Pak. J. Bot., 39(4): 1245-1250, 2007.

IDENTIFYING GENETIC VARIATION IN *GOSSYPIUM* BASED ON SINGLE NUCLEOTIDE POLYMORPHISM

SAGHIR AHMAD[1], TIANZHEN ZHANG[3], NOOR-UL-ISLAM[4], TAYYABA SHAHEEN[5] AND MEHBOOB-UR-RAHMAN[5]

[1]*Cotton Botanist, Cotton Research Station, Vehari, Pakistan,* [1]*Department of Botany, University of Agriculture, Faisalabad, Pakistan,* [3]*Director, Cotton Institute, Vice-Dean Agriculture College, Nanjing Agriculture University China,*
[4]*Director, Cotton Institute, Faisalabad, Pakistan and*
[5]*Senior Scientific Officer, NIBGE, Faisalabd, Pakistan*

Abstract

Various molecular markers such as RAPD, RFLP, AFLP, SSR and EST-SSR are being employed to detect genetic diversity and for marker-assisted selection (MAS) in cotton. But little information is available regarding the use of Single Nucleotide Polymorphisms (SNPs) in this crop. In the present study, polymorphism in the ***FIF1*** gene, important in regulating cotton fiber development in *Gossypium barbadense* L., was studied among *G. hirsutum* cv. TM-1 and *G. barbadense* cv. Hai7124. ***FIF1*** genes from *G. hirsutum* cv. TM-1 and *G. barbadense* cv. Hai7124 were cloned on the basis of its reported sequence in *G. arboreum* L. Three bases substitutions were detected at the 188-, 283-, 413-bp position of the ***FIF1*** sequence of *G. barbadense* while *G. hirsutum* has the same sequence as in *G. arboreum*. The first base change (188 T-C) is located in the non-coding region, and the second (283 A-T) and the third (413 C-T) in the coding region, in which AGG (Arg) is changed to a TGG (Trp) in 283-SNP, and CCA (Pro) to a CTA (Leu) in 413-SNP. Thus it could be concluded that SNP molecular marker can be used to identify minute difference (which may or may not morphologically express) in germplasm of cotton.

Introduction

Cotton is the world's leading food and fiber crop (Fryxell, 1992). Molecular markers like Randomly Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), Simple Sequence Repeats (SSRs) and Expressed Sequence Tags Simple Sequence Repeats (EST-SSR) have been employed to estimate genetic diversity as well as for marker-assisted selection (MAS) in cotton (Rahman *et al*., 2002; Buyyarapu *et al*., 2006; Rahman *et al*., 2006; Rahman *et al*., 2007). But the available literature suggests little information regarding the use of Single Nucleotide Polymorphism (SNP) in cotton. Han *et al*., (2006) used SNP to study diversity in cotton genotypes, developing SNP markers and allocating their position on genetic maps. In other crops like soybean (*Glycine max* L.), Zhu *et al*., (2003) studied 25 different soybean genotypes and reported 280 SNPs in 76kb sequence**.** Rong *et al*., (2004) found that the rates of intra-specific DNA sequence variation within amplicons derived from the genetically mapped sequence-tagged sites in cotton are somewhat higher than that of human or other taxa. Presently SNPs are the most popular markers because they occur more frequently in genomes than microsatellites thus providing larger sets of markers near or in any locus of interest (Landegren *et al*., 2006). Some of the single base changes may also effect individuals phenotype by changing protein structure or expression level therefore these may represent candidate alterations for genetic mechanism in disease (Landegren *et al*., 2006). Repeated sequences (SSR) exhibit instability, i.e., mutation occasionally alter the size of an allele making inheritance more complicating (Webber & Wang, 1993), but SNPs are highly stable and often contributes directly to a phenotype (Kim *et al*., 2004). It is anticipated that with the

availability of efficient typing systems with large throughputs scanning of putative risk factors like minor changes at DNA level will be possible. Efforts are being made to develop high throughput genotyping methods to detect SNPs based on allele-specific hybridization. Various SNP genotyping techniques such as primer extension, oligonucleotide ligation, endonuclease cleavage, Invader, Taqman, or allelic-specific PCR have been developed, (Gupta *et al*., 2001; Gut, 2001; Shi, 2001; Syvänen, 2001)**.** An ideal SNP marker should be simple, low cost and reproducible. Mining for SNPs in cotton is just in infancy. Major dilemma in cotton is the huge genome size and occurrence of polyploidy in cotton. Data mining for SNPs has been done throughout the diploid genomes of cotton by direct sequencing of PCR products amplified by primers based on sequence tagged sites of cotton genome derived from genetic maps. A total of 1,164 candidate polymorphisms were detected (anonymous). Shah *et al*., (2004) used SSR primers to discover SNPs in the flanking regions of SSRs in cotton. Due to availability of SSR primers this technique may prove very economical. Lu *et al*., (2005) estimated the nucleotide diversity in cotton fiber genes among cultivated tetraploid cotton for the development of SNP-based cotton linkage map and quantitative trait loci detection. DNA sequencing in tetraploid cotton revealed heterogeneity of nucleotide diversity.

