J. Cotton Res. Dev. **32** (1) 30-37 (January, 2018) **ISSN No. 0972-8619**

Elucidation of genetic divergence in upland cotton employing simple sequence repeat markers

PANKAJ SHARMA*, R.S.SOHU AND DHARMINDER PATHAK

Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana 141004

**E-mail: pankaj01972@gmail.com*

Abstract : Assessment of genetic diversity of crop plants is considered as an important primary prerequisite for the successful genetic improvement programme. In this study, the genetic diversity among 30 genotypes of American cotton using 26 cotton specific simple sequence repeats (SSR) molecular markers. All accessions were grouped in three major clusters, cluster 1 consisted of sixteen genotypes, cluster 2 contain ten genotypes, and cluster 3 consists of four genotypes respectively. For an individual primer, the alleles amplified varied from two to six with an average of 4.03. PIC values ranged from 0.41 (NAU 2278) through 0.76 (NAU 3401 and CGR 5282). Genotypes LH 2376 and BS 37 were found to be most distant, which may be used as parental lines in heterosis breeding.

Key words*:* Genetic diversity, *G. hirsutum*, molecular markers

Gossypium genus consists of 50 species, cultivated from arid to semi-arid regions of the world (Rahman *et al*, 2012). Cotton is an often cross pollinated crop and consisted of four cultivated species including two from Africa-Asia (*Gossypium arboreum* and *Gossypium herbaceum*) and two from Americas (*Gossypium hirsutum* and *Gossypium barbadense*) (Wendel and Cronn 2003). Among the four cultivated species American cotton (*Gossypium hirsutum* L.) is major cultivated crop throughout worldwide. It is a tetraploid (2n= 52) species grown as a cash crop by Indian farmers. India is the largest producer of cotton followed by China. Higher productivity is the major objective of any crop breeding programme, but to achieve this there is need of identification and characterization of diverse strains which could be used as a donor parent to develop high yielding cultivars through hybridization (Acquaah, 2012). Different genetic approaches have been used to estimate the genetic diversity in germplasm accessions, breeding lines and segregating populations but most of these approaches were morpho-physiological based but influenced by environment (Badigannavar *et al.,* 2012). Now a days DNA markers are used as indispensible tools for characterising the genetic resources by detecting variation from DNA sequences and helps to solve the problem of environment effect. Simple sequence repeat markers (SSRs) are multi-allele genetic markers and are being extensively used in plant breeding programme especially for the assessment of molecular diversity as they are co-dominant in nature, very informative, experimentally reproducible and transferable among related species (Vieira *et al.*, 2016). Availability of SSR markers in cotton genome has been increased with the recent advancement in genome sequences technologies (Li *et al.*, 2015). The information about genetic diversity is an important part for

the gene pool selection and used for development of economically better cultivars. In this study, we employed SSR markers to reveal the information of genetic divergence among American cotton accessions.

Plant material : Thirty *G. hirsutum* genotypes, were used in this study from a group of 81 genotypes already studied on the basis of diversity studies using Mahalanobis D^2 analysis (Sharma *et al,* 2016). The seeds of 30 different accessions of *G. hirsutum* has different origin (Table 1). These genotypes were grown at experimental area of Cotton Section, Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana during *kharif,* 2012. The seeds of all 30 genotypes were planted in a partially balanced lattice design with two replications.

Genomic DNA extraction : Fresh leaves (2-3 days old) were used from plants for genomic DNA extraction. Genomic DNA was isolated from each genotype by using cetyl trimethyl ammonium bromide (CTAB) method (Saghai-Maroof *et al.*, 1984). Quality and quantity of DNA was checked through calculating the absorbance ratio at 260/280 by using Nanodrop 1000 version Spectrophotometer in the Department of Vegetable Science, PAU Ludihana.

