohan, P., Singh, V. V., Singh S., Singh, J., Balyan H. S., Gupta P. K. and Bhat S. R. 2010. Evaluation of SSR markers for the assessment of genetic diversity and fingerprinting of *Gossypium hirsutum* accessions. J. Plant Biochem. Biotechnol., **19**(2): 153-160.

Rakshit S., Santosh H. B., Sekhar J. C., Nath R., Shekhar M., Chikkappa G. K., Gadag R. N. and Dass S. 2011. Analyses of genetic diversity among maize inbred lines differing for resistance to pink borer and post-flowering stalk rot. J. Plant Biochem. Biotechnol., **20**(2): 173-181.

Rana M., Singh V. and Bhat K. 2005. Assessment of genetic diversity in upland cotton (*Gossypium hirsutum* L.) breeding lines by using amplified fragment length polymorphism (AFLP) markers and morphological characteristics. Genet. Resour. Crop Evol., **52**: 989-997.

Salunkhe S. N. and Deshmukh Y. A. 2009. Molecular characterization of elite cotton cultivars using ISSR markers. Asian J. Bioscience, **4**(2): 249-253

Santhy V. and Meshram M. 2015. Widening the character base for distinctness in cotton plants. Curr. Sci., **109**(11): 1913.

Sharifi M., Sheidai M. and Koohdar F. 2018. Genetic fingerprinting of date palm (*Pheonix dactylifera* L.) by using ISSR and cpDNA sequences. Indian J. Genet., **78**(4): 507-514. DOI: 10.31742/IJGPB. 78.4.13.

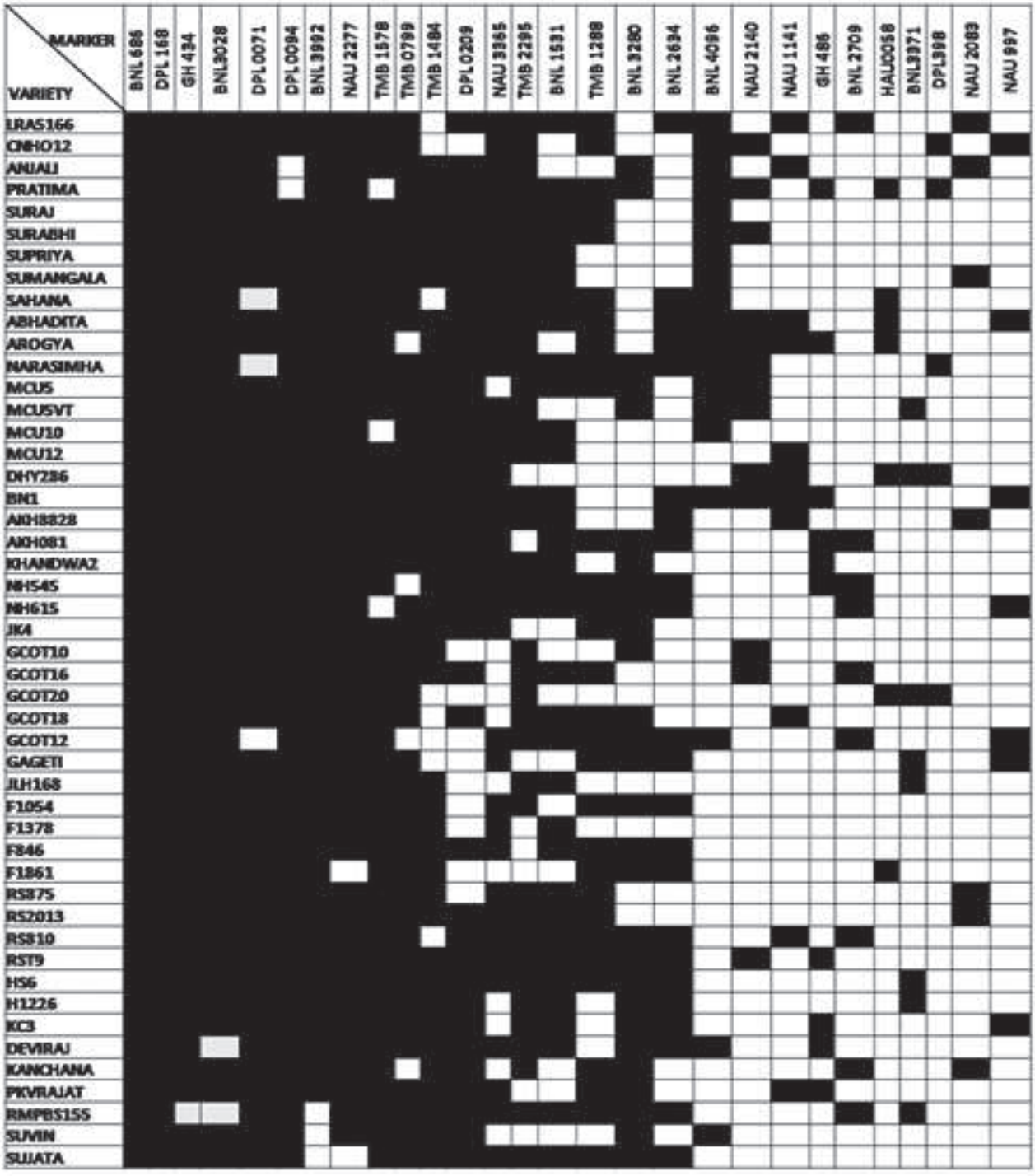
Singh N., Choudhury D. R., Singh A. K., Kumar S., Srinivasan K. and Tyagi R. K. 2013. Comparison of SSR and SNP Markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS ONE, **8**(12): e84136.

