Electronic Journal of Biotechnology ISSN: 0717-3458
http://www.ejbiotechnology.info
DOI: 10.2225/vol13-issue5-fulltext-3

Detection of single nucleotide polymorphisms in the conserved ESTs regions of *Gossypium arboreum*

Tayyaba Shaheen[1] · Yusuf Zafar[1] · Mehboob-ur-Rahman[1] ✉

1 Plant Genomics and Molecular Breeding Labs, National Institute for Biotechnology and Genetic Engineering, PO Box 577, Jhang Road, Faisalabad, Pakistan

✉ Corresponding author: mehboob_pbd@yahoo.com
Received February 21, 2010 / Accepted May 5, 2010
Published online: September 15, 2010

Abstract Exploring genetic variation in *Gossypium arboreum* L. germplasm is useful as it contains many important genes conferring resistance to different stresses. In limited earlier studies, low level of genetic diversity was found by using conventional DNA marker systems which may impede future genome mapping studies. In the present investigation, we explored the extent of Single Nucleotide Polymorphisms (SNP) among 30 conserved regions of Expressed Sequence Tags (EST) of low copy genes between two genotypes of *G. arboreum*. A total of 27 SNPs including 21 substitutions and 6 Insertions and deletions (Indels) in 7804 bp were found between these genotypes with a frequency of one SNP per 371 bp and one Indel after every 1300 bp. Out of these SNPs, 52% were transitions, whilst 48% SNPs were transversion. In conclusion, SNPs are expedient markers that can explore polymorphism in highly conserved sequences where other markers are not effective.

Keywords: conserved regions, *Gossypium arboretum*, SNPs

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are single base change or small insertions and deletion (Indels) in homologous DNA fragments. SNPs are the most abundant source of polymorphisms which have potential to be used in association mapping studies (Ayeh, 2008). For example, human genome contains ~9-10 million SNPs, of which 3.1 million have been identified (The International Hapmap Consortium, 2007), thus are preferred over the other marker assays. SNPs are useful for characterizing allelic variation, quantitative trait locus (QTL) mapping, and implementing marker-assisted selection (MAS) in plant breeding.

In multiple investigations, Expressed sequence tags (ESTs) have been used as a source for identifying SNPs in many plant species such as maize (*Zea mays* L.)

(Ching et al. 2002: Barbazuk et al. 2007), rice (*Oryza sativa* L.) (Nasu et al. 2002) and soybean (*Glycine max* L. Merr.) (Zhu et al. 2003: Choi et al. 2007). Recently, new re-sequencing approaches (array-based methods) have been developed to identify SNPs. For example, in *Arabidopsis* more than one million non-redundant SNPs were identified which can be used in disequilibrium mapping studies (Clark et al. 2007). However, reports on identification of SNPs in cotton are meager because of its huge genome size coupled with the polyploidy nature of the cultivated cotton (allotetraploid), which requires the distinction of allelic SNPs, from paralogs (Rahman et al. 2009).

Mining SNPs in diploid genomes is more feasible due to low level of complexity in diploid genomes (Wang et al. 2005; Shaheen et al. 2006). *G. arboreum* is a diploid cultivated cotton species has been present in Pakistan, in a domesticated form, since before 6,000 BC (Moulherat et al. 2002). The species has adaptive features like a deep root system, resistance to insect pests/diseases and indehiscent bolls, which can be utilized in isolation of important genes (Arpat et al. 2004). To the extent of our knowledge, SNPs have not been reported in the nuclear genome of *G. arboreum* accessions.

In the present study, conserved coding regions, least prone to mutations (Koornneef et al. 2004), have been selected to identify SNPs in low copy number genes (preferable single copy) in two *G. arboreum* genotypes (evolved from two different breeding programs) as the other marker systems can not be effective to explore these regions (Semagn et al. 2006). This preliminary information will set a stage for developing high genetic linkage maps which will be useful for trait mining.

