The *Journal of Cotton Science 8:112–123 (2004)*
http://journal.cotton.org, © The Cotton Foundation 2004

MOLECULAR BIOLOGY AND PHYSIOLOGY

EST-SSR: A New Class of Genetic Markers in Cotton

Samina N. Qureshi, Sukumar Saha*, Ramesh V. Kantety, and J. N. Jenkins

ABSTRACT

Recent advances in genomic technologies have generated a large number of expressed sequence tags (ESTs) in cotton. Many of these ESTs are available in public databases, which offer an opportunity to identify simple sequence repeats (SSR) in ESTs by data mining. These sequences may provide an estimate of diversity in the expressed portion of the genome and may be useful for comparative mapping, for tagging important traits of interest, and for additional map-based cloning of important genes. One hundred and thirty-three SSR-containing ESTs (EST-SSRs) were identified by analyzing 9,948 sequences belonging to *Gossypium hirsutum* L. in GenBank. The EST-SSR sequences and the related EST sequences without SSRs were clustered to reduce redundancy and to develop consensus sequences. Primers were designed for 84 of these EST-SSRs and were tested for their ability to amplify and detect diversity among three lines of *G. hirsutum* and one line of *G. barbadense* L. An average of three amplicons was obtained per primer pair. The intraspecies polymorphism rate among the *G. hirsutum* cotton cultivars was 26% and interspecific polymorphism between *G. hirsutum* and *G. barbadense* was 52%. The presence of SSRs in the EST-SSR markers was confirmed by cloning and sequencing of the amplified products of randomly selected primer pairs from four different lines. To explore the potential use of the EST-SSR loci for comparative mapping, these sequences were compared against different plant species using BLAST assuming an e-value of $1E^{-10}$ or less as a significant sequence similarity. About 74% of the EST-SSRs were from fiber-related tissues in *G. hirsutum*, whereas 26% were from other tissues, such as cotton bolls and cottonseed. Fifty-five percent of these EST-SSR sequences matched sequences in *G. arboretum* L., and 19% of the sequences showed considerable sequence similarity with sequences in the *Arabidopsis thaliana* (L.) Heynh. genome. In this manuscript, information about the primer sequence, repeat motif, and the degree of polymorphism of cotton EST-SSR markers is reported. A cost-effective strategy to develop EST-SSR markers by exploiting EST databases is demonstrated for the first time in cotton.

Cotton (*Gossypium* spp.) is the leading textile fiber in the world and one of the most important oilseed crops. Much of cotton's production efficiency is attributable solely to genetic improvement through conventional breeding methods. Cotton has a narrow genetic base, so there is little variation available for use in the development of new, higher yielding cotton cultivars. One of the major limitations to the application of genomic technology in cotton improvement is the paucity of informative DNA markers.

DNA markers linked to agronomic traits could increase the efficiency of breeding by significantly reducing the number of backcross generations required and by decreasing expensive, tedious, and subjective phenotypic selection. The two major limiting factors in the use of molecular markers for both quantitative trait locus (QTL) analysis and marker-assisted selection programs in cotton are: 1) the limited number of suitable markers available in the public sector, and 2) the lack of knowledge of how these markers are associated with economically important QTLs. Selection of suitable markers is one of the key factors for the success of any molecular breeding program. DNA markers must be based on a simple and efficient detection system, be highly polymorphic, and be distributed across the genome at random. A saturated genetic map of suitable DNA markers could expedite genetic improvement in cotton.

S. N. Qureshi, Department of Plant and Soil Science, P.O. Box 9555, Mississippi State University, Mississippi State, MS 39762; S. Saha and J. N. Jenkins, USDA-ARS, Crop Science Research Laboratory, P.O. Box 5367, Mississippi State, MS 39762; Ramesh V. Kantety, Department of Crop & Soil Sciences, 726 Bradfield Hall, Cornell University, Ithaca, NY 14853
* Corresponding author: SSaha@msa-msstate.ars.usda.gov

There are numerous published reports on cotton genetic mapping using restriction fragment length polymorphisms (RFLPs) (Brubaker et al., 1999; Jiang et al., 2000; Paterson et al., 2003; Reinisch et al., 1994; 1999; Saha et al., 1998; Shappley et al., 1996, 1998; Ulloa and Meredith, 2000; Wright et al., 1998), but the RFLP technique is labor intensive, time consuming, and expensive. Scientists from different countries initiated the International Cotton Genome Initiative (ICGI) to coordinate future cotton genomics research (Brubaker et al., 2000). ICGI has stressed the need for more portable, publicly available, PCR-based framework markers (e.g. SSRs) to expedite genomic research in cotton. Given that the genome size of cotton is 2,200 Mb/1C (Arumuganathan and Earle, 1991) with an approximate total genetic map length of 5,000 cM (Paterson and Smith, 1999; Reinisch et al., 1994), several hundred markers are needed to achieve a saturated genetic map of cotton.

In many crops, simple sequence repeats (SSRs) are considered to be one of the markers of choice for genome mapping because they are 1) PCR-based, 2) usually co-dominant, 3) usually multiallelic and hyper-variable, and 4) randomly dispersed throughout the genome. Microsatellite variation is thought to be due to slippage of the DNA polymerase during replication or unequal crossing over resulting in differences in the copy number of the core nucleotide sequence (Yu and Kohel, 1999).

More than 500 microsatellite clones, containing mostly (GA)n repeats, were first identified in cotton at the Brookhaven National Laboratory. The primer-pairs, designated BNL, for the amplification of these loci were made available to the cotton research community. An additional 150 (GA)n repeat loci, designated CM, were isolated at Texas A&M University (Connell et al., 1998; Reddy et al., 2000). An additional 300 SSR markers (JESPR) were identified using SSR-enriched genomic libraries (Reddy et al., 2001). Liu et al. (2000a) reported the chromosomal assignment of several BNL microsatellite markers using cytogenetic aneuploid stocks. Several papers have been published on the use of both JESPR and BNL SSR markers in genetic mapping of cotton. Zhang et al. (2003) identified SSR markers associated with a major QTL for fiber strength. These scientists also developed a molecular linkage map of 624 marker loci (510 SSRs and 114 RAPDs) into 43 linkage groups covering 3,314 cM (Zhang et al., 2002) and identified SSR markers linked to a fertility restorer gene in cotton (Liu et al., 2003). A molecular linkage map of important fiber traits using SSR and RAPD markers based on chromosome specific recombinant (F_2) lines was also developed (Ren et al., 2002). Our research group has used many of these SSRs in genetic studies of cotton (Gutiérrez et al., 2002; Liu et al., 2000a; 2000b).