Tyagi P., Gore M. A., Bowman D. T., Campbell B. T., Udall J. A. and Kuraparthy V. 2014. Genetic diversity and population structure in the US upland cotton (*Gossypium hirsutum* L.). Theor. Appl. Genet., **127**: 283-295.

Venugopalan M. V., Kranthi K. R., Blaise D., Lakde S. and Shankaranarayanan K. 2014. High density planting system in cotton – The Brazil experience and Indian initiatives. Cotton Res. J., **5**(2): 172-185.

Yu Z. J., Fang D. D., Kohel R. J., Ulloa M., Hinze L. L., Perci R. G., Zhang J., Chee P., Scheffler B. E. and Jones D. C. 2012. Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. Euphytica, **187**: 203-213.

Zhang Y. C., Kuang M., Yang,W. H., Xu H. X., Zhou D. Y., Wang Y. Q., Feng, X. A., Su C. and Wang F. 2013. Construction of a primary dna fingerprint database for cotton cultivars. Genet. Mol. Res., **12**(2): 1897-906.



Supplementary Fig. 1. Diagrammatic representation of heterozygosity in 48 genotypes as detected by 28 SSR markers

Supplementary Table S1. Genetic dissimilarity among cotton varieties of India as revealed by SSR markers

Jaccard's dissimilarity index	LRA5166	CNHO12	Anjali	Pratima	Suraj	Surabhi	Supriya	Sumangala	Sahana	Abhadita	Arogya	Narasimha	MCU5	MCU5VT	MCU10
CNHO12	0.31														
ANJALI	0.26	0.34													
PRATIMA	0.34	0.29	0.31												
SURAJ	0.24	0.29	0.32	0.22											
SURABHI	0.25	0.27	0.31	0.21	0.17										
SUPRIYA	0.25	0.26	0.23	0.29	0.24	0.22									
SUMANGALA	0.20	0.32	0.23	0.28	0.19	0.24	0.17								
SAHANA	0.25	0.25	0.26	0.20	0.23	0.21	0.21	0.27							
ABHADITA	0.25	0.18	0.30	0.23	0.23	0.23	0.22	0.23	0.17						
AROGYA	0.32	0.32	0.27	0.29	0.35	0.30	0.29	0.34	0.23	0.23					
NARASIMHA	0.30	0.23	0.27	0.25	0.23	0.23	0.23	0.23	0.22	0.19	0.25				
MCU5	0.27	0.27	0.30	0.25	0.20	0.23	0.25	0.24	0.22	0.17	0.30	0.23			
MCU5VT	0.35	0.30	0.29	0.30	0.30	0.28	0.20	0.30	0.27	0.29	0.33	0.31	0.27		
MCU10	0.26	0.29	0.31	0.23	0.15	0.19	0.21	0.19	0.23	0.22	0.27	0.20	0.20	0.27	