In the present study, we used *GaMYB2*/Fiber Factor 1 (*FIF1*) gene, which is predominantly expressed early in developing cotton fibers in *Gossypium barbadense* L. The first intron of both *GL1* and *GaMYB2* plays a key role in trichomes patterns, it acts as an enhancer in trichome and a repressor in nontrichome cells, generating a trichome specific pattern of MYB gene expression. Any change /disruption of a MYB motif conserved in intron 1 of *GL1*, *WEREWOLF* and *GaMYB2* genes affects trichome production, which suggests that GaMYB2 may play a key role in regulating fiber development in cotton (Wang *et al*., 2004).

Materials and methods

Isolation of DNA: The plant material comprising two tetraploid cotton cultivars i.e., TM-1 and a *Verticillium*-resistant cultivar Hai7124, was used for identifying polymorphism on single nucleotide basis. At six leaf stage, the young leaves were collected from the field and DNA was extracted using the CTAB (cetyltrimethyl ammonium bromide) method (Paterson *et al*., 1993).

Primer designing: To obtain PCR products corresponding to the *FIF1* gene from both of the TM-1 and Hai7124, primers FIF1L5’CTTCCCTTCTCACTCTTTGC3’and FIF1R5’TAGGCCAAGAAACATTTGGT3’ were designed to cover its full ORF sequences region based on the reported sequence of *FIF1* gene cloned from *G. arboreum* L. (GenBank Accession Number: AY626160) using the Primer3 program online. The PCR product/ reaction mixture (30 µl) was prepared using 30 ng of genomic DNA, 10 pmol of each forward and reverse primer, 150 µmol/L of dNTPs, 1.5 mmol/L $MgCl_2$, 1 × reaction PCR buffer and 1 U of *Taq* DNA polymerase (TaKaRa, Dalian, China). The reaction mixture was denatured at 95°C for 5 min and subjected to 30 cycles of 94°C for 30 s, 57°C for 1/min., 72°C for 1/ min and 72°C for 10 min., on a PTC-225 (MJ Research, USA). PCR products were visualized by electrophoresis on 1.0% (*w*/*v*) agarose gel. The products amplified from TM-1 and Hai7124 were eluted from the gel with Mini DNA Rapid Purification kit (Huashun, China). The DNA thus recovered was tested for quality. For this purpose 4 µl DNA from both cultivars TM-1 and Hai7124 was taken in 1.5 ml tube alongwith 2µl 10x loading buffer. It was run on 1.0% (*w*/*v*) agarose. After checking the quality of DNA, the process for gene cloning started.

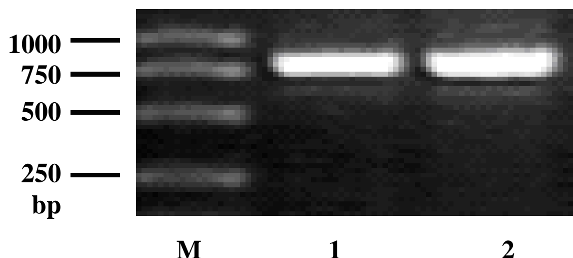

Fig. 1. Gel electrophoresis results of TM-1 (1) and Hai7124 (2).

Cloning and sequencing of the gene: Gene was cloned into T easy cloning vector and sequencing was done on ABI automated sequencer. DNA sequences from TM-1 and Hai7124 were compared to identify SNP for the *FIF1* gene by alignment of obtained sequences using the Clustal W (www.ebi.ac.uk/clustalw/).