Several thousand markers will be required to develop a high-resolution saturated genetic map of cotton. In the past, SSRs were either developed from SSR-enriched libraries or identified by surveys of genomic sequence databases (Akkaya et al., 1992; Bell and Ecker, 1994; Condit and Hubbell, 1991; Morgante and Oliver, 1993; Wu and Tanksley, 1993). A new alternative source of microsatellites comes from data mining sequence information from the publicly available EST databases of important crops (Kantety et al., 2002; Smulders et al., 1997).

Several labs, including the NSF-funded "Cotton Genome Center" at Univeristy of California-Davis and Clemson University, initiated large-scale cDNA analyses of the cotton genome, which have provided an opportunity to use partially characterized cDNA clones for identifying expressed sequence tags containing simple sequence repeats (EST-SSRs). Cardle et al. (2000) reported that the average frequency of SSRs was one every 6.04Kb in genomic DNA and one in every 14 Kb for ESTs in *A. thaliana*. With the considerable potential for comparative mapping that EST-SSRs offer (Eujayl et al., 2002; Herron et al., 1998; Kantety et al., 2002; Scott et al., 2000; Sorrells 2000a; 2000b), we decided to develop these markers for cotton.

To date, information on the presence of tandem repeats in the expressed sequences is available for several important plant species, including *A. thaliana* (Delseny et al., 1999), rice (Sasaki et al., 1994; Yamamoto and Sasaki, 1997), oil palm (Billotte et al., 2001) and other grasses (Kantety et al., 2002). Analysis of 739,258 ESTs from the publicly available databases of the major monocot and dicot crop species, including cotton, revealed that SSRs were present in the ESTs at a frequency of 3.4% and 1.7% in the sequences of monocots and dicots, respectively (Saha et al., 2003). There are some studies in which SSRs were derived from public databases for use in plants, for example in rice (Cho et al., 2000), and grape (Scott et al., 2000; Effie et al., 2000), but there is no information on EST-SSR markers in cotton. Saha et al. (2003) showed that SSR primers, derived from cotton genomic microsatellite libraries, could

be used in differential screening of cDNAs from tetraploid cotton species, which indicates that some of the genomic SSR primer-pairs were actually derived from coding sequences.

Using expressed sequences of *G. hirsutum* from GenBank, we report 84 primer pairs designed for the amplification of EST-SSR markers in cotton. This method of marker development by data mining is cost effective and complementary to other ongoing large scale sequencing projects in cotton. This study also provided an opportunity to investigate the type and distribution of repeat motifs in the expressed sequences of cotton and the genetic diversity in the expressed portion of cotton genome.

MATERIALS AND METHODS

Identification of EST-SSR and primer design. The EST-SSRs in cotton were identified according to the method of Kantety et al. (2002). A total of 9,948 EST sequences and 190 partial or full-length cDNA sequences of known genes that belong to *G. hirsutum* were downloaded from GenBank (National Center for Biotechnological Information; http://www.ncbi.nlm.nih.gov/Genbank/index.html). One hundred thirty-three sequences, containing at least four di-, tri, tetra-, penta- or hexanucleotide repeats, were identified using Perl script (Kantety et al., 2002).

Primers were designed for the flanking regions of the SSR using a web-based software, "Primer3" (Rozen and Skaletsky, 1998), and based on the criteria of 50% GC content, a minimum melting temperature of 50º C, and absence of secondary structure. Primers ranged from 18–27 nucleotides in length and amplified products of 100–400 bp. The primers were synthesized by Sigma-Genosys (Woodlands, TX). When possible, primers were designed within the 5′ or 3′ untranslated region (UTR)(or near to the start or stop codon within coding DNA) closest to the repeat motif.

Isolation and amplification of genomic DNA. Cotton genomic DNA was isolated from young leaves using a cetyl trimethyl ammonium bromide-based extraction method (Altaf-Khan et al., 1997). Polymerase chain reactions (PCR) were conducted in a total volume of 10 µL, using 50 ng of cotton DNA, 1X Perkin-Elmer PCR Buffer II (10mM Tris-HCl, 50mM KCl, pH 8.3), 2.5-3.5 mM $MgCl_2$, 0.2 mM deoxy-nucleotide triphophates, 0.15 mM of each primer, and 0.4 units of AmpliTaq Gold (Perkin-Elmer Applied Biosystems; Foster City, CA). The amplification profile consisted of an initial period of denaturation, and AmpliTaq Gold activation at 94ºC for 7 min, followed by 9 cycles (step-1) of 94ºC for 15 s, 65ºC for 30 s and 72ºC for 60 s with a decrease in the annealing temperature by 1ºC in each cycle from 65ºC to 56ºC, and then 40 cycles (step-2) of 94ºC for 15 s, 55ºC for 30 s and 72ºC for 60 s. The PCR amplifications were followed by incubation at 72ºC for 7 min.

Electrophoresis of EST-SSR markers. A horizontal Metaphor agarose gel electrophoresis system (Cambrex; North Brunswick, NJ), capable of resolving two base-pair polymorphisms was initially used with unlabeled primers to evaluate amplification product polymorphisms. Samples were loaded on a horizontal gel containing 4% 1:1 agarose:Metaphor agarose in 0.5X TBE buffer (45 mM Tris- borate, 1 mM EDTA, pH 8) chilled to 4°C. Gels were stained with ethidium bromide (0.25ng/µL) and visualized under UV light.

EST-SSR primers fluorescently labeled (with Hex, FAM or NED fluorochromes) were also used. Multimix PCR products from these fluorescently labeled primers were separated and markers were analyzed by using ABI PRISM 310 Genetic Analyzer (Perkin-Elmer Applied Biosystems; Foster City, CA) following the manufacturer's protocol. The PCR products were diluted with water 1:20 for HEX and FAM, and 1:10 for NED-labeled primers. Three (µL) of diluted solution and 0.2 µL of ROX dye (an internal size standard) were added to 10 (µL) of formamide, denatured at 95ºC for 5 min, kept at 4ºC for 5 min, and loaded into an ABI PRISM 310 Genetic Analyzer (Perkin-Elmer Applied Biosystems). The sizes of the amplified bands were analyzed using the accompanying Gene Scan software.