MATERIALS AND METHODS

Experimental material used in this study was a local cultivar of *G. arboreum* 'Ravi' which was a selection from variety 465-D. Total genomic DNA was isolated by a method used by Iqbal et al. (1997). EST sequences of *G. arboreum* accession 8401 (Developed in India for long staple length) were obtained from cotton db EST (Udall et al. 2006).

Gene specific primers were designed based on conserved regions of ESTs showing homology with low copy genes, preferable one copy gene (Table 1). These ESTs were selected out of 1000 ESTs on basis of their homology with genes of known function and low copy number. Polymerase chain reaction (PCR) was performed in a total volume of 20 µl, using 2.5 µl (15 ng/µl) of cotton DNA, 10 x PCR buffer without $MgCl_2$ (10 mM Tris-HCl, 50 mM KCl, PH 8.3), 3 mM $MgCl_2$, 0.1 mM each of dATP, dGTP, dCTP and dTTP and 0.5 units of Taq DNA polymerase, 0.15 mM of each primer. Taq DNA polymerase together with 10 x PCR buffer, $MgCl_2$ and dNTPs were from MBI Fermentas. Polymerase chain reaction consisted of 35 cycles of 94ºC for 1 min, 94ºC for 30 sec, 50ºC for 30 sec, 72ºC extension for 1 min and final extension at 72ºC for 10 min. PCR products were resolved on 1% agarose to check amplification.

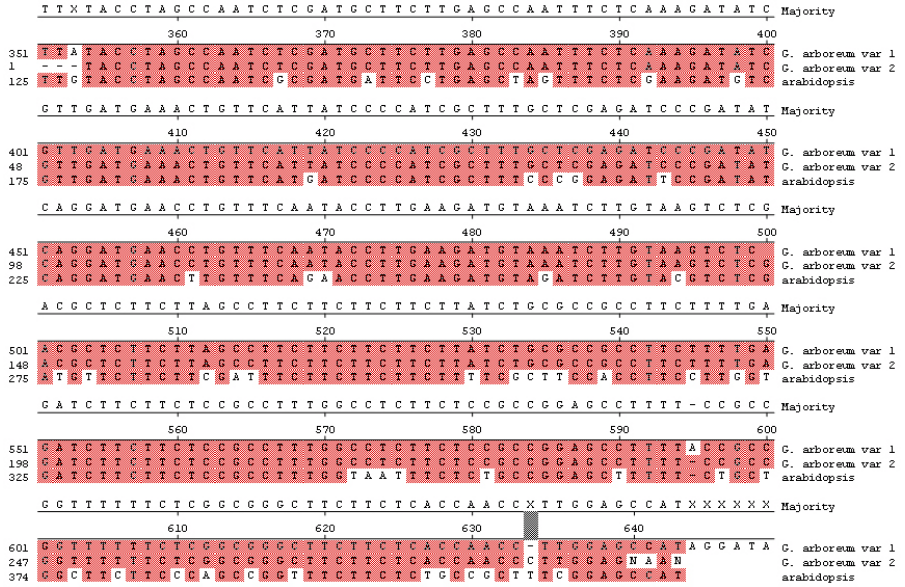

Fig. 1 Alignments to detect SNPs in conserved regions (Exon regions). 1) Histone H2B-3.

Sequencing of PCR products was done on ABI automated DNA sequencer. Sequences were edited manually. To avoid discrepancies in SNP detection, 4 runs of each of the product for sequencing were conducted. Sequences of *Arabidopsis thaliana* were obtained from GenBank. Thirty gene sequences from two diploid varieties were used for SNP detection (Table 1). Only those SNPs were considered which were detected repeatedly in all sequencing results. Consensus sequence of the 4 runs were used for alignment. DNASTAR (DNASTAR Inc., Madison, WI, USA) and Clustal v were used for sequence alignment (Figure 1).