Microsatellite markers and PCR amplification : In the present study, 26 cotton specific SSR primer pairs (Table 2) belonging to NAU and CGR series were employed for molecular analysis. These SSR primers were synthesised through Integrated DNA Technologies (IDT). PCR amplifications were carried out in a 20 il reaction mix consisting 2 il DNA at a concentration of 25 ng/il DNA template, 2 il of 10× PCR buffer, 4il of 1 mM dNTPs, 1.2 il of 25 mM $MgCl_2$, 2il of 5iM of each forward and reverse primer, 0.2 il Taq DNA polymerase and final volume (20 il) was made by adding deionized double distilled water. Amplification reactions were carried out as initial denaturation: 4 min at 94°C, 1 min at 94°C for denaturing, 1 min for annealing at 49-57°C and 1 min for DNA extension at 72 °C. Then a final extension at 72 °C for 10 min and hold at 4°C until the tubes were removed (Russel and Sambrook, 2000).

Resolution and visualisation of PCR products : The amplified fragments were separated electrophoretically using a denaturing gel consisting of 6 per cent polyacrylamide (19:1, acrylamide: bisacrylamide) and 7.5ml of TBE (10X). The PAGE gel was stained by using 50 il ethidium bromide and pre run (without loading PCR product) for 3 h for uniform staining. 10iL of each PCR product was loaded onto gel and was resolved by running the gel at 300V for 4-6 h. Finally gels were visualised under ultraviolet light and photographed using photo documentation system (Alphaimager HP, Alpha Innotech).

Data analysis : Total number of alleles was recorded for each microsatellite marker in 30 *Gossypium hirsutum* genotypes by assigning allele number as 1, 2, 3, 4 and so on. The amplified alleles were recorded as 1 (band present) or 0 (band absent) to construct binary matrix used for diversity analysis by using DARwin Computer software programme (Perrier

and Jacquemond Collet, 2006). Genetic distances between genotypes were calculated by using a dissimilarity matrix. Dissimilarly matrix for SSR primers was constructed using Dice coefficient of associations to find out genetic relationships. The data were subjected to unweighted pair groups method with arithmetic mean (UPGMA) analysis to generate dendrogram. The genetic divergence of each microsatellite locus was obtained by calculating the frequency of the microsatellite allele based on polymorphism information content (PIC) using the equation:

$$PIC = 1- \sum_{i=1}^{n}(P_{ij})^2 - \{\sum_{i=1}^{n}(P_{ij})^2\}^2 + \sum_{i=1}^{n}\{(P_{ij})^2\}^2$$

Where P_{ij} is the frequency of j^{th} allele in i^{th} primer and summation extends over 'n' patterns.

Allelic diversity generated by SSR markers : For the 30 cotton genotypes, 26 SSR primer pairs were used for genetic diversity analysis on the basis of scorable amplified bands. The number of bands amplified by each of the 26 microsatellite primers ranged from two to six on 6 per cent PAGE gel. A total of 105 alleles with an average of 4.03 alleles/locus were amplified in 30 genotypes. Maximum of six alleles were observed for NAU 2000 and NAU3401, five alleles were amplified for six primer pairs (NAU 1067, NAU 5166, CGR 5128, CGR 5282, CGR 5548 and CGR 5653) and three alleles were observed for seven primer pairs (NAU 3684, NAU 5152, NAU 5345, CGR 5152, CGR 5446, CGR 5747 and CGR 6550). Majority of the primers (10) amplified four alleles and one primer pair (NAU 2278) amplified two alleles (Table 2).

Polymorphic information content (PIC) values provide an estimate of the discriminating power of a marker by taking into account not only the number of alleles at a locus but also relative frequencies of these alleles. These values depend upon the genetic diversity among the genotypes being studied. PIC values obtained in the present study varied from 0.41 (NAU 2278) through 0.76 (NAU 3401, CGR 5282) with an average PIC value for 26 primers being 0.64. Twelve primers *viz.,* NAU 1067, NAU 2701, NAU