Sequencing of PCR products. To confirm the presence of the microsatellite repeats in the amplified fragments, 16 SSR products were amplified from four cotton lines amplified with four randomly selected unlabeled primer pairs. The amplified products were electrophoresed on a 2% agarose gel and the bands were eluted from the gel and purified on a column for cloning. Cloning was performed using the PCR4-TOPO TA Cloning Kit for Sequencing (Invitrogen Corporation; Carlsbad, CA) following the manufacturer's instructions. Sequencing of these clones was conducted using the BigDye terminator cycle sequencing system (Perkin-Elmer Applied Biosystems) and the M13-forward or M13-reverse

primers. Electrophoretic separation of sequencing products was performed on an ABI PRISM 3700 DNA Analyzer 96-capillary automated sequencer (Perkin-Elmer Applied Biosystems).

Sequence analysis: Contigs of double stranded DNA sequences from each clone, identification of overlaps, redundancy among clones, and vector sequence trimming were performed using Sequencher version 4.0.5 (Gene Codes; Ann Arbor, MI). Sequences of the EST-SSR products from different genotypes were aligned using ClustalW version1.8 (European Bioinformatics Institute; http://www.ebi.ac.uk/clustalw/index.html). To explore the potential utility of these microsatellite loci in genomic comparisons with *A. thaliana* and other plants, the sequences of these EST-SSRs were subjected to a basic local alignment search (BLAST) tool against the Arabidopsis database (TAIR, 2004) and other GenBank sequences assuming an e- value $1E^{-10}$ or less as a significant homology (Altschul et al., 1990).

Genetic diversity of polymorphic EST-SSR markers in cotton. The newly designed 84 EST-SSR primer pairs were first screened on a set of four diverse cotton genotypes (TM-1, HS-46, and MARCABUCAG8US-1-88 from *G. hirsutum*, and PIMA 3-79 from *G. barbadense*) for detecting polymorphism at both inter- and intra-specific level. TM-1 is considered as the genetic standard line in *G. hirsutum* and has been used in mapping many important traits. PIMA 3-79 is a double haploid line. On the basis of pedigree records, HS46 and MARCABUCAG8US-1-88 are distantly related (Calhoun et al., 1997), and have been used to produce intra-specific genetic maps for Upland cotton (Shappley et al., 1998). A few selected EST-SSR primers were also screened against one line each of diploid *G. arboretum* (A_2) and diploid *G. raimondii* Ulbr. (D_5)

RESULTS AND DISCUSSION

Data mining for SSR markers. We searched 9,948 EST sequences and 190 partial or full-length mRNA sequences from *G. hirsutum* for the presence of microsatellites. A total of 219 sequences, containing at least an SSR with a length of 18 bases, were identified and clustered using StackPACK 2.0 (Electric Genetics; Reston, VA)(Miller et al., 1999). After the first identification of the SSR containing ESTs, we searched for ESTs that overlapped with those that carried SSRs. The EST sequences were subjected to masking of low-complexity sequences, as well as other repeat sequences, using the RepeatMasker program (Institute for Systems Biology; http://www.repeatmasker.org) as per the method of Kantety et al. (2002). The EST-SSR sequences and the related EST sequences without SSR were clustered to reduce redundancy and to develop consensus sequences (available online at http://msa.ars.usda.gov/ms/msstate/csrl/saha.htm). A total of 133, non-redundant sequences were obtained from the clustering procedure and were used for designing primer pairs. Fifty-five (41%) of these sequences did not have sufficient flanking sequences to permit design of specific PCR primers. Seventy-eight non-redundant sequences were used for designing primers with 18-24 bases and approximately 50% GC content. A total of 84 primer pairs were designed, including six sequences for which two sets of primer pairs were designed, since two different repeat motifs were identified within the same sequence (Table 1).

The EST-SSRs contained diverse types of repeat motifs. About 18, 34, and 9% of the sequences contained di-nucleotide, tri-nucleotide, and tetra- or penta-nucleotide repeats, respectively. The majority of EST-SSR sequences contained tri-nucleotide repeats of pure structure (34%) followed by di-nucleotide pure repeat motifs (Table 1). Similarly, significant numbers of interrupted, imperfect, and compound repeats were also identified. Recent studies indicate that the predominant SSR motif in *G. hirsutum* was $(GA/CT)_n$ followed by $(AAG)_n$, while $(AAG)_n$ was the most common repeat in *G. arboreum* (Saha et al., 2003). The most common repeat in graminaceous species was $(GA/CT)_n$ (Kantety et al., 2002). Gupta et al. (1996) reported the most abundant tri-nucleotide repeat motifs in plants were $(AAG)_n$ and $(AAT)_n$ with some species-specific variations.

We found very low levels of redundancy in EST-SSR with other SSR sequences available in public databases, including BNL and JESPR sequences (Reddy et al., 2001), which indicates the possible presence of a large number of SSRs in the cotton genome. Only five EST-SSR sequences showed homology with the JESPER sequences in the BLAST searches. The sequences of EST-SSR primers MGHES-11, 41, 45, 51, and 54 were partly similar to JESPR-240, 139, 17, 173, and 278, respectively. BNL-243 and 2646 were similar to the MGHES-7 and MGHES-34, respectively. These results were confirmed by sequence comparison using Sequencher software.

Table 1. EST-SSR primer pairs from *Gossypium hirsutum* and their polymorphism within and across different cotton species