RESULTS AND DISCUSSION

We identified SNPs in the sequenced PCR amplified products of conserved regions of ESTs of *G. arboreum*. As the primers were based on conserved EST regions and the amplified product size was same as was expected from primers amplification. Hence, it can be concluded that the primers amplified only the exon regions. Development of new SNPs by re -sequencing of PCR amplicons with or without pre-screening has been reported in previous studies (Ayeh, 2008). In most plant species, SNPs have been detected by comparison of two accessions as in maize (Ching et al. 2002) and soybean (Zhu et al. 2003). Similarly, in tetraploid cotton, a PCR based direct DNA

Table 1. *G. arboreum* ESTs selected for SNPs study, their best blast homology, primer sequences, copy number of genes showing homology with ESTs and their position in *Arabidopsis thaliana*.

	EST	Best Blast homology	Primer sequence	Copy number in Arabidopsis	Chromosomal position in Arabidopsis
1	CON_005_03587	Histone H2B-3	GCTTAGCCAATTCACCAGGC GCTCCAAGGTTGGTGAGAAG	Single	Chr.2
2	CON_001_09243	NADPH-cytochrome P450 reductase	CACCCATTTAACCCTTCTCGC TGTATGTGTGTGGTGATGCC	Single	Chr.4
3	CON_002_00475	Werewolf	AGGGATTATGGGCAATGGAG GTGGTTTCAGTGGTTGTTCC	Single	Chr.5
4	CON_003_03042	Putative ATP binding protein	GCCATGGTGAATACTCTGAG TCGCCTTCAGAGAAGATGA	Single	Chr.1
5	CON_003_02564	Common regulatory factor	CCATCTCTTTCGCTATCTGC AGATCGAAGTAAGCTCGGAC	Single	Chr.1
6	CON_004_03320	Photolyase/blue light photoreceptor PHR2	AGCTCAACAAGGAGTGAG TGAACTGGCTGATGTTCG	Single	Ch.2
7	CON_031_04487	S-adenosyl-L-homocysteine hydrolase	GTCAATGACTCTGTCACC GATGGTAATGCGCTTGAC	Single	Chr.4
8	CON_002_01213	Mitochondrial small heat shock protein	CTCTCCATCACCCTAAAG GTGGAACAGAACACACTC	Single	Mitochondrial
9	CON_001_10630	S-adenosyl methionine synthetase	CCAATGTGATGAAGCTCC GGTGTACCTGAACCATTG	Single	Chr.1
10	CON_008_04131	Histone H2B1	AGAGAAGAAGCCTAAGGC TCACCAGGAAGTACAAGC	Single	Chr.1
11	CON_006_03927	Actin-depolymerizing factor 2	CAACCGAAAGCTATGAGG TGTAGGAAGGAAGGAAGC	Single	Chr.3
12	CON_001_11154	CONSTANS-like protein 1	ACCATGCTGGAGAAGACTGG AGACGGAAGTACGATGACGG	Single	Chr.5
13	CON_006_03877	Profilin	CTCCAGTTCTCAACGAATCC AGGTACAATCCAGTTGGAGC	Single	Chr.2
14	CON_002_01989	60S ribosomal protein L24	GCAAGAGGTACTTCCACAAC ACTTATCGCTTTCCACCACC	Single	Chr.2
15	CON_001_10272	Similar to ATP-citrate-lyase	CGAGATAACCACTACTCC CCTACACCACTTCAATGC	Single	Chr.1