MCU12	0.24	0.31	0.31	0.34	0.29	0.29	0.17	0.20	0.28	0.20	0.29	0.27	0.27	0.28	0.21
DHY286	0.37	0.29	0.36	0.34	0.32	0.30	0.29	0.33	0.33	0.25	0.36	0.26	0.35	0.30	0.27
BN1	0.28	0.28	0.26	0.26	0.26	0.20	0.16	0.21	0.22	0.22	0.26	0.24	0.29	0.29	0.21
AKH8828	0.22	0.31	0.25	0.31	0.26	0.20	0.19	0.19	0.27	0.22	0.28	0.25	0.29	0.29	0.18
AKH081	0.28	0.30	0.25	0.27	0.27	0.22	0.22	0.29	0.25	0.22	0.21	0.19	0.24	0.25	0.25
KHANDWA2	0.31	0.23	0.22	0.24	0.20	0.22	0.18	0.24	0.23	0.20	0.27	0.18	0.25	0.23	0.20
NH545	0.24	0.30	0.23	0.26	0.25	0.20	0.13	0.21	0.23	0.23	0.24	0.22	0.23	0.25	0.22
NH615	0.28	0.23	0.29	0.20	0.22	0.18	0.21	0.24	0.20	0.13	0.29	0.23	0.20	0.29	0.19
JK4	0.29	0.19	0.34	0.32	0.25	0.24	0.21	0.27	0.22	0.16	0.33	0.24	0.22	0.25	0.23
GCOT10	0.28	0.22	0.30	0.28	0.23	0.22	0.23	0.27	0.24	0.23	0.32	0.24	0.21	0.19	0.24
GCOT16	0.27	0.24	0.34	0.27	0.23	0.21	0.21	0.23	0.21	0.12	0.27	0.21	0.12	0.29	0.21
GCOT20	0.31	0.18	0.32	0.34	0.25	0.29	0.26	0.24	0.31	0.26	0.37	0.29	0.26	0.27	0.26
GCOT18	0.31	0.24	0.33	0.28	0.26	0.27	0.29	0.29	0.26	0.21	0.30	0.22	0.25	0.36	0.24
GCOT12	0.21	0.19	0.26	0.27	0.25	0.21	0.22	0.26	0.18	0.16	0.27	0.20	0.20	0.30	0.22
GAGETI	0.28	0.33	0.39	0.32	0.29	0.30	0.31	0.32	0.29	0.24	0.37	0.30	0.30	0.33	0.26
JLH168	0.27	0.27	0.22	0.32	0.26	0.26	0.20	0.22	0.28	0.23	0.27	0.22	0.21	0.24	0.18
F1054	0.30	0.31	0.38	0.34	0.27	0.25	0.29	0.30	0.30	0.25	0.32	0.31	0.25	0.33	0.23
F1378	0.33	0.31	0.28	0.33	0.28	0.25	0.28	0.33	0.29	0.29	0.26	0.26	0.28	0.26	0.24
F846	0.33	0.30	0.34	0.27	0.25	0.23	0.23	0.22	0.27	0.22	0.36	0.26	0.26	0.24	0.23
F1861	0.30	0.38	0.39	0.38	0.35	0.33	0.34	0.34	0.33	0.28	0.34	0.33	0.28	0.34	0.31
RS875	0.27	0.27	0.32	0.30	0.23	0.23	0.23	0.22	0.22	0.17	0.33	0.23	0.21	0.31	0.22
RS2013	0.33	0.30	0.34	0.30	0.18	0.23	0.32	0.25	0.29	0.23	0.31	0.24	0.24	0.36	0.22
RS810	0.31	0.29	0.31	0.29	0.23	0.21	0.29	0.29	0.26	0.23	0.32	0.22	0.26	0.34	0.22
RST9	0.24	0.29	0.34	0.27	0.23	0.23	0.27	0.24	0.25	0.20	0.27	0.21	0.19	0.31	0.19
HS6	0.33	0.36	0.40	0.30	0.21	0.25	0.32	0.24	0.30	0.25	0.35	0.25	0.28	0.34	0.24