EST-SSR#	GenBank accession(s)	Repeat motif	Forward primer (5' - 3')	Reverse primer (5' - 3')	Polymorphism	
					G. hirsutum	Interspecies
MGHES-1a	AI727289.1, AI730782.1, AW587458.1	$(TCA)_4$, $(CAC)_6$	GACATGAGGAAGCAGTTGAAAGG	GCATCACCTGAACAACATCCACC	M	M
MGHES-1b	AI727289.1, AI730782.1, AW587458.1	$(TCA)_4$, $(CAC)_6$	ACAGGGCAGCGTTTAATTTG	CACCTGAACAACATCCACCA	M	M
MGHES-2	AW187121.1, AW187322.1, AW187585.1, AW186799.2	$(CT)_{10}$, $(AG)_4$	TCTCTCAAAATCTCAAACCCAGA	GCTTAGGGCAAACCACTGAA	M	P
MGHES-3	AW187258.1, AW187678.1, AW187817.1, AW187855.1	$(TC)_{11}$	TCTCTCAAAATCTCAAACCCAGA	GCTTAGGGCAAACCACTGAA	M	M
MGHES-4	AI727595.1, AI730888.1	$(ACC)_8$, $(CT)_3$	GCCGGTTCCTTTGACCAC	CCCGCATCGTCATTAACTTT	M	M
MGHES-5	AI726568.1, AI727153.1, AI729937.1, AI730089.1	$(TC)_{10}$	ATTTGCGGGTGGAGAAGAC	TGGCGATTGAACAACAAAGA	M	M
MGHES-6	AW187078.1, AW187113.1	$(CCA)_7$	TCGCTTGACTTTCCATTTCC	AACCCTCGGGATTATCGTCT	P	P
MGHES-7	AI054962.1	$(CT)_{11}$	CCTTCTTCAACACCAATCTCc	TGCATTTCTGCTGAGTACCG	M	M
MGHES-8	AW187539.1, AW187811.1	$(ATT)_{11}$	CAAGCGATTGTTTCTCATCC	CGTCATCATAAACCAACGTGC	M	P
MGHES-9	AW187644.1, AW187666.1	$(CT/G)_n$	TCGAGAAATTTGGCTTCACC	GTGTTGGATGTAGCGGGAGT	M	M
MGHES-10	AW186906.1, AW186907.1, AW187848.1	$(CT)_{10}$	CTGATTCCACTCTCAAAACCAC	CTACTTTCCATCAGATCCCC	M	P
MGHES-11a	AI731348.1, AI731677.1	$(AG)_{12}$, $(GA)_3$, $(TTA)_3$	CGACTCCTCGACTCGCTATT	GCGCCACATACATCTCTCC	P	P
MGHES-11b	AI731348.1, AI731677.1	(AG)12, (TTA)3, (CGG)3	CATCATGGCTTTCCGTTTTT	CCAGGATTGGTAAACCCGTA	P	P
MGHES-12	AI728839.1, AI728842.1, AI731175.1	$(AC)_{10}$	GTTTCCAGGACAGAAAGGTGTC	GAGTTCCCAGTTACAGAGGC	M	P
MGHES-13	M16936.1	$(GA)_3$	CAGGGGAGCCATTGTTAGAA	CAGGGGTCCTGTGTTTCAGT	P	P
MGHES-14	AI727582.1, AI727584.1, AI730137.1, AI730170.1, AI731240.1	$(GAC)_{11}$, $(AGA)_3$	GAGGAGGCTGTGGTTGAAGA	ATGGTGACCCTGCTTACACC	P	P
MGHES-15	AI725775.1, AI727033.1, AI729351.1, AI731079.1	$(AAAC)_5$	AATCGAAGCGTTTCATCACC	CGAAGATCTTGGACAGACGA	M	P
MGHES-16	AI727680.1, AI731036.1	$(CT)_{10}$, $(TCT)_4$	ACCCCAATACAACCCCATTT	GCAGAGAAAAGGGACAGAGG	P	P
MGHES-17	AI727685.1, AI729986.1	$(CAT)_6$, $(TTC)_4$	AACCCTTCTTTTCCCCCTTT	TCTTCACCGATGCCATTGTA	M	M
MGHES-18	AI725573.1, AI729781.1	$(AT)_{13}$	GCCATCAATTGGTGAAGCAT	ATGCCTCGGTGAGAAAATTG	M	P
MGHES-19	AI054625.1, AI726167.1, AI728404.1	$(AGA)_5$, $(AGC)_3$	CACCGATCAGATAGCAGCAG	TGGCGTCTCAGAGATGAAGT	M	M
MGHES-20	AI054814.1, AI055346.1	$(CCA)_9$	CGAACCCTAGCTTTCAGTCG	AGTCGACGGCTTCAGATTGT	M	M
MGHES-21	AW187936.1	$(GA)_{14}$	TTTTTCGGGCTATGCTTTTG	GGGGTTGACATGTCCTATGC	P	P
MGHES-22	M73072.1, M16905.1	$(AGA)_7$, $(GAA)_3$, $(AGG)_4$	GGAACAGAGGCAACTGAGGA	TCGAAGGCACAGAGAAGGTT	M	P
MGHES-23	M88324.1, M88323.1	$(TA)_n$	AGCCGCATCACTTTTTGCTA	TCAAAAACAGAAGCACCAAGG	M	M
MGHES-24	AW587511.1	$(GCA)_6$, $(CAA)_3$	CGCAACAACTGATGCAACTC	AACCGATACCTCCGCTTCTT	M	M
MGHES-25	AW187493.2	$(TTA)_7$, $(TTC)_3$	TGAGGAACCAAGCAAAAACC	CTTGGGCAACTTCCAAGGTA	M	M
MGHES-26	AW187338.2	$(TTGT)_5$, $(GC)_3$	AAGGGGAGGTTTTGTGTAAGG	GACAAGAACCAGCTCCCAAA	M	M
MGHES-27	AW187804.1	$(TCT)_7$, $(AAC)_4$, $(TTC)_3$	TAATGGCGGCTAAACCTTCA	GGGGTTGCCTTCTTCTGTAG	M	P
MGHES-28	AW187747.1	$(CAG)_9$, $(TCA)_3$, $(ACA)_3$	CCTGCAAACGCTATTGATCC	CCCAGACTGGTGATGATGAA	M	M
MGHES-29	AW187478.1	$(TCA)_6$, $(TCT)_3$	TTTCCATTTTCTCCTGCTTCA	TCAGCTCATGGGAGCAAATA	M	P
MGHES-30a	AW187330.1	$(CT)_{13}$	GAATTGAATCATTTTCTCTGCAA	GGTCCACCCTTTCCTTTCTC	M	P
MGHES-30b	AW187330.1	$(CG/GC/GT)_n$, $(CGG)_4$, $(GCG)_3$	GGAGAAAGGAAAGGGTGGAC	GCAACACCAGAGAACACACG	M	P
MGHES-31	AW186938.1	$(CAT)_9$	AAGTTAGCGGCTTCTTGTGG	GGGTCAGAACTGGACAAGGA	M	P
MGHES-32	AW186919.1	$(TCT)_7$, $(AAG)_{25}(CTA)_4$	CGTCGCTTCCTTTGCTAAAC	GTCGGGTTAATTGCAAATCG	P	P
MGHES-33	AW186845.1	$(CT)_{8}(TC)_{8}(TCTCTT)_4$	TTTTTGGGCTTTCTTTTCTCTC	CCAATTACGCATGTTCAACG	P	M
MGHES-34	AW186834.1	$(CCA)_3$, $(CAC)_4$	TTCCTCCCTTCCTTCTCTCC	TTCCATTGTCATCGTTTCCA	M	M
MGHES-35	AI731986.1	$(TTCTTT)_3$, $(TG)_{14}$	TCGAACGGCTCGTTAAATCT	CAGCAAAGAGTGGTTCTCTGG	P	P
MGHES-36	AI731904.1	$(CTT)_8$, $(TCA)_4$	CGACAGCGAGTGTGAAACAT	GGTGGGAAAAACGCAAACT	M	M
MGHES-37	AI731713.1	$(AAG)_{11}$	GAAAATCCCAATTTCCACGA	TCATGATACCAATTTTGCTTCG	M	P
MGHES-38a	AI731600.1	$(ACC)_5$, $(CT)_4$	TTTGTTTTTGCAGCCCTTCT	GGGTCAAAGGAAGTGAACCA	M	P
MGHES-38b	AI731600.1	$(CAG)_{12}$, $(CAT)_5$, $(CCA)_3$	CCACCACCACTCTCCTTCTC	TCTTCTGCTGGTGCAATGAC	M	M
MGHES-39	AI730750.1	$(GCC)_6$, $(TCT)_3$	CACAACCCTAGCTAATCCCAAT	ACGTCGCTTCTGAGATTCGT	M	M
MGHES-40	AI730490.1	$(TTC)_6$	CGCGTTCCCAACTTATTTGT	GGTGCTCCCGGATTAGATTT	P	P