16	CON_002_02239	Contains similarity to protein kinase~gene	TGGCCGTAACAATCATGG TGCTTCCTTGCATCACAG	Single	Chr.1
17	CON_001_10467	putative transcription factor BHLH6	CTTCATCTTCAAGCCCAC ACGAACACCTTCGAGAAG	Single	Chr.1
18	CON_002_00682	60S ribosomal protein L4/L1	ACATAGTCACCTTCGTCC CGTCAGATATGACCAGAG	unknown	Chr.3
19	CON_005_03604	Histone H3.3	TCCGAGTCGTATCAGAAG CAGTTTCCTGATGAGGAG	Single	Chr.4
20	CON_001_05388	Luminal binding protein (BiP)	AGGATTGATGCTCGCAACAG CTCATCTCACCTACCTATGC	Single	Chr.5
21	CON_001_06492	Cystein proteinase	CCTCTTGTTGATGACTGG CATGTCTGGTTCAGTGTG	Single	Chr.2
22	CON_004_03318	Outer membrane lipoprotein - like	ACATGGGAAAGTGGTACGAG CTCACATCATACCCCAAGTG	Single	Chr.5
23	CON_001_00045	14-3-3-like protein	CCATCTCTTTCGCTATCTGC AGATCGAAGTAAGCTCGGAC	Single	Chr.1
24	CON_001_10976	Nucleotide-binding protein	GCTAGATCTCATTGCTGC GACTGTAGGGTACTGAAC	Single	Chr.1
25	CON_002_02249	RAD23 protein	CATCCCTTGTCTCTACACC ACAAGCTGTGCAAGTCACAC	Single	Chr.1
26	CON_013_04353	Elongation factor 1-gamma	GGACCTTGGTTAATCAGC CCGATCACCAACATCTTC	unknown	Chr.1
27	CON_001_12763	E2, Ubiquitin-conjugating enzyme, putative	GATCGTAACGGAGGTGGTAG GCAGGCAATTGGATCCTTCC	Single	Chr.5
28	CON_003_02574	Beta-galactosidase 1 like protein	GGCCATACTTGAACCTTG GTTAGGTTAGTGCCACAG	unknown	Chr.3
29	CON_009_04160	Alcohol dehydrogenase	TGTGGAGAGTGTAGGTGAGG GTGGACATTCCGCAGCTTAG	Single	Chr.1
30	CON_001_12406	vacuolar H+-ATPase subunit E	CGAGTAGGAGTTCAACATCG GTGCCGCTTTGAAATCTGAC	Single	Chr. 4

sequencing technique was used to identify SNPs in different fiber related genes (Lu et al. 2005).

Approximately 7804 bp of *G. arboreum* DNA was sequenced and a total of 27 polymorphisms were identified (21 SNPs and 6 single base Indels). The calculated SNP frequency was one single nucleotide change every 371 bp, whilst, Indels occurred less frequently, about one every 1300 bp or the rate of variation per nucleotide (0.27%). In previous studies with other molecular markers genetic diversity among *G. arboreum* varieties has been assessed. Randomly amplified polymorphic DNA (RAPD) estimated 59% to 76% similarity (Kumar et al. 2008) and 47.05% to 98.73 (Rahman et al. 2008), and with SSRs 58% to 87% similarity (Liu et al. 2006), 52% to 98% polymorphism information content (PIC) (Guo et al. 2006) was observed. Commonality among all above studies is the detection of low genetic diversity among *G. arboreum* accessions.

In tetraploid cotton, a total of 94 SNPs including 36 single-base changes (38.3%) and 58 indels (61.7%) were identified in 16 fiber gene fragments with an average frequency of one SNP per 500 bp DNA which was lower than that in coding sequences of many other plant species (Lu et al. 2005). Another SNP study revealed the rate of variation per nucleotide was 0.35% between *G. hirsutum* and *G. barbadense* (one SNP every 286 bp) (Rong et al. 2004), which is a higher frequency as compared to frequency observed in this study. In the *FIFI* gene, regulating the fiber development in *G. barbadense*, three SNPs were reported while comparing with the corresponding gene in *G. arboreum* and *G. hirsutum* with a frequency 1SNP/270 bp interspecifically (Ahmad et al. 2007). In another investigation, one SNP per 77 bases in the six *R2R3-MYB* genes were reported in different cotton genomes (An et al. 2008).

In a total of 21 substitutions, 11 (52%) were transitions and 10 were transversions (Table 2). Such commonalties were also found in rice (61.8%) (Feltus et al. 2004) and citrus (52.7%) (Novelli et al. 2004). Alterations of this type could be attributed to the actions of 5-methylcytosine deamination (Feltus et al. 2004). Out of 21 substitutions 14 were present in ORF region, of which 9 (64%) were synonymous, and 7 were present in non ORF region. These values are comparable with many previous studies conducted in maize (72% synonymous, Ching et al. 2002). Our results explicate a comparable ratio of synonymous to nonsynonymous mutations in cotton (1.8) as compared to (2.6) in soybean (Zhu et al. 2003), and 1.7 in melon (Morales et al. 2004). The synonymous substitutions are frequently found in ORF regions as these are not detrimental to the plant (Morales et al. 2004). In the present study, majority of Indels (67%) were found in non ORF regions because these are not tolerable in ORF regions (Liston and Briedis, 1995).