Table 1. List of plant material used for molecular diversity analysis

S .No	Genotype	Origin of genotype
1	LH 2379	Ludhiana
2	LH 2376	Ludhiana
3	LH 2368	Ludhiana
4	LH 2355	Ludhiana
5	LH 2332	Ludhiana
6	LH 2306	Ludhiana
7	LH 2255	Ludhiana
8	LH 900	Ludhiana
9	F 2228	Faridkot
10	F 2170	Faridkot
11	F 2164	Faridkot
12	F 846	Faridkot
13	F 505	Faridkot
14	CSH 3158	Sirsa
15	CSH 3129	Sirsa
16	CSH 3114	Sirsa
17	CSH 10	Sirsa
18	HS 286	Sirsa
19	HS 283	Sirsa
20	H 1246	Hisar
21	LRA 5166	Coimbatore
22	LRK 516	Coimbatore
23	TCH 1728	Coimbatore
24	MR 786	M R Seeds
25	MR 68	M R Seeds
26	GJHV 500	Junagadh
27	KH 1001	Khandwa
28	BS 37	Bhawanipatnam
29	P 2150	New Delhi
30	RS 2013	Sriganganagar

Table 2. Allele amplification and PIC values for the SSR primers screened using 30 *Gossypium hirsutum* accessions

S.No.	Primer	Number of alleles amplified	PIC value	S.No.	Primer	Number of alleles amplified	PIC value
1	NAU 1067	5	0.75	14	CGR 6550	3	0.67
2	NAU 3684	3	0.47	15	CGR 5282	5	0.76
3	NAU 1081	4	0.62	16	CGR 5128	5	0.70
4	NAU 2701	4	0.71	17	CGR 5729	4	0.62
5	NAU 2000	6	0.71	18	CGR 5446	3	0.60
6	NAU 2278	2	0.41	19	CGR 5022	4	0.65
7	NAU 5152	3	0.59	20	CGR 5548	5	0.60
8	NAU 3324	4	0.75	21	CGR 5747	3	0.59
9	NAU 5318	4	0.72	22	CGR 5888	4	0.65
10	NAU 5166	5	0.75	23	CGR 5862	4	0.70
11	NAU 5428	4	0.72	24	CGR 5653	5	0.73
12	NAU 3401	6	0.76	25	CGR 6864	4	0.43
13	NAU 5345	3	0.58	26	CGR 5152	3	0.49
					Total	**105**	**-**
					Average	**4.03**	**0.64**

2000, NAU 3324, NAU 5318, NAU 5166, NAU 5428, NAU 3401, CGR 5282, CGR 5128, CGR 5862 and CGR 5653 recorded PIC value of at least 0.70. Primer NAU 2278 registered minimum PIC value of 0.41 in the present investigation. In *G. hirsutum,* a range of PIC values have been reported for SSR markers as 0.0 to 0.57 (Rungis *et al.*, 2005), 0.28 to 0.91 (Wei *et al.,* 2010) and 0.00 to 0.77 (Thiyagu *et al.,* 2010).

Cluster analysis : The dendrogram showing genetic relationships among 30 genotypes based on 26 microsatellite markers is presented in Fig. 2. The UPGMA cluster analysis showed that all the 30 cotton genotypes were clustered into three main groups. First cluster is the largest one comprising of 16 genotypes; second cluster consisted of 10 genotypes; and third cluster which is the smallest one, included four genotypes. Finally, the 30 genotypes were divided into six sub-groups at similarity coefficient of 0.1. The largest sub-group (Cluster 1) comprised of 10 genotypes and the smallest sub groups (cluster 5 and 6)