Results and Discussion

PCR amplicons resolved on agarose gel could not show any polymorphism in *FIF1* gene of TM-1 (*G. hirsutum*) and Hai7124 (*G. barbadense*). The similar results have been reported by Han *et al*., (2006). When compared with standard molecular marker, DL 2000, the bands of TM-1 and Hai7124 were found to be 810 bps. In order to find the difference on the basis of single nucleotide, ***FIF1*** genes from TM-1 and Hai7124 were cloned, sequenced and then using Clustal W programme, the full length nucleotide sequence of both the cultivars was aligned and analyzed for presence of SNPs. The results indicated three bases substitutions: at 188bp position, TM-1 gene contains T base, whereas 7124 had C base at same position (Table 1). The second change recognized was at 283 bp, A base in TM-1 was substituted by T (7124), whereas C-T substitution was observed at 413 bp (Table 1). The first base change (188 T-C) is located in the non-coding region, and the second (283 A-T) and the third (413 C-T) in the coding region, in which AGG (Arg) is changed to a TGG (Trp) in 283-SNP, and CCA (Pro) to a CTA (Leu) in 413-SNP. *Gossypium arboreum* L., containing A genome is thought to evolve from its ancestor 6-11 MYA and then reunited with D genome to produce tetraploid cotton species nearly 1.1-1.9 MYA (Wendel & Cronn, 2003). Results of the present study indicate that all variations are present in *G. barbadense*, it is still to be explored that either these variations have any contribution to superb fibre qualities of *G. barbadense*. It can be demonstrated from the results that *GaMYB2*/Fiber Factor 1 (*FIF1*) gene in *G. hirsutum* has not been affected during course of evolution while in case of *G. barbadense* nucleotide changes are present at three different positions. Presence of these variations only in *G. barbadense* indicates that these variations have been acquired after the polyploidization event has taken place. Although other molecular markers are important for estimation of diversity and evolution, SNPs are expedient in detection of variations, which are not detected by other molecular markers (Batley *et al*., 2003). SNPs can be used to develop gene specific markers, to study frequency of variations in genomes and to study evolution. In preliminary results 1 SNP per 37 bp were detected in some house

keeping genes in diploid genomes of cotton (Shaheen & Co-workers, unpublished data). It can thus be concluded from these results that SNP is a handy tool for depicting differences at single base level, which cannot be detected by other molecular markers.

Table 1. Results of the three *FIF1* genes amplified from TM-1 in *G. hirsutum,* Hai7124 in *G. barbadense*, and *FIF1* gene in *G. arboreum* L., through the program Clustal W online. Presence of SNPs is represented by bold letters.