SNPs can prove very effective in MAS if a SNP marker is found associated with target trait. Moreover SNPs are highly stable markers which may contribute directly to phenotype which can further be utilized by plant breeders for MAS to identify individual plants containing a combination of alleles of interest from large segregating populations (Batley and Edwards, 2007). The SNPs identified in this study can further be utilized for traits association and MAS.

Table 2. Thirty *G. arboreum* ESTs showing amplified size of the amplicons, the number of SNPs detected for each EST, and their sequence and position in the amplicon.

Sr #	Best Blast homology	Conserved region	Amplicons size	No. of SNPs/ Indels	SNP/ Indel location/ position
1	Histone H2B-3	350-650	300	2	+A (595) - C (634)
2	NADPH-cytochrome P450 reductase	450-550	100	0	-
3	Werewolf	201-570	400	1	- G (562)
4	putative ATP binding protein	30-150	120	2	T→C (30) A→T (32)
5	common regulatory factor	262-562	300	1	A→G (558)
6	Photolyase/blue light photoreceptor PHR2	20-200	180	0	-
7	S-adenosyl-L-homocysteine hydrolase	246-663	420	3	C→A (249) C→T (412) C→A (646)
8	Mitochondrial small heat shock protein	139-440	300	0	-
9	S-adenosyl methionine synthetase	193-332	120	1	G→A (229)
10	Histone H2B1	190-500	310	0	-
11	Actin-depolymerizing factor 2	300-500	200	0	-
12	CONSTANS-like protein 1	162-440	280	1	+A (178)
13	Profilin	35-540	500	0	-
14	60S ribosomal protein L24	134-441	300	1	C→G (89)
15	Similar to ATP-citrate-lyase	79-424	350	0	-
16	Contains similarity to protein kinase~gene	10-450	440	1	-A (12)
17	putative transcription factor BHLH6	111-296	200	1	-T (159)
18	60S ribosomal protein L4/L1	169-346	200	0	-
19	Histone H3.3	169-392	225	0	0
20	Luminal binding protein (BiP)	46-411	365	2	A→C (49) C→T (50)
21	Cystein proteinase	267-356	100	0	0
22	Outer membrane lipoprotein – like	40-428	400	0	0
23	14-3-3-like protein	275-560	300	0	-
24	Nucleotide-binding protein	148-230	200	1	G→C (204)
25	RAD23 protein	23-105	80	0	-
26	Elongation factor 1-gamma	848-1010	160	3	G→C (899) A→G (920) A→G (947)
27	E2, Ubiquitin-conjugating enzyme, putative	10-378	370	0	-
28	Beta-galactosidase-1	151-380	240	2	C→A (379) T→A (380)
29	Alcohol dehydrogenase	299-476	200	2	G→ T (431) A→G (462)
30	vacuolar H+-ATPase subunit E	70-214	144	3	G→A (137) G→A (153) A→G (217)

SNPs identification in polyploids have been simplified with the Illumina BeadArray technology coupled with GoldenGate assay without the need of a prior PCR amplification step *e.g.*, in polyploid wheat pure lines (Akhunov et al. 2009). About 89 and 84% of SNPs in tetraploid and hexaploid wheat, respectively, could be converted into successful genotyping assays. The Illumina BeadArray platform, represents an excellent tool for studying genetic architecture of complex traits, association mapping

and, with proper safeguards, evolutionary forces that shape the genetic diversity of polyploids like cotton as well (Akhunov et al. 2009).