Table 1. continued

EST-SSR#	GenBank accession(s)	Repeat motif	Forward primer (5' - 3')	Reverse primer (5' - 3')	Polymorphism	
					G. hirsutum	Interspecies
MGHES-41	AI730273.1	$(CCA)_8$	GAAGGAGGGCGAAAAACATA	TTGGAGATTACGCGACCTTC	M	P
MGHES-42	AI730036.1	$(AGA)_6/(GT)_n$	GAAGAGCAGGTGGACCTTGA	CCCCTCATTAGCATCAGAGC	M	P
MGHES-43a	AI729689.1	$(GT)_{10}, (TCT)_3$	AGGACTTGTCCACGTGCTTC	TTTGATTCTTTTCGGCTGCT	M	M
MGHES-43b	AI729689.1	$(A)_n$	AAGCGTTCACACCATGACTTC	GGGAATCTCCGGGTTAAA	P	P
MGHES-44	AI729661.1	$(GAA)_{10}, (TCA)_3, (CAT)_3$	ACCACTTGGGATTGGTTCAA	GAGGCCACCACATATCGTTT	P	P
MGHES-45	AI728339.1	$(GTT)_6$	TGTGTCGTTTTCTGGGACCT	TATCACCGACATCTCCACCA	P	P
MGHES-46	AI727625.1	$(CGT)_9$	CGATTTCCATTCCACACCTC	GCATTGCAATCGAAACACAT	P	P
MGHES-47	AI727363.1	$(TCC)_5, (CAT)_7$	CCCCATCAGAAGGAGTGCTA	TGAATGATACCGCAGGGATT	M	M
MGHES-48	AI726417.1	$(CAG)_6$	AAAGGGAGATTGAAGCAGCA	CACCACCAAATCATCTGCAT	M	P
MGHES-49	AI725914.1	$(TTCC)_2, (CT)_{10}$	GGGGTCTCACTCAAATGCTC	TGGTGAGGGCTTAATCATGG	M	P
MGHES-50	AI725600.1	$(AGA)_{12}, (GAG)_4$	TGCACTAAATTCACCCACCA	GGGACCGAAGAAAGGAAGAA	P	P
MGHES-51	AI725599.1	$(ACAA)_5, (TA)_4, (AT)_5$	GCCCTTCAAACCAAACGTAA	GGCTGCTTCAAAAGCATAGA	M	M
MGHES-52	AF165924.1	$(CAC)_5, (AGG)_3, (CCA)_3$	GGAAGGAGGAGGCAGTGATA	GGGAAGATGAAACCGGTAGG	P	P
MGHES-53	AI055612.1	$(GT)_{10}, (AT)_3$	ACAAGGACCAGCAGTTTTGG	TCAAGTCCAAGTACTGCAATGAA	M	M
MGHES-54	AI055443.1	$(AAAAT)_3, (GAA)_3, (AT)_{10}$	AGCCCTTCACTTTCCCCTTT	CGTCGCCCGTACCATAAC	M	M
MGHES-55	AI055070.1	$(CAT)_5$	CGAACCCTAGCTTTCAATCG	CGGCTTCAATTGTACGGTCT	M	M
MGHES-56	AI055001.1	$(AT)_7$	ACCAGGACTGGGCTGAGATA	GAACGTATTTCCACAAGTCTAGCA	M	M
MGHES-57	AI054842.1	$(AGA)_6, (GA)_3$	CCACCCAGTTTGGAAAGTCA	TCTCCACTGGACTGCAACTG	P	P
MGHES-58	AI054761.1	$(CTT)_{10}, (CT)_3$	TCTCCATGTATCCACCCACA	GCATCGTGAGAGAAGTGAAGG	P	P
MGHES-59	AF336286.1	$(AGC)_5$	GCAAACCCAACCAGAGTCAT	AACCACTGCTGTTGTTGCTG	M	P
MGHES-60	AF336283.1	$(AT)_{15}/(TA)_5$	TCCATGGACCCAGAAGAAGA	TCAGTCTGCAACTCTTCCACA	M	P
MGHES-61	AF336282.1	$(TCT)_7, (TG)_5, (TC)_3$	CCAGTCCTTGCCTCCATTTA	TCAGTCTGCAACTCTTCCACA	M	M
MGHES-62	AF336280.1	$(CA)_{10}, (AG)_4$	TGCATCTGATCTAATTGTTGGTG	TGTTCCTCACAGCAAGAGCA	M	M
MGHES-63	D88414.1	$(TTTTA)_4$	GCGGACAATTGGTGCTTAAT	CCGCAAACTGACTCTAATTTTTC	M	P
MGHES-64	L04497.1	$(TGATT)_4$	CCGTTATTCTTTCTCTATTCTACCTG	ACATCTAAGCAGCCCAGCAG	M	M
MGHES-65a	M16891.1	$(AGA)_3, (GA)_3, (TG)_3$	CAATGTGTGAGGGAATGCAG	CCTTGAAGCAAACCTTTGGA	M	M
MGHES-65b	M16891.1	$(AACAAG)_3, (GAA)_5, (AAC)_3$	TCTTCCAGGCAAAGCTCATT	TAAAGACCGAACCCGTCATC	M	M
MGHES-66	L17308.1	$(TCC)_3, (CTC/CCA)_3$	TCCTCCTCCCACTTCATCAC	GACTGTGGCTGGAGGAGAAG	M	P
MGHES-67	AF336277.1	$(GGA)_5, (CT)_3$	GACTACCCACCCACCAAAGA	CCCGAGTTTTCGCTACTGAT	M	M
MGHES-68	AI725806.1, AI729485.1	$(TTC)_7$	CTTCGCCCATTTCTTACCTG	GAGAAGGGTCGGGAGAGATT	M	M
MGHES-69	AI055489.1	$(AT)_{11}, (CA)_4$	AAGAGGGCTTGAAGCTGTTG	TATAGGCGAAATGGGTACGG	M	M
MGHES-70	AW186947.1	$(GCA)_7$	CCAATAGGACTTTGGGTTTGG	CTTGCGAGGATCAGAAAAGC	P	P
MGHES-71	AI731721.1	$(CAC)_4/(CCA)_5/(CCA)_3$	ATCACCACCTCCACCATCTC	CTCCGATTACAGGTGGCAGT	P	M
MGHES-72	AI729015.1	$(TTA)_3$	CCCTTCGTTGTTTCCTCAAC	GGGCTCATCCTCTTCGACTT	M	M
MGHES-73	AI728199.1	$(ATTC)_5, (TCC)_3$	CCCGATATCCTTAGCCTTTT	AGTCGGAGGTGATGGTTAGG	P	P
MGHES-74	AI726497.1	$(AC)_{11}$	AAACCCCCAGGGGAAAAA	GGGGCCATGTTTCTGTACC	M	P
MGHES-75	AI054830.1	$(CAA)_8$	GCATATGTCGAAGATGCTACCC	AGCATCAGCAGTAGGCCAAG	M	P
MGHES-76	AF336284.1	$(GA)_{20}$	ATGGGGTCATCCTCCATTTT	TTTCTCCATTTGCCTTCACC	M	M
MGHES-77	AF165925.1	$(AAG)_6$ AAA	GCACAGGGTGAGGAATTGAG	TGCACCAATATCACTTCTTTGC	M	M
MGHES-78	X52135.1	$(CT/AG)_n$	TGCTGGCAACTACCTGAGTG	GATTCCAGGAACCACAATGG	M	M