FIF1	CTTCCCTTCTCACTCTTTGCCTCTTCTCACTGTCGGCTAATAATGGCTCCAAAGAAGGAT	60
TM-1	CTTCCCTTCTCACTCTTTGC TCTTCTCACTGTTGGCTAATAATGGCTCCAAAGAAGGAT	60
7124	CTTCCCTTCTCACTCTTTGCCTCTTCTCACTGTTGGCTAATAATGGCTCCAAAGAAGGAT	60
FIF1	GGAGTGAGCAAAAGGGTTTTTAACAAAGGTTCTTGGACAGCTGAGGAAGATAGAAGATTG	120
TM-1	GGAGTGAGCAAAAGGGTTTTTAACAAAGGTTCTTGGACAGCTGAGGAAGATAGAAGATTG	120
7124	GGAGTGAGCAAAAGGGTTTTTAACAAAGGTTCTTGGACAGCTGAGGAAGATAGAAGATTG	120
FIF1	GCTAAATATATTGAGATTCATGGCGCAAAGAGATGGAAAACAATCGCCATTAAATCAGGT	180
TM-1	GCTAAATATATTGAGATTCATGGCGCAAAGAGATGGAAAACAATCGCCATTAAATCAGGT	180
7124	GCTAAATATATTGAGATTCATGGCGCAAAGAGATGGAAAACAATCGCCATTAAATCAGGT	180
FIF1	AATTTACTTTCTGTTGAAGAGAAACTCCATTTTTTGGAGCATTGGAATTAACGGGTTATT	240
TM-1	AATTTAC**T**TTCTGTTGAAGAGAAACTCCATTTTTTGGAGCATTGGAATTAACGGGTTATT	240
7124	AATTTAC**C**TTCTGTTGAAGAGAAACTCCATTTTTTGGAGCATTGGAATTAACGGGTTATT	240
FIF1	TTGTTTTGATGTATAGGTTTGAATCGATGCGGCAAGAGTTGCAGGTTGAGATGGTTGAAC	300
TM-1	TTGTTTTGATGTATAGGTTTGAATCGATGCGGCAAGAGTTGC**A**GGTTGAGATGGTTGAAC	300
7124	TTGTTTTGATGTATAGGTTTGAATCGATGCGGCAAGAGTTGC**T**GGTTGAGATGGTTGAAC	300
FIF1	TACTTGAGACCTAACATTAAGAGAGGCAACATATCAGATGAAGAAGAGGACTTAATTATT	360
TM-1	TACTTGAGACCTAACATTAAGAGAGGCAACATATCAGATGAAGAAGAGGACTTAATTATT	360
7124	TACTTGAGACCTAACATTAAGAGAGGCAACATATCAGATGAAGAAGAGGACTTAATTATT	360
FIF1	AGGCTTCATAAACTGCTGGGGAACAGGTGGTCTTTGATTGCTGGGAGACTTCCAGGGCGA	420
TM-1	AGGCTTCATAAACTGCTGGGGAACAGGTGGTCTTTGATTGCTGGGAGACTTC**C**AGGGCGA	420
7124	AGGCTTCATAAACTGCTGGGGAACAGGTGGTCTTTGATTGCTGGGAGACTTC**T**AGGGCGA	420
FIF1	ACAGACAATGAAATTAAGAACTACTGGAATTCCCATTTGAGCAAGAAAATAATAAACCAT	480
TM-1	ACAGACAATGAAATTAAGAACTACTGGAATTCCCATTTGAGCAAGAAAATAATAAACCAT	480
7124	ACAGACAATGAAATTAAGAACTACTGGAATTCCCATTTGAGCAAGAAAATAATAAACCAT	480
FIF1	GATGTCAGAACAGAACAAACTTCCTCCTCGGAACAAATTGTGCCTCACAAAGCATGGGAA	540
TM-1	GATGTCAGAACAGAACAAACTTCCTCCTCGGAACAAATTGTGCCTCACAAAGCATGGGAA	540
7124	GATGTCAGAACAGAACAAACTTCCTCCTCGGAACAAATTGTGCCTCACAAAGCATGGGAA	540
FIF1	ACTGTCCATATGGAAGAAGAAGAGGTAGTAAAAGGAAGTGATGAAATTGAAAACTCTGAA	600
TM-1	ACTGTCCAGATGGAAGAAGAAGAGGTAGTAAAAGGAAGTGATGAAATTGAAAACTCTGAA	600
7124	ACTGTCCAGATGGAAGAAGAAGAGGTAGTAAAAGGAAGTGATGAAATTGAAAACTCTGAA	600
FIF1	TTCAGCATTGATGTGGACGAATTCTTTGACTTCACAACGGAAGGTTGCTTTACTTTGGAT	660
TM-1	TTCAGCATTGATGTGGACGAATTCTTTGACTTCACAACGGAAGGTTGCTTTAGTTTGGAT	660
7124	TTCAGCATTGATGTGGACGAATTCTTTGACTTCACAACGGAAGGTTGCTTTAGTTTGGAT	660
FIF1	TGGGTGAATAAGTTCCTTGAACTTGATGATCAACAGGATCCATTAGCAATGGTATAATAA	720
TM-1	TGGGTGAATAAGTTCCTTGAACTTGATGATCAACAGGATCCATTAGCAATGGTATAATAA	720
7124	TGGGTGAATAAGTTCCTTGAACTTGATGATCAACAGGATCCATTAGCAATGGTATAATAA	720
FIF1	GTTTGTAATTAACATGTTTCCTTGGATAAATAAATAAAGAGATGTTCTGAGTCTAAAAAT	780
TM-1	GTTTGTAATTAACATGTTTCCTTGGATAAATAAATAAAGAGATGTTCTGAGTCTAAA---	777
7124	GTTTGTAATTAACATGTTTCCTTGGATAAATAAATAAAGAGATGTTCTGAGTCTAAA---	777
FIF1	AAATAAAGAGATGTTCTGAGTCTAAATTTCTAGGCTATAACCAAATGTTTCTTGGCCTA	839
TM-1	--------------------------TTTCTAGGCTATAACCAAATGTTTCTTGGCCTA	810
7124	--------------------------TTTCTAGGCTATAACCAAATGTTTCTTGGCCTA	810

References

Batley, J., R. Mogg, D. Edwards, H. O'Sullivan and K.J. Edwards. 2003. A high-throughput SNuPE assay for genotyping SNPs in the flanking regions of *Zea mays* sequence tagged simple sequence repeats. *Mol. Breed*., 11: 111-120.

Buyyarapu, R., S. Sha, J. Yu, K. Soliman, G. Sharma and R. Kantety. 2006. Development and mapping of single nucleotide polymorphism markers based on candidate genes and expressed sequence tags in *Gossypium* species. *Plant & Animal Genomes* XIV *Conference*. Jan. 14-18, Town & County Convention Center, San Diego, CA.

Fryxell, P.A. 1992. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea*., 2: 108-165.

Gupta, P.K., J.K. Roy and M. Prasad. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Sci.*, 80: 524-535.

Gut, I.G. 2001. Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat.*, 17: 475-492.

Han, Z.G., W.Z. Guo, X.L. Song and T.Z. Zhang. 2006. *Mapping of One Fiber Factor 1 gene based on single nucleotide polymorphism in Gossypium*. Ph.D. Thesis. National Key Lab. Of Crop Genetics and Germplasm Enhancement, Cotton Res. Inst. Nanjing Agri. Uni. Nanjing, China.

Kim, M.Y., B.K. Ha, T.H. Jun, E.Y. Hwang, K. Van, Y.I. Kuk and S.K. Lee. 2004. Single nucleotide polymorphism discovery and linkage mapping of lipoxygenase-2 gene *Lx2* in soybean. *Euphytica.*, 135: 169-177.

Landegren, U., M. Nilsson and P.Y. Kwok. 2006. Reading bits of genetic information: methods for single nucleotide polymorphism analysis. *Genome Res*., 1998 Cold Spring Harbor Lab. Press. 769-776.

Lu, Y., J. Curtiss, J. Zhang, R.G. Percy and R.G. Cantrell. 2005. Discovery of Single Nucleotide Polymorphisms in Selected Fiber Genes in Cultivated Tetraploid Cotton. *National Cotton Council Beltwide Cotton Conference*, Abstract, p. 946.

Shah, M.R., S.P. Kumpatla, M. Gupta and S.A. Thompson. 2004 SSR-based SNP discovery: an economical approach for the development of SNP markers in cotton. *Town & Country Convention Center San Diego*, CA. p. 301

Paterson, A.H., C. Brubaker and J.F. Wendel. 1993. A rapid method for extraction of cotton (*Gossypium* spp.) Genomic DNA suitable for RFLP or PCR analysis. *Plant Mol Bio Rep.*, 11: 122-127.

Rahman, M., D. Hussain and Y. Zafar. 2002. Estimation of genetic divergence among elite cotton cultivars–genotypes by DNA fingerprinting technology. *Crop Sci*., 42: 2137-2144.

Rahman, M., T. Yasmin, H.L. Shaheen, T. Shaheen, N. Tabassum and Y. Zafar. 2006. Harvesting genetic diversity in the cotton genome. *National symposium on biotechnology for economic prosperity*. July 24-26, 2006 Green Retreat Hotel, Nathiagali, Pakistan. pp. 38.

Rahman, M., T. Yasmin, N. Tabassum, I. Ullah, M. Asif and Y. Zafar. 2007 Studying the extent of genetic diversity among *Gossypium arboreum* L., genotypes/cultivars using DNA fingerprinting. (in press, *Genetic Res & Crop Evo*).

Rong, J. K., C. Abbey, J.E. Bowers, C.L. Brubaker, C. Chang, P.W. Chee, T.A. Delmonte, X. Ding, J.J. Garza, B.S Marler, C. Park, G.J. Pierce, K.M. Rainey, V.K. Rastogi, S.R. Schulze, N.L. Trolinder, J.F. Wendel, T.A. Wilkins, T.D. Williams-Coplin, R.A. Wing, R.J. Wright, X. Zhao, L. Zhu, A.H. Paterson. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics.*, 166: 389-417.

Shi, M.M. 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin Chem.*, 47: 164-172.

Syvänen, A.C. 2004. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Rev. Genet*., 2: 930-942

Wang, S., J. Wang, N. Yu, C. Li, B. Luo, J. Gou, L. Wang and X. Chen. 2004. Control of Plant Trichome Development by a Cotton Fiber MYB Gene. *Plant Cell.,* 16(9): 2323-2334.

Webber, J.L. and C. Wong. 1993. Mutation of short tandem repeats. *Hum. Mol. Genet*., 2: 1123-1128.

Wendel, J. F. and R.C. Cronn. 2003. Polyploidy and the evolutionary polyhistory of cotton. *Adv. Agron*., 78: 139-186.

Zhu, Y.L., Q.J. Song, S.M. Hyten, E.W. Fickus, N.D. Young, P.B. Cregan. 2003. Single-nucleotide polymorphism in soybean. *Genetics*., 163: 1123-1134.

(Received for publication 21 April 2007)