In conclusion SNPs are an effective tool for whole genome survey and are potent markers to survey conserved regions where other markers may not prove very effective which will pave the way of developing dense genetic maps. Microarray based SNP genotyping can be a very effective tool but it is just in preliminary stages. These methods can improve the pace of genotyping in cotton.

Financial support: This research was funded by Higher Education Commission (HEC) through a project under the scheme of PYI award (2007-2010) and indigenous PhD scholarship scheme through a PhD student grant.

REFERENCES

AHMAD, Saghir; ZHANG, Tianzhen; ISLAM, Noor-Ul-Islam; SHAHEEN, Tayyaba and RAHMAN, Mehboob-Ur. Identifying genetic variation in *Gossypium* based on single nucleotide polymorphism. *Pakistan Journal of Botany*, August 2007, vol. 39, no. 4, p. 1245-1250.

AN, Chuanfu; SAHA, Sukumar; JENKINS, Johnie N.; MA, Din-Pow; SCHEFFLER, Brian E.; KOHEL, Russell J.; YU, J.Z. and STELLY, David M. Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome localization and linkage mapping. *TAG Theoretical and Applied Genetics*, May 2008, vol. 116, no. 7, p. 1015-1026. [CrossRef]

AKHUNOV, Eduard; NICOLET, Charles and DVORAK, Jan. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *TAG Theoretical and Applied Genetics,* May 2009, vol. 119, no. 3, p. 507-517. [CrossRef]

ARPAT, Aladdin B.; WAUGH, Mark; SULLIVAN, John P.; GONZALES, Michael; FRISCH, David; MAIN, Dorrie; WOOD, Todd; LESLIE, Anna; WING, Rod and WILKINS, Thea. Functional genomics of cell elongation in developing cotton fibers. *Plant Molecular Biology*, April 2004, vol. 54, no. 6, p. 911-929. [CrossRef]

AYEH, Kwadwo Owusu. Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs): Emerging molecular marker tools for improving agronomic traits in plant biotechnology. *African Journal of Biotechnology*, February 2008, vol. 7, no. 4, p. 331-341.

BARBAZUK, W. Brad; EMRICH, Scott and SCHNABLE, Patrick S. SNP mining from maize 454 EST sequences. *CSH Protocols,* July 2007, vol. 2007, no. 7. [CrossRef]

BATLEY, J. and EDWARDS, D. SNP applications in plants. In: ORAGUZIE, N.C.; RIKKERINK, E.H.A.; GARDINER, S.E. and DE SILVA, H.N., eds. *Association mapping in plants.* New York, Springer, 2007, p. 95-102.[CrossRef]

CHING, Ada; CALDWELL, Katherine S.; JUNG, Mark; DOLAN, Maurine; SMITH, Oscar S.; TINGEY, Scott; MORGANTE, Michele and RAFALSKI, Antoni J. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics*, October 2002, vol. 3, no. 19, p. 3-19. [CrossRef]

CHOI, Ik-Young; HYTEN, David L.; MATUKUMALLI, Lakshmi K.; SONG, Qijian; CHAKY, Julian M.; QUIGLEY, Charles V.; CHASE, Kevin; LARK, K. Gordon; REITER, Robert S.; YOON, Mun-Sup S.; HWANG, Eun-Young; YI, Seung In; YOUNG, Nevin D.; SHOEMAKER, Randy C.; VAN TASSELL, Curtis P.; SPECHT, James E. and CREGAN, Perry B. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics*, May 2007, vol. 176, no. 1, p. 685-696. [CrossRef]

CLARK, Richard M.; SCHWEIKERT, Gabriele; TOOMAJIAN, Christopher; OSSOWSKI, Stephan; ZELLER, Georg; SHINN, Paul; WARTHMANN, Norman; HU, Tina T.; FU, Glenn; HINDS, David A.; CHEN, Huaming; FRAZER, Kelly A.; HUSON, Daniel H.; SCHÖLKOPF, Bernhard; NORDBORG, Magnus; RÄTSCH, Gunnar; ECKER, Joseph R. and WEIGEL, Detlef. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, July 2007, vol. 317, no. 5836, p. 338-342. [CrossRef]

FELTUS, F. Alex; WAN, Jun; SCHULZE, Stefan R.; ESTILL, James C.; JIANG, Ning and PATERSON, Andrew H. An SNP resource for rice genetics and breeding based on

subspecies *Indica* and *Japonica* genome alignments. *Genome Research*, September 2004, vol. 14, no. 9, p. 1812-1819. [CrossRef]

GUO, Wang-Zhen; ZHOU, Bao-Liang; YANG, Lu-Ming; WANG, Wei and ZHANG, Tian-Zhen. Genetic diversity of landraces in *Gossypium arboreum* L. race *sinense* assessed with simple sequence repeat markers. *Journal of Integrative Plant Biology,* September 2006, vol. 48, no. 9, p. 1008-1017. [CrossRef]

IQBAL, M.J.; AZIZ, N.; SAEED, N.A.; ZAFAR, Y. and MALIK, K.A. Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *TAG Theoretical and Applied Genetics*, January 1997, vol. 94, no. 1, p. 139-144. [CrossRef]

KOORNNEEF, Maarten; ALONSO-BLANCO, Carlos and VREUGDENHIL, Dick. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology*, June 2004, vol. 55, p. 141-172. [CrossRef]

KUMAR, Mukesh; KUMAR, Rajiv and CHAUDHARY, Lakshmi. Genetic diversity evaluation of some *G. arboreum* (diploid cotton) varieties by RAPD analysis. *International Journal of Biotechnology & Biochemistry,* October 2008, vol. 4, no. 1, p. 23-31.

LISTON, Peter and BRIEDIS, Dalius J. Ribosomal frame shifting during translation of measles virus P protein mRNA is capable of directing synthesis of a unique protein. *Journal of Virology*, November 1995, vol. 69, no. 11, p. 6742-6750.

LIU, Diqui; GUO, Xiaoping; LIN, Zhongxu; NIE, Yichun and ZHANG, Xianlong. Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. *Genetic Resources and Crop Evolution,* September 2006, vol. 53, no. 6, p. 1145-1152. [CrossRef]

LU, Y.; CURTISS, J.; ZHANG, J.; PERCY, R.G. and CANTRELL, R.G. Discovery of single nucleotide polymorphisms in selected fiber genes in cultivated tetraploid cotton. *National Cotton Council Beltwide Cotton Conference*, 2005, p. 946.

MORALES, M.; ROIG, E.; MONFORTE, A.J.; ARÚS, P. and GARCIA-MAS, J. Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). *Genome*, April 2004, vol. 47, no. 2, p. 352-360. [CrossRef]

MOULHERAT, Christophe; TENGBERG, Margareta; HAQUET, Jérôme-F. and MILLE, Benoît. First evidence of cotton at Neolithic Mehrgarh, Pakistan: analysis of mineralized fibres from a copper bead. *Journal of Archaeological Science*, December 2002, vol. 29, no. 12, p. 1393-1401. [CrossRef]

NASU, Shinobu; SUZUKI, Junko; OHTA, Rieko; HASEGAWA, Kana; YUI, Rika; KITAZAWA, Noriyuki; MONNA, Lisa and MINOBE, Yuzo. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research*, 2002, vol. 9, no. 5, p. 163-171. [CrossRef]

NOVELLI, Valdenice Moreira; TAKITA, Marco Aurélio and MACHADO, Marcos Antonio. Identification and analysis of single nucleotide polymorphisms (SNPs) *in citrus*. *Euphytica*, March 2004, vol. 138, no. 3, p. 227-237. [CrossRef]

RAHMAN, Mehboob-Ur; YASMIN, Tahira; TABASSUM, Nabila; ULLAH, Ihsan; ASIF, Muhammad and ZAFAR, Yusuf. Studying the extent of genetic diversity among *Gossypium arboreum* L., genotypes/ cultivars using DNA fingerprinting. *Genetic Resources and Crop Evolution,* May 2008, vol. 55, no. 3, p. 331-339. [CrossRef]

RAHMAN, Mehboob-ur; ZAFAR, Yusuf and PATERSON, Andrew H. Gossypium DNA markers: types, numbers and uses. In: PATERSON, Andrew H. ed. *Genetics and Genomics of Cotton*, New York, Springer, 2009, vol. 3, p. 1-39.