#MGHES represents Mississippi USDA/ARS (M), *Gossypium hirsutum* (GH), EST-SSR (ES). Suffix a or b indicates multiple primer pairs for the same sequence.

EST-SSR polymorphism. All of the 84 primer pairs amplified products from the three *G. hirsutum* lines and the *G. barbadense* line (Table 1). Sixty primer pairs produced amplicons of the expected size. An average of three amplicons was obtained per primer pair because some of the primer pairs amplified more than two PCR products. More than one amplicon might result from amplification of homoeologous loci of the A and D genomes in the tetraploid cotton species (AD genome) and/or due to the amplification of paralogous loci within the genome.

Polymorphism of EST-SSRs within *G. hirsutum* was 26% (Table 1). Forty-four primer pairs (52%) were polymorphic between *G. hirsutum* and *G. barbadense*. Moreover, the monomorphic EST-SSRs may have potential for detecting single nucleotide polymorphisms (SNPs) that could also be used as markers to map important genes. Reddy et al. (2001) observed 21% polymorphism within *G. hirsutum* and 49% polymorphism between *G. hirsutum* and *G. barbadense* for SSR markers derived from genomic sequences. The level of polymorphism detected across the four cotton genotypes in this study was similar to the results of Reddy et al. (2001).

DNA Star software (DNA Star; Madison, WI) was used to locate the position of the SSRs in the EST sequences. Fifty-eight percent of EST-SSRs were in the 5′UTR region, 36% in the coding DNA sequence, and only 6% in 3′UTR region. Tri-nucleotide pure repeats had the highest level of polymorphism (69%) followed by di-nucleotide repeats. The imperfect repeats had the lowest frequency of polymorphism among all classes of repeats in this study. Interrupted and imperfect SSRs in the EST sequences generated lower polymorphism, presumably due to the initial stages of mutational decay with less chance for polymerase slippage (Smulders et al., 1997). Most of the repeat motifs with a higher number of repeats showed higher polymorphism in agreement with the previously reported positive correlation between the repeat length and polymorphism (Weber, 1990).

Sequencing of amplified products. Selected PCR products were cloned and sequenced to further confirm the presence of the SSR in the EST markers (Table 2, Fig. 1). A total of 16 PCR products, amplified with a set of four randomly selected EST-SSR primer pairs (MGHES-25, 30a, 41 and 58) using genomic DNA from four different cotton lines (TM-1, HS46 and MARCABUCAG8US-1-88, PIMA 3-79), were selected for sequencing. Both monomorphic and polymorphic markers based on the randomly selected primers were used for sequencing. Each of the amplified products was cloned and 24 clones from each amplified product were sequenced. Almost all of the sequences of cloned products gave products within the expected size range (Table 2). All of the cloned products have the same motif as the original EST sequence but with variable numbers of repeats, confirming the presence of SSR motifs in the amplified products (Fig. 1). A multiple sequence alignment with sequences for the original *G. hirsutum* EST (AI054761), its homologous *G. arboreum* EST (BQ404189), and the sequence of cloned PCR products from four different cotton lines revealed several point mutations, such as one base insertions, deletions, or transitions, in addition to the variable repeat numbers of the SSR motif (CTT)n (Table 2, Fig. 1). Overall, the sequence alignments showed that priming sites were conserved, but revealed the presence of variable number of microsatellite repeats (Fig. 1). For example, Pima 3-79 (*G. barbadense*)

Table 2. Sequence analysis of selected amplified products from four different cotton lines using EST-SSR primers

EST-SSR primer pair	Cotton genotype[z]	Expected product size (bp)	Observed product size (bp)	Repeat motif of original sequence	Repeat motif of sequenced product
MGHES-25	Pima 3-79	200	200	$(TTA)_7$	$(TTA)_6$
	TM-1		203		$(TTA)_7$
	HS 46		203		$(TTA)_7$
	MAR-1-88		203		$(TTA)_7$
MGHES-30a	Pima 3-79	165	177	$(CT)_{13}$	$(CT)_{13}$
	TM-1		167		$(CT)_{13}$
	HS 46		167		$(CT)_{13}$
	MAR-1-88		167		$(CT)_{13}$
MGHES-41	Pima 3-79	198	204	$(CCA)_8$	$(CCA)_8$
	TM-1		201		$(CCA)_7$
	HS 46		201		$(CCA)_7$
	MAR-1-88		201		$(CCA)_7$
MGHES-58	Pima 3-79	228	223	$(CTT)_{10}$	$(CTT)_9$
	TM-1		232		$(CTT)_{10}$
	HS 46		232		$(CTT)_{10}$
	MAR-1-88		229		$(CTT)_9$

[z] **TM-1 is considered genetic standard line of *G. hirsutum*; HS 46 and MAR-1-88 (MARCABUCAG8US-1-88) are two improved lines of *G. hirsutum;* Pima 3-79 is a double haploid line of *G. barbadense.***

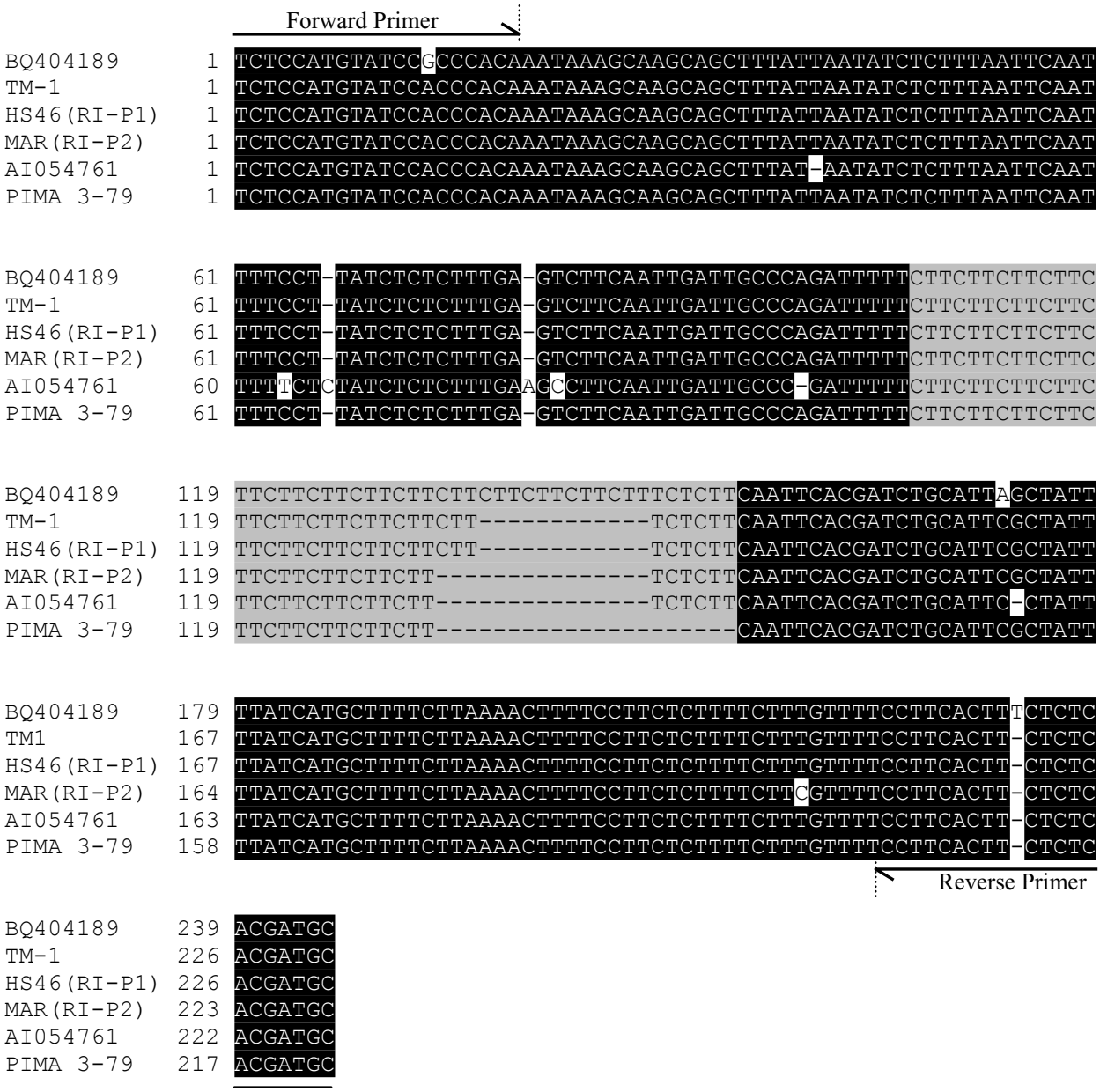

Figure 1. Analysis of sequences derived from different cotton lines that were amplified using the primer pair MGHES-58. Sequence alignment of the four cloned amplification products, the original *G. hirsutum* sequence (AI054761) from which the primer pair had been developed, and a *G. arboreum* EST from Genbank (BQ404189). The CTT repeat motifs were shown with gray background and the sequences were aligned in the 5′-3′ orientation. TM-1, HS46, and MAR (MARCABUCAG8US-1-88) are genotypes of *G. hirsutum,* and Pima 3-79 is a genotype of *G. barbadense*.

had a larger deletion of six bases at the end of one SSR compared with the other lines, which distinguished it from the *G. hirsutum* lines. The homologous region of the *G. arboreum* EST (BQ404189) had four extra repeat units, a transition of A to G in the forward primer region and an insertion of T in the reverse primer region (Fig. 1). There was little clone to clone variation in the sequences derived from a single line (data not shown), which may be caused by Taq polymerase error, by allelic variations due to heterozygosity, by amplification of a mixture of homologous and homoeologous products, or by PCR-mediated recombination (Cronn et al., 2002).

The EST-SSR primers successfully amplified other species of cotton, *G. arboreum* (A_2) and *G. raimondii* (D_5), and detected 52% polymorphism between these two species. Recently, genomic SSRs derived from *G. hirsutum* were used for mapping in other species, such as *G. barbadense*, *G. nelsonii* Fryx., and *G. australe* F. Muell. (Liu et al. 2000a;

2000b; Qureshi et al., 2001). These results indicate that there is a high degree of sequence conservation between the different diploid and tetraploid species of *Gossypium*. The ability to amplify these markers by PCR, conservation of sequence across different species of cotton, a high polymorphism rate, and the potential for sequence based comparative mapping make the EST-SSR markers ideal for comparative mapping in cotton.

To explore the potential utility of the EST-SSR markers for comparative mapping, the non-redundant EST-SSR sequences were compared with sequences of other cotton species, as well as the sequences of other crops and model organisms, using BLAST assuming an e-value $<1E^{-10}$ as a significant homology. About 74% of the EST-SSR loci had sequence similarity with a gene expressed in fiber tissue of *G. hirsutum*. In contrast, 85% of all the *G. hirsutum* ESTs originate from fiber tissue. Forty eight percent of the non-redundant EST-SSR sequences matched with ESTs from *G. arboretum* derived from fiber tissues. This result highlights the differences in expression profiles of cotton fiber tissues from *G. hirsutum* and *G. arboreum*.

These EST-SSR markers can be useful for comparative mapping with the Arabidopsis genome because about 19% of the sequences showed significant sequence similarity with sequences in the *A. thaliana* genome. Reddy et al. (2001) reported that about 13% of genomic JESPR SSR sequences from cotton were similar to sequences in the Arabidopsis genome and is consistent with the finding that many of the genomic SSRs were derived from expressed portion of the genome (Saha et al., 2003). Soltis et al. (1999) reported a close evolutionary relationship between *Gossypium* (*Malvaceace*) and *Arabidopsis* (*Brassicaceae*) based on taxonomic and molecular phylogenetic evaluation. Despite the close relationship between cotton and Arabidopsis, significant matches were identified for only one fifth of the non-redundant EST-SSR sequences. Since 74% of the EST-SSR sequences originated in cDNA libraries of fiber tissue from *G. hirsutum*, the non-matching sequences might be unique to fiber tissues. The EST-SSR markers described here should prove suitable as comparative markers between the Arabidopsis and cotton genomes and could be used for comparative mapping between these two species based on sequence matching as observed in other plant species (Sorrells et al., 2003).

CONCLUSIONS

In conclusion, this study identifies and makes publicly available new PCR-based EST-SSR markers for cotton. These markers will complement the currently available genomic SSR markers. Since these markers were developed based on expressed sequences and they are conserved across other cotton species and other closely related plant species such as Arabidopsis, they may be useful for comparative mapping. In addition, these markers may be useful for studies of fiber traits because majority of these EST sequences were derived from fiber tissues and have fewer matches to other plant genomic and expressed sequences. This is the first report on the development of PCR based EST markers in cotton. These markers will be very useful in the future for map-based cloning of important genes and integrating physical and functional genome maps using BAC clones recently developed in other laboratories.

ACKNOWLEDGEMENT

We acknowledge the help of Mr. D. Dollar of USDA/ARS, Mississippi State, Dr. R. Katam and Dr. A. Khan, Plant and Soil Science Dept., Mississippi State University for technical help in this research. The authors also thank the late Dr. A. Zipf of Plant and Soil Sciences Dept., Alabama A&M University and Dr. O.U. Reddy of School of Agriculture, Alcorn State University for their helpful suggestions on the manuscript.

DISCLAIMER

Mention of trade names or commercial products or vendors in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

REFERENCES

Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean genetics. Genetics 132:1131–1139.

Altaf-Khan, M., J. McD Stewart, J. Zhang, R.G. Cantrell, and M.K. Wajahatullah. 1997. Molecular and morphological genetics of a trispecies F_2 population of cotton. p. 448-452. *In* Proc. Beltwide Cotton Conf., New Orleans, LA. 6-10 Jan. 1997. Natl. Cotton Counc.. Am., Memphis, TN.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J.Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Arumuganathan K., and E.D. Earle. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9:208-218.

Bell, C.J., and J.R. Ecker. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144.

Billotte, N., A.M. Risterucci, E. Barcelos, J.L. Noyer, Pamblard, and F.C. Baurens. 2001. Development, characterization and across-taxa utility of oil palm microsatellites. Genome 44(3):413-425.

Brubaker, C.L., A.H. Paterson, and J.F. Wendel. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184-203.

Brubaker, C., R. Cantrell, M. Giband, B. Lyon and T. Wilkins. 2000. Letter to Journal of Cotton Science community. J. Cotton Sci. 4(2):149-151. [Online]. Available at http://journal.cotton.org/2000/issue02/article13.pdf (verified 7 June 2004).

Calhoun, D.S., D.T. Bowman, and O.L. May. 1997. Pedigrees of upland and pima cotton cultivars released between 1970 and 1995. Bull. 1069. Mississippi Agric. For. Exp. Stn., Mississippi State, MS.

Cardle, A.L., L. Ramsy, D. Milbourne, M. Macaulay, D. Marshell, and R. Waugh. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847-854.

Cho, Y.G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S.R. McCouch, W.D. Park, N. Ayres and S. Cartinhour. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet 100:713-722.

Condit, R., and S.P. Hubbell. 1991. Abundance and DNA sequence of two-base repeat regions in tropical tree genomics. Genome 34:66–71.

Connell, J.P., S. Pammi, M.J. Iqbal, T. Huizinga, and A.S. Reddy. 1998. A high through-put procedure for capturing microsatellites from complex plant genomes. Plant Mol. Biol. Rep. 16:341–349.

Cronn, R., M. Cedroni, T. Haselkorn, C. Grover, and J.F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploidy cotton. Theor. Appl. Genet. 10:482-489

Delseny, M. 1999. Genomics: methods and early results. Ol. Corps Gras Lipides, 6(2):136-143.

Effie, A., G. Seaton, K. Scott, D. Shelton, M.W. Graham, P. Baverstock, L.S. Lee, and R. Henry. 2000. Analysis of grape ESTs: global gene expression patterns in leaf and berry. Plant Sci. 159:87-95.

Eujayl, I., M.E. