

**ASSESSING GENETIC DIVERSITY AMONG GOSSYPIUM ARBOREUM
L. GENOTYPES USING ISSR MARKERS****KHUSHBOO SETHI¹, PRIYANKA SIWACH^{1*}, SURENDER KUMAR VERMA²
AND MEGHA SIHAG¹**¹*Department of Biotechnology, Chaudhary Devi Lal University, Sirsa, Haryana, India*²*Central Institute of Cotton Research, Regional Station, Sirsa, Haryana, India.***ABSTRACT**

The diploid cultivated cotton (*Gossypium arboreum*) is an invaluable genetic resource for the improvement of primary cotton. First step for this is a systematic race wise characterization using molecular markers. Present study attempted to characterize 95 genotypes with ISSR markers. The 100 selected markers produced a total of 397 bands, of which 368 were polymorphic. The PIC values ranged from 0.23 to 0.5 (average 0.40). Resolving power and marker index were calculated to know the efficiency of markers. The mean number of effective alleles was 1.409 and the mean value of shannon's information index was 0.368. AMOVA test showed that 17% of molecular variance were due to 6 population groups and rest due to variance in each group. The six populations were distributed in two major clusters by UPGMA while PCoA analysis could not differentiate population groups accurately. Outcomes of this research will help in identification of right genotypes for any given breeding program.

KEYWORDS: Diploid Cotton, genetic diversity, *Gossypium arboreum*, ISSR, PCoA

*Corresponding author

PRIYANKA SIWACHDepartment of Biotechnology, Chaudhary Devi Lal University,
Sirsa, Haryana, India

INTRODUCTION

Cotton, 'the white gold', is the leading natural fiber crop and the corner stone of textile industries worldwide. It is also one of the principal crops of India, playing a vital role in the country's economic growth by providing substantial employment and making significant contributions to export earnings. It belongs to genus *Gossypium* (family Malvaceae) which contains more than 45 diploid species and 5 allotetraploid species. Out of which only four species are cultivated viz. *Gossypium arboreum*, *Gossypium herbaceum* (both are diploid, Old world cotton) and *Gossypium hirsutum* and *Gossypium barbadense* (both are tetraploid, New world cotton) as only these produce spinnable fibers¹⁻³. At present, tetraploid cotton (dominantly *G. hirsutum*) occupies a major fraction (>90%) of world cotton cultivation because of superior fiber quality and has achieved the status of primary cotton. But the primary cotton breeding gene pool presently has a narrow genetic base due to continuous breeding and selection activities. Latter has been found a major cause for genetic bottlenecks and loss of genetic diversity in crop plants including cotton⁴. There is a need to enrich the breeding gene pool of primary cotton with genetic diversity. *G. arboreum* is an important germplasm resource in cotton breeding programs due to its desirable traits for improvement of tetraploid cotton. *G. arboreum* has certain inherent qualities, which the tetraploids lack, like the ability to withstand drought and salinity^{5, 6} and is tolerant to several pests and disease⁷⁻¹¹. A huge collection of *G. arboreum* germplasm is available which is maintained at different centers worldwide¹². The exploitation of this secondary germplasm pool for enriching the genetic diversity for primary cotton pool will certainly produce far reaching effects. Domestication of *G. arboreum* initiated at Indus valley¹³ and from there, it spread to various locations all around resulting in evolution of different races viz. 'indicum', 'burmanicum', 'sinense', 'soudanense', 'bengalense', 'cernuum'^{14,15}, each race having its own characteristics features¹². The knowledge of genetic variation between and

within populations of *G. arboreum* is desirable for exploiting them in modern cotton improvement programs. DNA markers have emerged as efficient tools for molecular characterization of plant species. Polymerase chain reaction (PCR) based molecular markers, e.g. ISSR (Inter Simple Sequence Repeat), RAPD (Random amplified Polymorphic DNA), SSR (Simple Sequence Repeat), AFLP (Amplified Fragment Length Polymorphism) etc. are useful for various applications in the plant breeding. Among these, Inter Simple Sequence Repeats (ISSRs) are arbitrary markers in which only one primer is used. The ISSR technique involves amplification of a DNA segment present at an amplifiable distance between two identical microsatellite repeat regions oriented in opposite strands. This technique uses microsatellites, usually 16 to 25 bp long, as primers in the single primer PCR reaction targeting multiple genomic loci to amplify mainly the inter SSR sequences of different size¹⁶. The primers used can be repeats of di, tri, tetra or penta nucleotides anchored with one or two base sequences at 3' or 5' end¹⁷. Unanchored primers also can be used¹⁸. ISSR are reproducible markers with 92-95 per cent efficiency¹⁶. ISSR markers have been proved to be better than RAPD and other arbitrary markers systems for genetic diversity studies^{19, 20}. There are only a few studies using ISSR markers for *G. arboreum* germplasm analysis^{19, 21, 22}, though considerable work has been reported for *G. hirsutum*²³⁻²⁷. The present work was designed to study genetic diversity as well as genetic relationship among genotypes belonging to six different races of *G. arboreum* using ISSR markers.

MATERIALS AND METHODS

Plant materials and DNA extraction

Ninety five cotton genotypes, belonging to six races of *G. arboreum*, were selected for the present study (Table I). Among the selected genotypes, 65 genotypes belongs to race 'bengalense' (population group 1), 10

genotypes were from race 'cernuum' (population group 2) and 5 genotypes from each of the race 'indicum' (population group 3), 'soudanense' (population group 4), 'sinense' (population group 5) and 'burmanicum' (population group 6). The seeds were procured by Central Institute of Cotton research (CICR), Regional Station, Sirsa, Haryana, India from CICR, Headquarter Office, Nagpur as well as various CICR regional stations in different locations of India. The cotton plants were cultivated in two rows of 6 m length with 30 cm interplant distance in the experimental field of CICR, Sirsa, in a completely randomized design (CRD) with 3 replications. Single plant, having fresh and young leaves, was selected randomly from any of the three replicates of each genotype. Fresh and young leaves of selected plants were subjected to total genomic DNA extraction using CTAB method²⁸ with certain modifications. Quality and quantity of extracted DNA were examined by running on 0.8% agarose gel as well as by UV-Spectrophotometer method.

ISSR assay

In this study, 200 ISSR primers were used for initial screening. Out of which 100 primers gave informative banding patterns. The selected 100 primers were 15-20-mers which include 66% dinucleotide repeat motif, 24% trinucleotide repeat motif, 6% tetranucleotide repeat motif and 4% pentanucleotide repeat motif (Table II).

They were anchored at 5' end or 3' end by zero nucleotide or by one to three partially degenerated selective nucleotides. PCR amplification was performed in a volume of 20 µl containing 2 µl of DNA (50ng/µl), 0.4 µM of each primer, 200 µM of dNTPs, 0.5 U Taq polymerase and 1X PCR buffer. After a pre-denaturation step of 5 min at 95°C, amplification reactions were cycled forty times at 95°C for 1 min, at annealing temperature (optimized separately for each primer pair, generally Tm-5 °C) for 2 min and polymerization at 72°C for 1 min in a thermocycler (Bio-Rad, USA). The PCR products were visualized by running on 2% agarose gel, followed by staining with ethidium bromide. Finally the gel was photographed under the Gel Documentation system (Bio-Rad, USA). Fragment size was estimated by using a 100 bp molecular size ladder (Fermentas).

Data collection and polymorphism information contents

ISSR bands obtained were treated as binary characters and coded accordingly (presence=1, absence = 0). Primer banding characteristics such as number of scored bands (NSB), number of polymorphic band (NPB), Number of monomorphic bands (NMB) and percentage of polymorphic bands (PPB) were obtained. Polymorphism information content of a band (PIC_i) was calculated as follow²⁹:

$$PIC_i = 1 - \sum_j f_{ij}^2$$

where f_{ij} is the frequency of the j^{th} pattern of the i^{th} band (note that dominant markers have 2 patterns for a band as being present and absent). Then PIC of each primer was calculated as:

$$PIC = 1/n \sum_{i=1}^n PIC_i$$

where n is NPB for that primer.

Informativeness of a band (BI_i) was calculated as:

$$BI_i = 1 - (2 \times |0.5 - p|)$$

where p is the proportion of 95 accessions containing the band. Then, the resolving power (RP) of each primer was calculated as:

$$RP = \sum_{i=1}^n BI_i$$

where n is the NPB for that primer³⁰. Further we calculated mean resolving power for each primer as:

$$MRP = 1/n \sum_i BI_i$$

Marker index (MI) was calculated as product of PIC_i and an effective multiplex ratio (EMR), which is defined as product of the fraction of polymorphic loci and the number of polymorphic loci³¹.

Statistical analysis

Rectangular binary matrix for ISSR was used for statistical analyses. Several genetic diversity parameters were determined in plants with a different population group studied viz. number of ISSR alleles (N), number of different allele (N_a), effective number of allele (N_e), Shannons index (I) and expected heterozygosity (H_e). These were accompanied by Analysis of Molecular Variance (AMOVA) to reveal significant difference between various genotypes and population groups. UPGMA (Unweighted Paired Group using Mean Average) dendrogram of 6 population groups was drawn based on Nei's genetic distance, modified from Neighbor procedure of PHYLIP ver. 3.5. Ordination based on principle coordinate analysis (PCoA) was also done. All computations for determination of genetic parameters, clustering, AMOVA and PCoA analysis was done using softwares- POPGENE ver. 3.2 and GenAlex 6.5. In addition, efforts were made to find relationships between PIC, MRP, RP and MI using scatter matrix plot.

RESULTS AND DISCUSSION

ISSR diversity

Two hundred ISSR primers were used to amplify genomic DNA of 95 genotypes of *G. arboreum*. Of these, only hundred primers were found to suitable for genetic diversity analysis as others yielded poor or no polymorphism. This set of 100 primers yielded a total of 397 reproducible bands, of which 368 (92.6%) were found to be polymorphic. In previous studies on genetic diversity analysis of cotton genotypes

using ISSR marker, 78.3% and 100% of polymorphic bands were obtained^{26, 27}. During the present study, among the primers used, ISSR-18 and ISSR-32 produced the highest number of bands (15 and 12 respectively) while the lowest number of bands (2) were obtained for 15 primers (ISSR-17, 19, 38, 47, 49, 59, 62, 69, 74, 81, 82, 84, 87, 96 and 103). The highest number of polymorphic bands was observed with primer ISSR-18 and ISSR-32 (13 and 10 respectively). The mean number of bands per locus obtained was 3.97 while the number of bands per locus varied from 2 to 15. The size of the bands obtained ranged from a minimum of 200 bp to maximum of 1700 bp. For each ISSR primer NSB, NPB, NMB, PPB, PIC, RP, MRP and MI were calculated (Table II). Polymorphism information content (PIC) value gave the measure of polymorphism describes the usefulness of a genetic marker³². ISSRs are dominant markers which yield PIC values in the range of 0 to 0.5. The PIC values, obtained during the present study, ranged from 0.23 (ISSR-9) to 0.5 (ISSR-69, ISSR-82, ISSR-84) with an average of 0.40. Studies on genetic diversity in tetraploid and diploid genotypes using SSR and ISSR markers, observed mean PIC value=0.4396, (range from 0.0040 to 0.9993)²². The PIC range observed in tetraploid cotton genotypes using ISSR markers was in the range of 0.101- 0.378 with an average of 0.282²⁵. Also, other reports observed PIC range of 0.00-0.5 with an average of 0.321 using ISSR markers in tetraploid cotton³³. Percent polymorphism among ISSR primers ranged from 50% to 100% with an average of 91.35%. ISSR profile generated by ISSR-40 was represented in Fig I. Resolving power (RP)

provided an indication of the ability to distinguish between cultivars. Estimates of RP were found highest for the primers ISSR-18 (21.78), followed by primer ISSR-31 (17.97) and was lowest for primer ISSR-96 (1.04). The maximum mean resolving power (MRP) was found highest for primer ISSR-82 (3.0) and minimum for the primer ISSR-42 (0.39). Marker index (MI) which depicts the discriminating power for a marker was also calculated for each primer and was found to be highest for primer ISSR-18 (4.32) and lowest for primer ISSR-49 (0.18). The maximum and minimum values of the above parameters corresponded with the most efficient and least efficient primers, respectively. Scatter matrix plot of PIC, MRP, RP and MI were drawn to find out the relationships between these factors. But, low value of coefficient of determination (R^2) was observed due to which we were unable to find any correlation between these factors. In contrast to our study, a strong linear relationship MI and RP ($R^2=0.89$) and similarly between MRP and PIC ($R^2=0.97$) was obtained³⁴. Analysis of genetic diversity over all loci showed the mean number of effective alleles (N_e) = 1.409, mean value of Shannon's information index (I) = 0.368 and a mean value of 0.243 expected heterozygosity (H_e). The highest value of N_e (1.806), I (0.593) and H_e (0.421) occurred in population group 2 (race 'cernuum'). The lowest value for the same parameters was observed in population group 5 (race 'sinense') (Table III). AMOVA analysis (after 999 numbers of permutations) was performed among population and within population group. The analysis indicated that 17% of molecular variance were due to 6 population groups and rest 83% was due to genetic variations among accessions in each population group (Table IV). A significant difference among population groups, among individuals of a group and within individuals ($p=0.001$) was observed, depicting the presence of genetic variability among accessions present in each population group. In comparison to our study, previous studies in diploid and tetraploid genotypes using ISSR markers observed 61% of genetic variation among populations selected and 39% due to within population variation²⁷.

This high percentage of genetic variation may be due to the different genomic constitution of diploid and tetraploid species. Also, in another study, the high value of genetic variation (34%) between genotypes and 66% within genotypes was observed in *G. hirsutum* genotypes²⁶. UPGMA dendrogram based on Nei's genetic distance distributed 6 population groups into 2 main clusters. Population group 1 (race 'bengalense'), 3 (race 'indicum'), 4 (race 'soudanense'), 5 (race 'sinense') and 6 (race 'burmanicum') formed the first main cluster whereas single population group 2 (race 'cernuum') formed the second main cluster (Fig II). The genetic distance coefficient among 6 population groups ranged from 0.038-0.163 (Table V). The minimum coefficient (0.038) was observed between race 'soudanense' and race 'bengalense' which means that they were closely related. The maximum coefficient of genetic distance (0.163) was between race 'cernuum' and race 'indicum' which means they are distant. On the basis of evolution of races, race 'indicum' has been found to be a most primitive perennial form in western India and the dispersal of this race to various regions have evolved the other races like 'burmanicum', 'soudanense', 'sinense' and 'bengalense'. Race 'cernuum' have found to evolve independently in the Assam hills of North-East India and Chittagang hills of Bangladesh. It was evident from the dendrogram's first cluster that all the 5 races except race 'cernuum' were found close to each other as they were evolved from a single race 'indicum' whereas the second cluster which includes race 'cernuum', has been found as a separate group. This study is in accordance with the domestication and evolution of *G. arboreum*. So the present study indicates that ISSR, though multi-locus in nature, are good markers systems for studying evolutionary relationship. Likewise, Randomly amplified polymorphic DNA (RAPD) markers are commonly used for phylogenetics study in plants.³⁵ Principal coordinate analysis (PCoA) is a technique which highlights the similarities and differences in the given data by reducing the number of dimensions without much loss of information. PCoA was performed of cotton

genotypes after 999 iterations and a plot was generated (Fig III). The results could not differentiate six population groups of *G. arboreum* accurately. Accessions from population group 2 (race 'cernuum') were well separated from rest of groups due to its independent evolution. Accessions from population groups 1, 3, 4, 5 and 6 were found scattered and overlapped between the 2 coordinates.

CONCLUSION

Plant breeders desire to use *Gossypium arboreum* as an invaluable genetic resource for improving both diploid and tetraploid cotton production. No study till date has been reported for characterization of all six races of *G. arboreum* using ISSR markers due to non availability of race specific markers. The present study has been done to characterize all the six races of *G. arboreum* with the help of ISSR markers which will help in identifying the

genotypes for their further utilization in hybridization program to obtain the highest level of germplasm diversity.

ACKNOWLEDGEMENT

Financial support provided by the University Grants Commission (UGC), New Delhi, India in the form of Major Research Project to Priyanka Siwach (Principal Investigator) and Surender Kumar Verma (Co-Principal Investigator) is gratefully acknowledged. The authors are grateful to the Director, CICR, Sirsa and Director CICR, Nagpur for providing the germplasm. We also acknowledge the infrastructural facilities provided by Chaudhary Devi Lal University, Sirsa, Haryana, India for all the lab work. Acknowledgement is also extended to CICR, Sirsa for allowing the use of the experimental field facility.

Conflict of Interest : None

REFERENCES

1. Fryxell PA, A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea*. 2: 108-165, (1992).
2. Percival AE, Wendel JF, Stewart JM, Taxonomy and germplasm resources. Pp. 33-63 in W.C. Smith and J.T. Cothren, eds., *Cotton: Origin, History, Technology and Production*. John Wiley & Sons, New York, (1999).
3. Ulloa M, Brubaker C, Chee P, Cotton. In: Kole C (ed) *Genome mapping & molecular breeding*. vol 6: technical crops. Springer, New York, pp 1-49, (2007).
4. Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC, Genetic distance among selected cotton genotypes and its relationship with F2 performance. *Crop Sci*, 42: 1841-1847 (2002).
5. Maqbool A, Abbas W, Rao AQ, Irfan M, Zahur M, *Gossypium arboreum* GHSP26 enhances drought tolerance in *Gossypium hirsutum*. *Biotechnol Progr*, 26: 21-25 (2010).
6. Tahir MS, Khan NUI, Sajid Ur R, Development of an interspecific hybrid (Triploid) by crossing *Gossypium hirsutum* and *G. arboreum*. *Cytologia*, 76: 193-199 (2011).
7. Dhawan AK, Simwat GS, Sidhu AS, Field reaction of some varieties of Asiatic Cotton (*Gossypium arboreum* L.) to sucking and boll worm pests. *J Res Punjab Agri Uni*, 28: 57-62 (1991).
8. Nibouche S, Brevault T, Klassou C, Dessauw D, Hau B, Assessment of the resistance of cotton germplasm (*Gossypium* spp.) to aphids (Homoptera, Aphididae) and leafhoppers (Homoptera : Cicadellidae, Typhlocybinae): methodology and genetic variability. *Plant breed*, 127: 376-382 (2008).
9. Wheeler TA, Gannaway JR, Keating K, Identification of resistance to *Thielaviopsis*

- basicola* in diploid cotton. Plant Dis, 83: 831-833 (1999).
10. Mehetre SS, Shinde GC, Pardeshi SU, Status and strategies of host plant resistance for biotic stress in cotton. In: B.M. Khadi, I.S. Katergeri, S.S. Patil, H.M. Vamadevaiah, B.R. Patil and S.M. Manjula (Eds), Proceedings of international symposium on "Strategies for Sustainable Cotton Production-A Global Vision"1, Crop Improvement, 23-25 November 2004, University of Agricultural Sciences, Dharwad, Karnataka (India), pp.31-47 (2004).
 11. Akhtar KP, Haidar S, Khan MKR, Ahmad M, Sarwar N, Evaluation of *Gossypium* species for resistance to cotton leaf curl Burewela virus. Ann Appl Biol, 157: 135-147 (2010).
 12. Kulkarni VN, Khadi BM, Manjula S, Lalitadas M, Deshapande A, Narayanan SS, The Worldwide Gene Pools of *Gossypium arboreum* L. and *G. herbaceum* L. and their Improvement A.H. Paterson (ed.), Genetics and Genomics of Cotton, Plant Genetics and Genomics: Crops and Models 3, 3: 69-100 (2009).
 13. Hutchinson JB, New evidence on the origin of the old world cotton. Heridity, 8: 225-241 (1954).
 14. Silow RA, The genetics of species development in old world cottons. J Genet, 46: 62-77 (1944).
 15. Brubaker CL, Bourland FM, Wendel JF, The origin and domestication of cotton. In: C.W. Smith and J.T. Cothren (Eds) Cotton. origin, history, technology and production. John Wiley, New York pp. 3-31 (1999).
 16. Reddy MR, Sarla M, Siddiq EA, Inter Simple Sequence Repeat (ISSR) polymorphism and its application in plant breeding. Euphytica, 128: 9-17 (2002).
 17. Zeitkiewicz E, Rafalski A, Labuda D, Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics, 20: 176-183 (1994).
 18. Gupta PK, Varsheney RK, The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, 113: 163-185 (2000).
 19. Khandagale GB, Dongre AB, Kalpande HV, Salunkhe SN, Molecular evaluation of elite cotton cultivars using DNA markers. In: Proceedings of World Cotton Research Conference-4, Lubbock TX, 10-14 September, (2007)
 20. Myskow B, Milczarski, Masoje P, Comparison of RAPD, ISSR and SSR markers in assessing genetic diversity among rye (*Secale cereale* L.) inbred lines. Plant Breed Seed Sci, 62: 107-115 (2010).
 21. Dongre AB, Bhandarkar M, Banerjee S, Genetic diversity in tetraploid and diploid cotton (*Gossypium* spp.) using ISSR and microsatellite DNA markers. Indian J Biotechnol, 6: 349-353 (2007).
 22. Bardak A, Bolek Y, Genetic diversity of diploid and tetraploid cottons by SSR and ISSR markers. Turk J Field Crops, 17: 139-144 (2012).
 23. Dongre AB, Parkhi V, Gahukar S, Characterization of cotton (*Gossypium hirsutum*) germplasm by ISSR, RAPD markers and agronomic values. Indian J Biotechnol, 3: 388-393 (2004).
 24. Noormohammadi Z, Shojaei-Jesvaghani F, Sheidai M, Farahani F, Alishah O, Inter simple sequence repeat (ISSR) and random amplified polymorphic DNA (RAPD) analyses of genetic diversity in Mehr cotton cultivar and its crossing progenies. Afr J Biotechnol, 10:11839-11847 (2011).
 25. Abdi AA, Sofalian O, Asghari A, Shokrpour M, Baqheri H, Seyyed Masoumi SY, Inter Simple Sequence Repeat (ISSR) markers to study genetic diversity among cotton cultivars in associated with salt tolerance. Not Sci Biol, 4(4): 120-126 (2012).
 26. Sheidai M, Noormohammadi Z, Shojaei-Jesvaghani F, Baraki SG, Farahani F, Alishah O, SSR (Simple sequence repeat) and ISSR (Inter simple sequence repeat) analysis of genetic diversity in tissue regenerated plants of cotton. Afr J Biotechnol, 11(56): 11894-11900 (2012).
 27. Noormohammadia Z, Taghavi E, Foroutan M, Sheidai M, Alishah O, Structure

- analysis of genetic diversity in tetraploid and diploid cotton genotypes. *Inter J Plant Ani Envir Sci*, 3: 79-90 (2013).
28. Saghai-Maroo MA, Soliman KM, Jorgensen RA, Allard RW, Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *PNAS*, 81: 8014-8018 (1984).
 29. Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorells ME, Optimising parental selection for genetic linkage maps. *Genome*, 36: 181-186 (1983).
 30. Prevost A, Wilkinson MJ, A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet*, 98(1): 107-112 (1999).
 31. Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R, Comparisons of PCR based marker systems for the analysis of genetic relationships in cultivated potato. *Mol Breed*, 3(2): 127-136 (1997).
 32. Botstein DRL, White Skolnick M, Davis RW, Construction of genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet*, 32: 314-331 (1980).
 33. Noormohammadi Z, Farahani YHA, Sheidai M, Baraki SG, Alishah O, Genetic diversity analysis in opal cotton hybrids based on SSR, ISSR and RAPD markers. *Genet Mol Res*, 12: 256-269 (2013).
 34. Kayis SA, Hakki EE, Pinarkara E, Comparison of effectiveness of ISSR and RAPD markers in genetic characterization of seized marijuana (*Cannabis sativa* L.) in Turkey. *Afr J Agric Res*, 5(21): 2925-2933 (2010).
 35. Mathew S, Molecular profiling of bauhinia accessions using RAPD markers. *Int J Pharma Bio Sci*, 1(4) : B199-207 (2010).