INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
12–656/2013/15–5–968–972
http://www.fspublishers.org

Full Length Article

Assessment of Genomic Diversity of Cotton (*Gossypium hirsutum*) Genotypes using Simple Sequence Repeats Markers through Genetic Analysis Software

Beenish Ehsan[1], Asma Haque[1], Muhammad Younas[2], Tayyaba Shaheen[1]*, Tayyaba Huma[1], Summera Sattar[1], Sobia Idrees[1] and Zafar Iqbal[2]
[1]*Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan*
[2]*Biotechnology Lab, Ayub Agriculture Research Institute, Faisalabad, Pakistan*
*For correspond: tayaba_pgmb@yahoo.com

Abstract

The present study was designed to evaluate genetic diversity among twenty cotton genotypes by employing thirty one SSR markers that were reported from Brookheaven National Laboratories. Out of thirty-one primers, six SSR primers were found polymorphic at different SSR loci. Maximum number of loci (33), were amplified by MNH-93, while minimum (22) by 238-F. Total percentage of informative loci was 36.8%. Genetic diversity matrix of these genotypes was calculated by NTSYS-pc 2.2 software. Genetic similarity coefficients ranged from 0.474 to 0.974. The highest dissimilarity (0.474) was observed between genotype 298-F and CIM-496. Cluster analysis as performed using un-weighted pair group method of arithmetic average algorithm showed that Nei and Li's coefficients ranged from 0.63 to 0.97. The dendrogram of cotton lines showed two major clusters except 238-F forming an out-group from other genotypes, thus showing maximum dissimilarity. Genotypes MNH-93 and K-68/9 showed maximum similarity coefficient of 0.97. Information obtained from the present study can be used to select divergent parents for future breeding programs.

Keywords: Genetic diversity; Cotton; polymorphism; Molecular markers

Introduction

Estimation of genetic diversity and relationships in plants is necessary for crop improvement. Characterization of genetic similarity among genotypes is a valuable source to select parental combinations for maintaining genetic diversity in a breeding program (Becelaere *et al.,* 2005). Genetic diversity studies have been carried out in cotton by employing different marker techniques e.g. amplified fragment length polymorphism (AFLP) (Abdalla *et al.,* 2001; Rana *et al.,* 2005; Li *et al.,* 2008), random amplified polymorphic DNA (RAPD) (Xu *et al.,* 2001; Chaudhary *et al.,* 2010) and simple sequences repeats (Qayyum *et al.,* 2009; Arunita *et al.,* 2010).

Molecular markers have played an important role in revealing the structure of genetic diversity in a wide range of crops. Recent advancements have made possible to gain various goals including gene targeted surveys, investigation of population dynamics, advantageous in selection for domestication (Glaszmann *et al.,* 2010). Microsatellite or short tandem repeats or simple sequences repeats are monotonous repetitions of very short (one to five) nucleotide motifs, which occur as interspersed repetitive elements in all eukaryotic genomes (Tautz and Renz, 1984).

Throughout the past decade, association mapping has been successfully applied in human genetics. It has proved a powerful and promising technique for mapping complex quantitative traits of economic significance in plants than conventional method of linkage mapping (Al-Maskri *et al.,* 2012). According to Blenda *et al.* (2006), there are various uses of microsatellites for breeders of plants. Some of these are selective breeding improvement, genetic diversity estimation, introducing novel genes into breeding germplasm from exotic germplasm, cultivar protection and locating qualitative and quantitative trait loci. For developing SSR markers, sequences flanking microsatellites, are considered and flanking sequences are conserved among individuals.

Cotton is an important crop grown for having natural textile fiber directly and used as oil seed as by-product. It serves as the backbone of textile industry (Rahman *et al.* 2012). Cotton has contributed significantly in economic development of Pakistan by contributing more than half of the total exports (GOP, 2010). Estimation of genetic relationships among crop species helps to provide information about genetic diversity. Accurate assessment of patterns and levels of genetic diversity can be very important in crop breeding for various applications (Mohammadi and Prasanna, 2003). This can be important to introduce genes of interest from diverse germplasm into the existing germplasm (Thompson *et al.,* 1998). Studies have

To cite this paper: Ehsan, B., A. Haque, M. Younas, T. Shaheen, T. Huma, S. Sattar, S. Idrees and Z. Iqbal, 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

been reported exhibiting more polymorphic potential of SSR markers than other molecular markers (Belaj *et al.,* 2003; Ferreira *et al.,* 2004) and also have been reported for showing low diversity in cotton (Gutierrez *et al.,* 2002; Rungis *et al.,* 2005).

Knowledge of genetic diversity is important in the context of gene pool selection for further breeding purposes and development of better cultivars. The present study was conducted using microsatellites, with the purpose to reveal the genetic diversity among cotton genotypes, estimation of relatedness and to characterize genotypes with distinct fingerprints.

Materials and Methods

Plant Materials

Twenty commercial cotton genotypes/varieties collected from different research institutes in Pakistan belonging to *G. hirsutum,* were studied. The parentages and research institutes of genotypes/varieties are given in (Table 1).

DNA Extraction

Fresh leaves 2-3 days old, which were grown on the top of the branches not fully expanded (4-5 g) were taken from plants. For each variety, genomic DNA was extracted by a modified cetyl trimethyl ammonium bromide (CTAB) method (Iqbal *et al.,* 1997). After extraction, quality of DNA was determined by running on 0.8% (w/v) agarose gel and DNA was quantified by NanoDrop-1000 version 3.3.1 spectrophotometer (Nanodrop, USA).

SSR Analysis

A total of 31 microsatellite primers (BNL113, BNL116, BNL118, BNL119, BNL137, BNL150, BNL193, BNL226, BNL243, BNL252, BNL285, BNL390, BNL530, BNL542, BNL569, BNL580, BNL597, BNL598, BNL632, BNL673, BNL827, BNL834, BNL840, BNL852, BNL946, BNL1045, BNL1047, BNL1053, BNL1059, BNL1061, BNL1064) reported from Brookhaven National Laboratories (BNL) were chosen randomly across the cotton genome. The primer sequences were obtained from cotton marker database (http://www.cottonmarker.org/). They were custom synthesized by Gene Link Co. USA.

In the PCR reaction mixture reagents (10X PCR Buffer ($MgCl_2$+$(NH_4)_2SO_4$), 50 mM $MgCl_2$, (1.5+1.5) µL of each Primer (forward+reverse) in concentration of 30 ng/µL, 2.0 µL of genomic DNA (30 ng/µL), 4.5 µL of 0.2 Mm of total $dNTP_S$ and 0.3 µL *Taq* DNA polymerase in the concentration of 5 unit/µL) were mixed and final volume of 20 µL was made by adding deionized double distilled water. Reagents were purchased from Fermentas, USA.

Amplification reactions were carried out using the Eppendorf Mastercycler (Germany). Mastercycler was programmed to work in the sequence as initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturing at 94°C for 30 sec, annealing at 55°C for 30 sec and extension at 72°C for 1 min. Then a final extension at 72°C for 10 min and hold at 22°C until the tubes were removed.

Concentration of PCR amplicons was illustrated using 3.0% (w/v) high resolution agarose gel and polyacrylamide gel electrophoresis to separate them on molecular weight basis. Visualization was performed using ultraviolet transilluminatoar and gel pictures were photographed by using GeneSnap version 7.6.03 of Syngene gel documentation system and GeneTool analysis system.

Data Analysis

The molecular size of each fragment was estimated using Gene Ruler 100 bp DNA ladder plus (Fermentas, USA) as standard. These were compared with the fragment size

Table1: List of cotton genotypes used in study and their origin

Varieties	Parentage	Research Institute
CIM-496	CIM-425 X 755-6/93	Central Cotton Research Institute, Multan
4F	Selection from stray plants of American Cotton	Cotton Research Institute, Faisalabad
LSS	Selection from 4 F single plant	Cotton Research Institute, Faisalabad
289-F	4 F – Selection	Cotton Research Institute, Faisalabad
199-F	Selection from 4-F-98	Cotton Research Station, Multan
238-F	289 F/43 – selection	Cotton Research Institute, Faisalabad
218-F	Unknown	Unknown
216-F	4 F – selection	Cotton Research Institute, Faisalabad
298-F	Unknown	Unknown
L-11	181 F – Selection	Cotton Research Institute, Faisalabad
262-F	Unknown	Unknown
ACD-4	Unknown	Unknown
MS-39	Natural hybrid in L 11 (Lasani 11)	Cotton Research Station, Multan
MS-40	A single variant selected from AC 252	Cotton Research Station, Multan
149-F	124FxBabdal	Cotton Research Station, Multan
Qalandri	(M 4 x *G. anomalum*) x Karnak	Cotton Research Station, Tandojam
B-557	268F x L 5	Cotton Research Institute, Faisalabad.
K-68/9	199 F x Wilds	Cotton Research Station, Ghotki
MNH-93	389/65 x C 158	Cotton Research Station, Multan
Rahmani	*G. hirsutum*21xMcNaire-TH14920	Cotton Research Station, Tandojam

reported in the literature and scored as present (1) or absent (0). Both monomorphic and polymorphic primers were included in the analysis and only reproducible fragments were considered. The data were used to estimate genetic similarity on the basis of number of shared amplification products (Nei and Li, 1979). A dendrogram based on these similarity coefficients was constructed by using unweighted pair group method of arithmetic means (UPGMA). This similarity matrix was analyzed using numerical taxonomy and multivariate analysis system (NTSYS-pc 2.2) and clustered with UPGMA algorithm to determine the genetic relationships among the 20 cotton genotypes.

Results

Out of thirty one primer pairs, six were able to reveal genetic diversity among twenty cotton genotypes (Table 2). Six polymorphic primer pairs were used to determine the genetic variability in the twenty cotton genotypes. The DNA fragments were amplified in the range of 100-500 base pairs. Maximum number of loci (about 33) was amplified by MNH-93, while minimum (about 22) by 238-F. Average number of loci per genotype were about 28.9. A total of 38 loci were amplified. Total number and percentage of informative loci were 14 and 36.8%, respectively while the total percentage of informative primers was 19.35%. Average number of loci per primer was 1.22. Different primers produced a different level of polymorphism among the different varieties (Fig. 1 and 2). Cluster analysis was performed based on similarity values (Fig. 3). The cluster analysis revealed significant genetic variation. The dendrogram from SSR analysis formed two major groups (Fig. 3). Cotton genotype 238-F remained isolated forming an out-group from the rest of the genotypes showing dissimilarity from the major group consisting rest of the genotypes. Based on genetic similarity, the major group comprising rest of the genotypes was further divided into two sub groups A and B. These two subgroups contained further sub-clusters. The genotypes present in the same sub-cluster are more alike than others. Nei and Li's coefficients ranged from 0.63 to 0.97, with an average of 0.80 accounting for 50% variation among the 20 genotypes. Sub group A contained 11 genotypes whereas genotype 189-F remained isolated forming an out-group from the rest of the genotypes in this sub group whereas other sub-clusters are formed by genotypes CIM-496 and 4-F, LSS and 218-F, 289-F and B-557 and the genotypes 216-F and Qalandri. Genotypes MNH-93 and K-68/9 were most similar to each other among the whole genotypes, accounting for similarity coefficient 0.97. These two genotypes are of same era. Sub group A contained 7 genotypes whereas genotype 298-F remained isolated, forming an out-group from the rest of the genotypes in this sub group, whereas genotypes L-11 and 149-F, 262-F and MS-39 formed sister groupings. ACD-4 and MS-40 were most dissimilar genotypes.

Table 2: List of Polymorphic Primers

Primer Name	Product Size Range
BNL150	121-131 bp
BNL193	100-120 bp
BNL 285	235-250 bp
BNL 530	130-141 bp
BNL 1047	148-176 bp
BNL 1061	140-160 bp

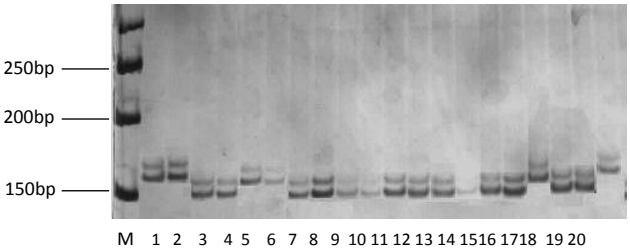

Fig. 1: PAGE picture for 20 cotton varieties (V1-V20) with primer BNL1047. Varieties (V1 and V2, V5-V7, V17 and V20) CIM-496, 4F, 199-F, 238-F, 218-F, B-557 and Rahmani showed polymorphism

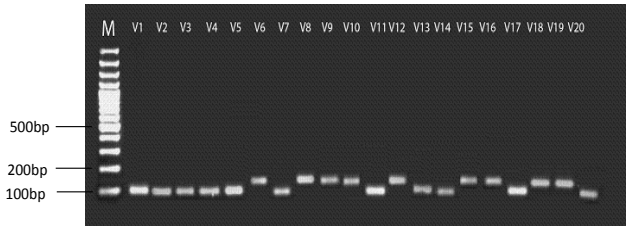

Fig. 2: Gel picture for 20 cotton varieties (V1-V20) with primer BNL193. Varieties (V6, V8-V10, V12, V15 and V16, V18 and V19) 238-F, 216-F, 298-F, L-11, ACD-4, 149-F, Qalandri, K-68/9 and MNH-93 showed polymorphism

Discussion

Zhang *et al.* (2008) described that the recent development of cotton SSR markers in large amount has raised more effort in molecular characterization of cotton germplasm originating from particular cotton breeding programs in the world. Genetic diversity was observed among twenty cotton varieties by applying thirty one SSR markers in present studies. Six SSR markers were able to show genetic diversity among twenty cotton varieties. Total percentage of informative loci was 36.8% and total percentage of informative primers was 19.35%. It has been reported in earlier studies that SSR markers target only one locus at a time and are useful source of observing molecular diversity and phylogenetic relationship (Liu *et al.,* 2006). There are many reports on efficient utilization of SSR markers in cotton (Rahman *et al.,* 2002) Recently, simple sequence repeats have been employed efficiently for distinguishing cellulase synthase genes in cotton (Lin *et al.,*

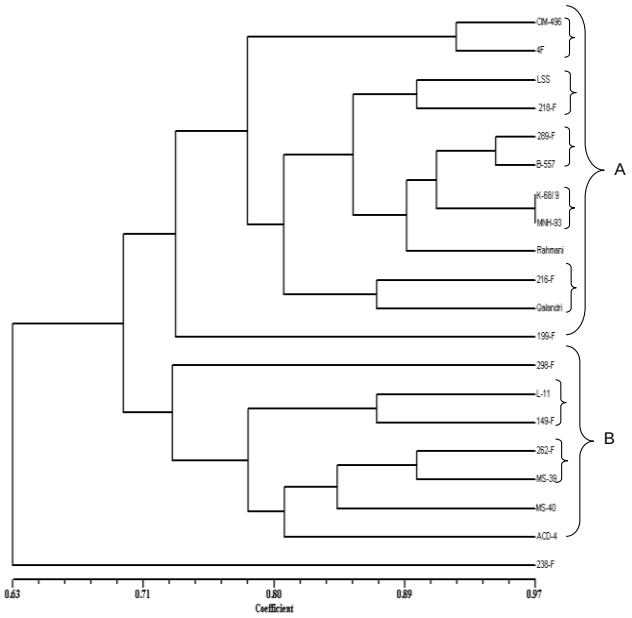

Fig. 3: Dendrogram of 20 cotton genotypes developed from SSR data using the unweighted pair group method of arithmetic means (UPGMA). The scale represents the Nei and Li's coefficients of similarity

2012) and gene flow evaluation from upland to mocó cotton (Dantas *et al.*, 2012).

In the present study, six SSR markers were able to discriminate among different varieties but none of them was able to distinguish all the varieties. Markers showed a considerable genetic diversity among varieties. According to Powell *et al.* (1996), SSR loci give good discrimination between closely related individuals in some cases even when only a few loci are employed. SSR markers have proved beneficial for the diversity analysis in cultivars regenerated using tissue culture technique (Sheidai *et al.*, 2012).

In present study dendrogram showed a unique relationship among genotypes and therefore genotypes can be identified by relationship among themselves. MNH-93 and K-68/9 are most similar among whole genotypes. The results of the present study can further be utilized for selection of parental genotypes for breeding programs to develop better genotypes.

In conclusion, the results indicated less genetic diversity among genotypes. The information, we gained in the present study provides us a useful guidance for cotton breeding programs. The cultivars used in this study can be used as parents for broadening the genetic base and also can be used for the evolution of new high yielding varieties to increase the productivity and quality of cotton.

References

Abdalla, M., O.U.K. Reddy, K.M. El-Zik and A.E. Pepper, 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor. Appl. Genet.*, 102: 222–229

Al-Maskri, A.Y., M. Sajjad and S.H. Khan, 2012. Association mapping: a step forward to discovering new alleles for crop improvement. *Int. J. Agric. Biol.*, 14: 153–160

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

Becelaere, G.V., E.L. Lubbers, A.H. Paterson and P.W. Chee, 2005. Pedigree- vs. DNA marker-based genetic similarity estimates in cotton. *Crop Sci.*, 45: 2281–2287

Belaj, A., Z. Satovic, G. Cipriani, L. Baldoni, R. Testolin, L. Rallo and I. Trujillo, 2003. Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theor. Appl. Genet.*, 107: 736–744

Blenda, A., J. Scheffler, B. Scheffler, M. Palmer, J. Lacape, J.Z. Yu, C. Jesudurai, S. Jung, S. Muthukumar, P. Yellambalase, S. Ficklin, M. Staton, R. Eshelman, M. Ulloa, S. Saha, B. Burr, S. Liu, T. Zhang, D. Fang, A. Pepper, S. Kumpatla, J. Jacobs, J. Tomkins, R. Cantrell and D. Main, 2006. CMD: a cotton microsatellite database resource for *Gossypium* genomics. *BMC Genomics*, 7: 132

Chaudhary, L., A. Sindhu, M. Kumar, R. Kumar and M. Saini, 2010. Estimation of genetic divergence among some cotton varieties by RAPD analysis. *J. Plant Breed. Crop Sci.*, 2: 39–43

Dantas, A.C.A., P.A.V. Barroso, L.V. Hoffmann, M.F. Alves and F.P. Andrade, 2012. SSR markers to detect gene flow from upland to mocó cotton. *Rev. Ciênc. Agron.*, 43: 163–169

Ferreira, A.M., A.R. Wagner, V. Carneiro, G.P. Brandao and M. Melo, 2004. Genetic variability of Brazilian toxoplasma gondii strains detected by random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) and simple sequence repeat anchored-PCR (SSR-PCR). *Infect. Genet. Evol.*, 4: 131–142

Glaszmann, J.C., B. Kilian, H.D. Upadhyaya and R.K. Varshney, 2010. Accessing genetic diversity for crop improvement. *Curr. Opin. Plant. Biol.*, 13: 167–173

GOP, 2010. *Economic Survey of Pakistan: Finance Division*. Government of Pakistan

Gutierrez, O.A., S. Basu, S. Saha, J.N. Jenkins, D.B. Shoemaker, C.L.A. Cheatham and J.R.J.C. Mccarty, 2002. Genetic distance among selected cotton genotypes and its relationship with F_2 performance. *Crop Sci.*, 42: 1841–1847

Iqbal, M.J., N. Aziz, N.A. Saeed, Y. Zafar and K.A. Malik, 1997.Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor. Appl. Genet.*, 94: 139–144

Li, Z., X. Wang, Y. Zhang, G. Zhang, L. Wu, J. Chi and Z. Ma, 2008. Assessment of genetic diversity in glandless cotton germplasm resources by using agronomic traits and molecular markers. *Front. Agric. Chin.*, 2: 245–252

Lin, Z., Y. Wang, X. Zhang and J. Zhang, 2012. Functional Markers for Cellulose Synthase and Their Comparison to SSRs in Cotton. *Plant Mol. Biol. Rep.*, 30: 1270–1275

Liu, D., X. Guo, Z. Lin, Y. Nie and X. Zhang, 2006. Genetic diversity of Asian cotton (*Gossypium arboretum* L.) in China evaluated by microsatellite analysis. *Genet. Resour. Crop Evol.*, 53: 1145–1152

Mohammadi, S.A. and B.M. Prasanna, 2003. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop Sci.*, 43: 1235–1248

Nei, M. and W.H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA*, 76: 5269–5273

Powell, W., M. Morgante, C. Andre, M. Hanafey, M.J. Vogel, S.V. Tingey and A. Rafalski, 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellites) markers for germplasm analysis. *Mol. Breed.*, 2: 225–238

Qayyum, A., N. Murtaza, F.M. Azhar and W. Malik, 2009. Biodiversity and nature of gene action for oil and protein contents in *Gossypium hirsutum* L. estimated by SSR markers. *J. Food Agric. Environ.*, 7: 590–593

Rahman, M., T. Shaheen, N. Tabbasam, M.A. Iqbal, Y. Zafar and A.H. Paterson, 2012. Cotton genetic resources- A review. *Agron. Sustain. Dev*., 32: 419–432

Rahman, M., T.A. Malik, N. Aslam, M. Asif, R. Ahmad, I.A. Khan and Y. Zafar, 2002. Optimization of PCR conditions to amplify microsatellite loci in cotton (*Gossypium hirsutum* L.) genomic DNA. *Int. J. Agric. Biol*., 4: 282–284

Rana, M.K., V.P. Singh and K.V. Bhat, 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

Rungis, D., D. Llewellyn, E.S. Dennis and B.R. Lyon, 2005. Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Aust. J. Agric. Res.*, 56: 301–307

Sheidai, M., Z. Noormohammadi, F.S. Jeshvaghani, S.G. Baraki, F. Farahani and O. Alishah, 2012. Simple sequence repeat (SSR) and inter simple sequence repeat (ISSR) analyses of genetic diversity in tissue culture regenerated plants of cotton. *Afr. J. Biotechnol.*, 11: 11894–11900

Tautz, D. and M. Renz, 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res*., 12: 4127–4138

Thompson, J.A. and R.L. Nelson, 1998. Utilization of diverse germ-plasmfor soybean yield improvement. *Crop Sci*., 38: 1362–1368

Xu, Q.H., X.L. Zhang and Y.C. Nie, 2001. Genetic diversity evaluation of cultivars (*G. hirsumtum* L.) from the Changjiang river valley and Tellow river valley by RAPD markers. *Acta Genet. Sin.*, 28: 683–690

Zhang, H., Y. Li, B. Wang and P.W. Chee, 2008. Recent advances in cotton genomics. *Int. J. Plant Genomics*, 2008: 742304

(Received 11 August 2012; Accepted 07 may 2013)