H1226	0.30	0.36	0.36	0.32	0.19	0.21	0.30	0.29	0.31	0.26	0.33	0.27	0.26	0.32	0.18
KC3	0.33	0.35	0.34	0.29	0.23	0.30	0.28	0.26	0.30	0.23	0.29	0.25	0.25	0.33	0.19
DEVIRAJ	0.26	0.32	0.34	0.25	0.16	0.20	0.21	0.22	0.26	0.21	0.28	0.22	0.24	0.26	0.13
KANCHANA	0.34	0.36	0.45	0.42	0.43	0.41	0.35	0.37	0.37	0.32	0.43	0.35	0.34	0.39	0.38
PKVRAJAT	0.34	0.33	0.35	0.28	0.24	0.26	0.30	0.35	0.25	0.25	0.31	0.27	0.29	0.26	0.30
RMPBS155	0.24	0.25	0.31	0.26	0.17	0.21	0.21	0.24	0.18	0.16	0.31	0.19	0.17	0.25	0.22
SUVIN	0.37	0.38	0.44	0.41	0.36	0.36	0.37	0.38	0.40	0.38	0.43	0.36	0.39	0.36	0.35
SUJATA	0.32	0.38	0.42	0.39	0.33	0.33	0.34	0.36	0.35	0.34	0.41	0.30	0.39	0.40	0.34

Supplementary Table S1. Genetic dissimilarity among cotton varieties of India as revealed by SSR markers

Jaccard's dissimilarity index	MCU 12	DHY 286	BN1	AKH 8828	AKH 081	KHAN DWA2	NH 545	NH 615	JK4	GC OT	GCOT 16	GCOT 20	GCOT 18	GC 12	GAGE TI
CNHO12															
ANJALI															
PRATIMA															
SURAJ															
SURABHI															
SUPRIYA															
SUMANGALA															
SAHANA															
ABHADITA															
AROGYA															
NARASIMHA															
MCU5															
MCU5VT															
MCU10															
MCU12															
DHY286	0.30														
BN1	0.26	0.28													
AKH8828	0.14	0.30	0.22												
AKH081	0.25	0.28	0.25	0.23											
KHANDWA2	0.25	0.23	0.21	0.24	0.15										
NH545	0.21	0.30	0.19	0.19	0.13	0.17									
NH615	0.25	0.28	0.20	0.22	0.19	0.15	0.18								
JK4	0.24	0.21	0.27	0.25	0.21	0.20	0.24	0.17							
GCOT10	0.24	0.34	0.23	0.29	0.23	0.16	0.23	0.19	0.22						
GCOT16	0.19	0.29	0.26	0.23	0.19	0.22	0.20	0.17	0.18	0.19					
GCOT20	0.28	0.31	0.31	0.25	0.31	0.24	0.29	0.24	0.24	0.26	0.26				
GCOT18	0.25	0.32	0.29	0.26	0.27	0.22	0.29	0.19	0.22	0.29	0.24	0.22			
GCOT12	0.27	0.30	0.25	0.21	0.21	0.21	0.17	0.14	0.18	0.23	0.20	0.27	0.22		
GAGETI	0.31	0.31	0.35	0.32	0.28	0.29	0.28	0.23	0.25	0.33	0.28	0.32	0.29	0.24	