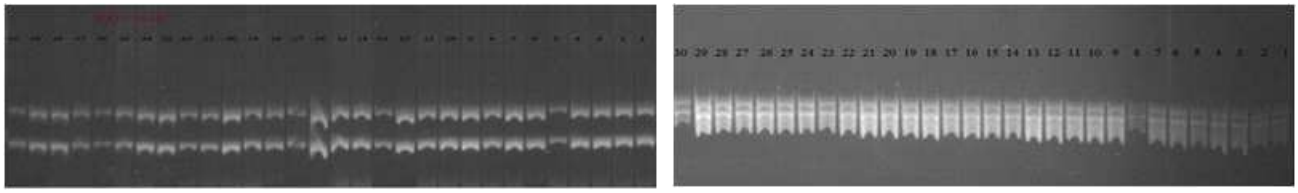

Fig. 1. PAGE Plates: Amplification of genomic DNA of 30 test American cotton genotypes with SSR primers NAU 5428 (Left) and NAU 1081 (Right) where Genotypes marked by order (1-30):LH 2306, LH 2332, LH 2255, LH 2355, LH 2376, LH 2368, LH 2379, LRK 516, LRA 5166, F 505, H 1246, F 846, LH 900, MR 786, CSH 10, F 2228, CSH 3158, HS 286, MR 68, CSH 3129, RS 2013, F 2170, F 2164, HS 283, CSH 3114, GJHV 500, P 2150, BS 37, KH 1001, TCH 1728

Table 3. Average estimates of genetic dissimilarity between 30 cotton genotypes using 26 SSR primers.

	LH 2306	LH 2332	LH 2255	LH 2355	LH 2376	LH 2368	LH 2379	LRK 516	LRA 5166	F 505	H 1246	F 846	LH 900	MR 786	CSH 10	F 2228	CSH 3158	HS 286	MR 68	CSH 3129	RS 2013	F 2170	F 2164	HS 283	CSH 3114	GJHV 500	P 2150	BS 37	KH 1001	TCH 1728
LH2306	0.00																													
LH2332	0.43	0.00																												
LH2255	0.47	0.37	0.00																											
LH2355	0.34	0.44	0.38	0.00																										
LH2376	0.45	0.30	0.41	0.43	0.00																									
LH2368	0.48	0.31	0.28	0.38	0.24	0.00																								
LH2379	0.26	0.42	0.38	0.20	0.46	0.39	0.00																							
LRK516	0.32	0.46	0.38	0.22	0.45	0.42	0.14	0.00																						
LRA5166	0.31	0.42	0.35	0.29	0.41	0.43	0.15	0.19	0.00																					
F505	0.31	0.49	0.41	0.25	0.53	0.45	0.13	0.17	0.16	0.00																				
H1246	0.37	0.52	0.44	0.29	0.49	0.43	0.16	0.22	0.18	0.17	0.00																			
F846	0.27	0.44	0.35	0.29	0.52	0.38	0.19	0.21	0.18	0.18	0.25	0.00																		
LH900	0.35	0.46	0.41	0.30	0.43	0.33	0.19	0.19	0.23	0.17	0.18	0.23	0.00																	
MR786	0.55	0.43	0.29	0.39	0.51	0.33	0.31	0.27	0.35	0.33	0.38	0.33	0.35	0.00																
CSH10	0.25	0.46	0.33	0.29	0.43	0.45	0.19	0.17	0.16	0.18	0.25	0.17	0.20	0.35	0.00															
F2228	0.35	0.49	0.35	0.29	0.51	0.46	0.20	0.14	0.22	0.18	0.27	0.23	0.20	0.26	0.17	0.00														
CSH3158	0.29	0.46	0.31	0.25	0.43	0.40	0.16	0.09	0.12	0.18	0.23	0.19	0.21	0.28	0.13	0.15	0.00													
HS286	0.29	0.46	0.37	0.29	0.48	0.40	0.19	0.17	0.20	0.18	0.19	0.19	0.21	0.35	0.13	0.15	0.15	0.00												
MR68	0.37	0.51	0.35	0.23	0.53	0.43	0.15	0.13	0.20	0.12	0.22	0.22	0.21	0.26	0.16	0.13	0.12	0.14	0.00											
CSH3129	0.43	0.38	0.38	0.46	0.45	0.27	0.40	0.33	0.42	0.42	0.47	0.31	0.32	0.27	0.31	0.30	0.33	0.29	0.36	0.00										
RS2013	0.47	0.48	0.52	0.50	0.44	0.51	0.37	0.38	0.46	0.41	0.37	0.41	0.41	0.39	0.39	0.40	0.45	0.41	0.44	0.37	0.00									
F2170	0.42	0.46	0.38	0.38	0.51	0.42	0.29	0.29	0.40	0.36	0.35	0.38	0.35	0.33	0.28	0.31	0.36	0.34	0.34	0.27	0.22	0.00								
F2164	0.49	0.47	0.34	0.36	0.55	0.38	0.25	0.27	0.35	0.26	0.33	0.31	0.33	0.16	0.33	0.23	0.28	0.33	0.22	0.25	0.31	0.26	0.00							
HS283	0.42	0.51	0.41	0.35	0.55	0.41	0.19	0.19	0.29	0.24	0.26	0.26	0.25	0.26	0.24	0.18	0.20	0.22	0.20	0.31	0.36	0.28	0.20	0.00						
CSH3114	0.39	0.47	0.35	0.37	0.44	0.41	0.31	0.30	0.23	0.32	0.33	0.28	0.27	0.42	0.23	0.29	0.23	0.28	0.30	0.38	0.42	0.35	0.40	0.25	0.00					
GJHV500	0.48	0.37	0.39	0.53	0.36	0.44	0.53	0.48	0.42	0.56	0.49	0.48	0.46	0.51	0.42	0.49	0.44	0.48	0.51	0.41	0.45	0.44	0.47	0.56	0.36	0.00				
P2150	0.38	0.51	0.41	0.29	0.55	0.49	0.21	0.21	0.27	0.24	0.32	0.28	0.25	0.35	0.24	0.22	0.24	0.24	0.20	0.40	0.41	0.32	0.28	0.22	0.25	0.49	0.00			
BS37	0.43	0.55	0.49	0.40	0.56	0.46	0.28	0.26	0.35	0.31	0.33	0.35	0.23	0.43	0.31	0.26	0.31	0.31	0.29	0.37	0.45	0.35	0.32	0.23	0.31	0.48	0.19	0.00		
KH1001	0.39	0.45	0.35	0.39	0.42	0.46	0.33	0.27	0.26	0.32	0.37	0.26	0.27	0.40	0.25	0.23	0.21	0.26	0.30	0.38	0.47	0.45	0.38	0.26	0.20	0.40	0.20	0.28	0.00	
TCH1728	0.41	0.49	0.40	0.33	0.53	0.41	0.25	0.24	0.34	0.24	0.31	0.28	0.27	0.38	0.32	0.29	0.28	0.30	0.28	0.43	0.42	0.35	0.31	0.24	0.27	0.49	0.18	0.22	0.27	0.00

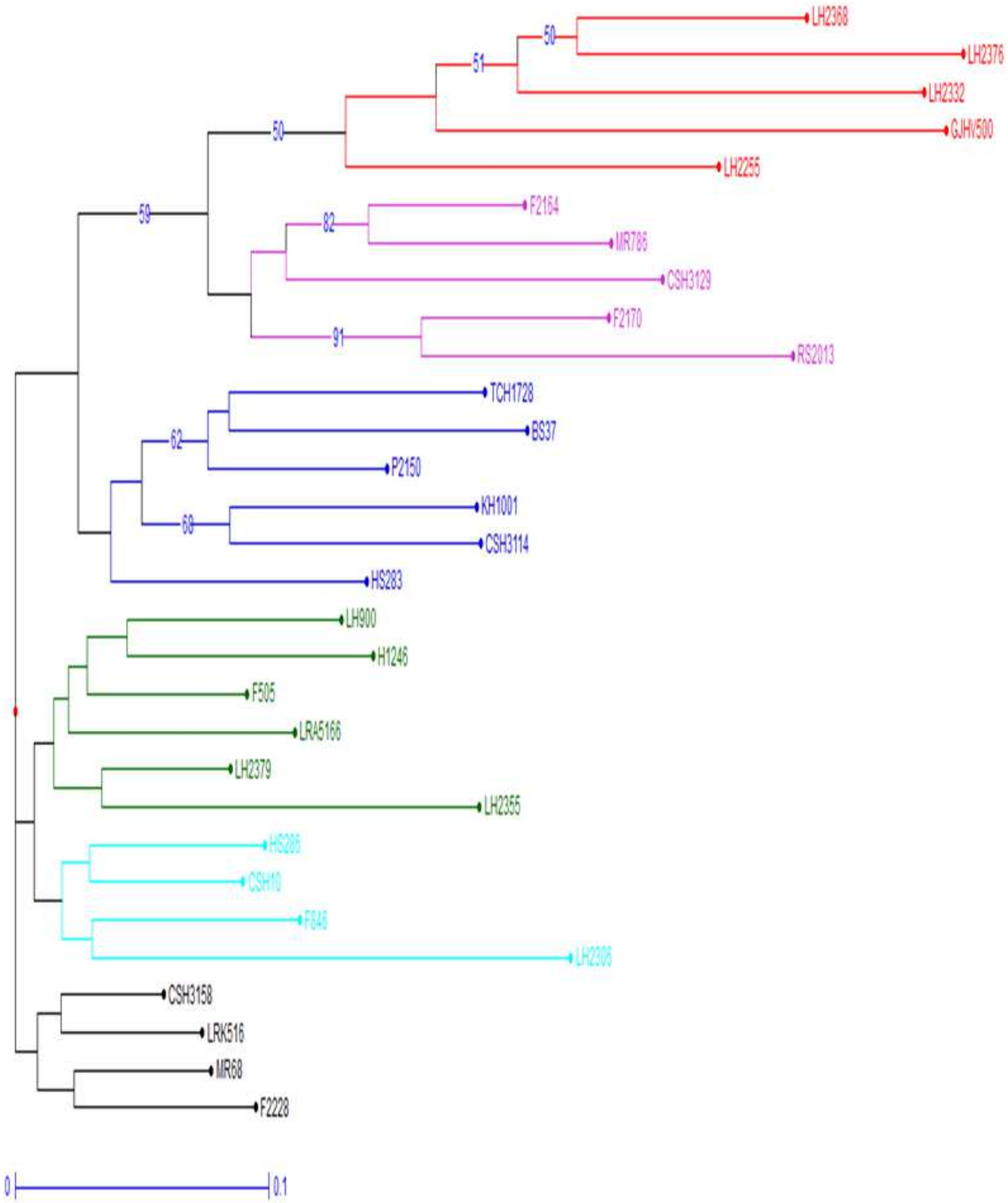

Fig. 2. Dendrogram obtained by SSR marker analysis using DARwin 5.0

comprised of two genotypes each.

The clustering of genotypes into different groups did not follow any specific pattern. For example, genotypes originating from Ludhiana, Faridkot, Sriganganagar, Junagadh, Sirsa were included in cluster 1. Of the six genotypes in cluster 3, three namely LH 900, LH 2355 and LH 2379 were developed at Ludhiana; LRA 5166 at Coimbatore; H 1246 at Hisar and F 505 was developed at Faridkot. Similarly, based on SSR analysis, Guo *et al.,* (2006) observed no relationship between genetic diversity and geographical distribution in a set of 109 accessions of *G. arboreum*. According to Wang *et al.,* (2007) clustering patterns could not be correlated to the geographic origin, pedigree and common parentage of the American cotton cultivars.

In the present investigation, the dissimilarity coefficients ranged from 0.09 to 0.56 signifying dissimilarity among the *G. hirsutum* genotypes evaluated (Table 3). Genotypic pairs having utmost genetic dissimilarity of 56% were LH 2376 and BS 37, while the minimum genetic dissimilarity of (9%) was observed between the lines LRK 516 and CSH 3158. Furthermore, genotypes LH 2376, BS 37, LH 2306, MR 786 and F 2164 recorded high values of dissimilarity coefficients. Likewise, dissimilarity coefficients in American cotton have been reported to vary from 0.0 to 0.41 by Bertini *et al.,* (2006).

CONCLUSIONS

The overall outcomes of the investigation suggest that SSR are effective for assessing genetic diversity, relatedness. The information, we gained in the present study provides us a useful guidance for cotton breeding programmes. The cultivars used in this study can be used as parents for broadening the genetic base and also can be used for the development of new high yielding varieties to increase the productivity and quality of cotton. Genotypes LH 2376 and BS 37 which were observed to be quite distant (56%) in the present investigation, may serve as the parental lines for the heterosis breeding.

REFERENCES

Acquaah, G. 2012. *Principles of Plant Genetics and Breeding*. John Wiley and Sons, Ltd, UK. Pp. 227-27.

Badigannavar, A., Myers, G. and Jones, D. 2012. Molecular Diversity Revealed by AFLP Markers in upland cotton genotypes. *J Crop Improv* **26**: 627–40.

Bertini, C., Schuster, I., Sediyama, T., Barros, G. E. and Moreira, M. A. 2006. Characterization and genetic diversity analysis of cotton cultivars using microsatellites.*Genet Mol Biol* **24** : 321-29

Guo, W. Z., Zhou, B. L., Yang, L. M., Wang, W. and Zhang, T. Z. 2006. Genetic diversity of landraces in *G. arboreum* L. race *sinense* assessed with simple sequence repeat markers. *J Integ Plant Biol* **48**:1008-17.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R., Ma, Z., Shang, H., Ma, X., Wu, J. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol* **33**: 524-30.

Perrier, X. and Jacquemond-Collet, J. P. 2006. DARwin software. http://darwin.cirad.fr/darwin/

Rahman, M. U., Shaheen, T., Tabbasam, N., Iqbal, M. A., Ashraf, M., Zafar, Y. and Paterson, A. H. 2012. Cotton genetic resources. *Agron Sustain Dev* **32**: 419-32.

Rungis, D., Llewellyn, D., Dennis, E. S. and Lyon, B. R. 2005. Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Australian J Agric Res* **56**:301-07.

Russel, D. W. and Sambrook, J. 2000. *Molecular cloning a laboratory manual*. Vol 3 Cold Spring Harbour Laboratory Press, New York.

Saghai, Maroof M. A., Soliman, K. M., Jorgensen, R. A. and Allard, R. W. 1984. Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proc Natl Acad Sci USA* **81**:8014-19.

Sharma, P., Sohu, R. S. and Pathak, D. 2016. Assessment of Genetic Diversity in *Gossypium hirsutum* L. Germplasm using morphological characters following Mahalanobis D^2 Analysis. *J. Cotton Res. Dev* **30** : 1-5.

Thiyagu, K., Boopathi, N. M., Nadarajan, N., Gopikrishnan, A., Selvakumar, P., Magadum, S. and Ravikesava, R. 2010. Sampling and exploitation of genetic variation exist in locally adapted accessions using phenotypic and molecular markers for genetic improvement of cotton. *Gene Conserve* **10**:129-53.

Wang, X., Ma, J., Yang, S., Zhang, G. and Ma, Z. 2007. Assessment of genetic diversity among Chinese upland cottons with *Fusarium* and/ or *Verticillium* wilts resistance by AFLP and SSR markers. *Front Agric China* **1**:129-35.

Wendel, J. F. and Cronn, R. C. 2003. Polyploidy and the evolutionary history of cotton. *Adv Agron* **78**:139-86.

Wei, O., Liu, F., Li, S. H.,Wang, C. Y., Zhang, X. D., Wang, Y. H. and Wang, K. B. 2010. Analysis of genetic diversity of allotetraploid cotton based on SSR. *Acta Agron Sinica* **36** : 1902-09.

Vieira, M. L. C., Santini, L., Diniz, A. L. and Munhoz, C. F. 2016. Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol* **39** : 312-28

Received for publication : June 22, 2017
Accepted for publication : November 6, 2017