RONG, Junkang; ABBEY, Colette; BOWERS, John E.; BRUBAKER, Curt L.; CHANG, Charlene; CHEE, Peng W.; DELMONTE, Terrye A.; DING, Xiaoling; GARZA, Juan J.; MARLER, Barry S.; PARK, Chan-hwa; PIERCE, Gary J.; RAINEY, Katy M.; RASTOGI, Vipin K.; SCHULZE, Stefan R.; TROLINDER, Norma L.; WENDEL, Jonathan F.; WILKINS, Thea A.; WILLIAMS-COPLIN, T. Dawn; WING, Rod A.; WRIGHT, Robert J.; ZHAO, Xinping; ZHU, Linghua and PATERSON, Andrew H. A 3347-Locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics*, January 2004, vol. 166, no. 1, p. 389-417. [CrossRef]

SEMAGN, Kassa; BJØRNSTAD, Åsmund; SKINNES, Helge; MARØY, Anne Guri; TARKEGNE, Yalew and WILLIAM, Manilal. Distribution of DArT, AFLP and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. *Genome*, 2006, vol. 49, no. 5, p. 545-555. [CrossRef]

SHAHEEN, Tayyaba; RAHMAN, Mehboob and ZAFAR, Yusuf. Chloroplast RPS8 gene of cotton reveals the conserved nature throughout plant taxa. *Pakistan Journal of Botany*, December 2006, vol. 38, no. 5, p. 1467-1476.

THE INTERNATIONAL HAPMAP CONSORTIUM. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, October 2007, vol. 449, p. 851-862. [CrossRef]

UDALL, Joshua A.; SWANSON, Joshua M.; HALLER, Karl; RAPP, Ryan A.; SPARKS, Michael E.; HATFIELD, Jamie; YU, Yeisoo; WU, Yingru; DOWD, Caitriona; ARPAT, Aladdin B.; SICKLER, Brad A.; WILKINS, Thea A.; GUO, Jin Ying; CHEN, Xiao Ya; SCHEFFLER, Jodi; TALIERCIO, Earl; TURLEY, Ricky; McFADDEN, Helen; PAYTON, Paxton; KLUEVA, Natalya; ALLEN, Randell; ZHANG, Deshui; HAIGLER, Candace; WILKERSON, Curtis; SUO, Jinfeng; SCHULZE, Stefan R.; PIERCE, Margaret L.; ESSENBERG, Margaret; KIM, Hye Ran; LLEWELLYN, Danny J.; DENNIS, Elizabeth S.; KUDRNA, David; WING, Rod; PATERSON, Andrew H.; SODERLUND, Cari and WENDEL, Jonathan F. A global assembly of cotton ESTs. *Genome Research*, March 2006, vol. 16, no. 3, p. 441-450. [CrossRef]

WANG, Rui-Sheng; WU, Ling-Yun; LI, Zhen-Ping and ZHANG, Xiang-Sun. Haplotype reconstruction from SNP fragments by minimum error correction. *Bioinformatics*, February 2005, vol. 21, no. 10, p. 2456-2462. [CrossRef]

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VAN TASSELL, C.P.; MATUKUMALLI, L.K.; GRIMM, D.R.; HYATT, S.M.; FICKUS, E.W.; YOUNG, N.D. and CREGAN, P.B. Single-Nucleotide polymorphism in soybean. *Genetics*, March 2003, vol. 163, no. 3, p. 1123-1134.

How to cite this article:

SHAHEEN, T.; ZAFAR, Y. and RAHMAN, M. Detection of single nucleotide polymorphisms in the conserved ESTs regions of *Gossypium arboreum*. *Electronic Journal of Biotechnology*, September 2010, vol. 13, no. 5. http://dx.doi.org/10.2225/vol13-issue5-fulltext-3