JLH168	0.20	0.32	0.29	0.19	0.20	0.20	0.18	0.24	0.24	0.22	0.18	0.18	0.23	0.20	0.28
F1054	0.22	0.36	0.30	0.23	0.31	0.29	0.29	0.23	0.25	0.27	0.22	0.27	0.27	0.26	0.28
F1378	0.32	0.30	0.31	0.30	0.16	0.22	0.23	0.29	0.25	0.26	0.29	0.29	0.33	0.28	0.32
F846	0.23	0.26	0.29	0.20	0.22	0.24	0.23	0.18	0.20	0.26	0.18	0.24	0.25	0.25	0.25
F1861	0.30	0.33	0.38	0.31	0.27	0.35	0.28	0.29	0.29	0.34	0.27	0.38	0.32	0.29	0.20
RS875	0.23	0.35	0.27	0.24	0.25	0.25	0.25	0.20	0.20	0.17	0.14	0.28	0.26	0.21	0.27
RS2013	0.30	0.33	0.27	0.27	0.23	0.19	0.27	0.24	0.26	0.21	0.18	0.30	0.28	0.28	0.32
RS810	0.25	0.32	0.26	0.26	0.22	0.21	0.24	0.19	0.25	0.24	0.22	0.27	0.20	0.24	0.29
RST9	0.22	0.32	0.28	0.25	0.24	0.22	0.22	0.21	0.22	0.17	0.18	0.30	0.29	0.22	0.26
HS6	0.25	0.33	0.32	0.29	0.29	0.25	0.29	0.21	0.27	0.23	0.23	0.31	0.27	0.29	0.29
H1226	0.28	0.35	0.30	0.20	0.26	0.23	0.25	0.24	0.29	0.25	0.23	0.30	0.28	0.25	0.27
KC3	0.23	0.34	0.26	0.22	0.24	0.22	0.25	0.22	0.28	0.28	0.20	0.33	0.27	0.27	0.29
DEVIRAJ	0.18	0.30	0.22	0.16	0.22	0.17	0.22	0.18	0.27	0.23	0.18	0.28	0.27	0.25	0.27
KANCHANA	0.37	0.39	0.41	0.41	0.41	0.39	0.36	0.33	0.33	0.33	0.31	0.39	0.38	0.33	0.38
PKVRAJAT	0.31	0.35	0.33	0.33	0.19	0.20	0.28	0.25	0.23	0.20	0.25	0.33	0.27	0.31	0.26
RMPBS155	0.22	0.29	0.24	0.22	0.18	0.16	0.19	0.15	0.17	0.14	0.10	0.27	0.24	0.16	0.24
SUVIN	0.35	0.35	0.40	0.39	0.33	0.35	0.40	0.36	0.34	0.33	0.33	0.37	0.37	0.39	0.32
SUJATA	0.32	0.39	0.37	0.34	0.32	0.32	0.35	0.35	0.37	0.32	0.32	0.38	0.38	0.35	0.37

Supplementary Table S1. Genetic dissimilarity among cotton varieties of India as revealed by SSR markers

Jaccard's dissimilarity index	JLH 168	F 1054	F 1378	F 846	F 1861	RS 875	RS 2013	RS 810	RST 9	HS 6	H 1226	KC 3	Devi-raj	Kanch-ana	PKV-Rajat	RMPBS 155	SUVIN
CNHO12																	
ANJALI																	
PRATIMA																	
SURAJ																	
SURABHI																	
SUPRIYA																	
SUMANGALA																	
SAHANA																	
ABHADITA																	
AROGYA																	
NARASIMHA																	
MCU5																	
MCU5VT																	
MCU10																	
MCU12																	
DHY286																	
BN1																	
AKH8828																	
AKH081																	
KHANDWA2																	
NH545																	
NH615																	
JK4																	
GCOT10																	
GCOT16																	
GCOT20																	
GCOT18																	
GCOT12																	
GAGETI																	
JLH168																	
F1054	0.29																
F1378	0.20	0.29															
F846	0.24	0.21	0.31														
F1861	0.29	0.20	0.26	0.26													
RS875	0.19	0.23	0.28	0.22	0.28												
RS2013	0.26	0.28	0.29	0.24	0.35	0.19											
RS810	0.27	0.24	0.26	0.23	0.30	0.21	0.19										
RST9	0.25	0.20	0.26	0.25	0.28	0.18	0.25	0.20									
HS6	0.28	0.22	0.31	0.24	0.28	0.23	0.19	0.18	0.18								
H1226	0.23	0.27	0.30	0.26	0.26	0.26	0.22	0.25	0.21	0.19							
KC3	0.26	0.28	0.32	0.26	0.31	0.23	0.20	0.20	0.20	0.22	0.18						
DEVIRAJ	0.24	0.21	0.32	0.20	0.29	0.24	0.20	0.23	0.21	0.19	0.16	0.11					
KANCHANA	0.38	0.41	0.46	0.37	0.39	0.33	0.42	0.41	0.35	0.39	0.42	0.42	0.40				
PKVRAJAT	0.27	0.35	0.25	0.25	0.32	0.25	0.25	0.24	0.27	0.26	0.28	0.29	0.25	0.41			
RMPBS155	0.21	0.23	0.30	0.19	0.29	0.14	0.15	0.21	0.18	0.18	0.18	0.20	0.16	0.34	0.17		
SUVIN	0.30	0.40	0.35	0.34	0.38	0.34	0.34	0.37	0.37	0.35	0.38	0.40	0.33	0.45	0.29	0.33	
SUJATA	0.32	0.37	0.36	0.36	0.41	0.31	0.33	0.33	0.33	0.35	0.39	0.38	0.31	0.36	0.31	0.28	0.23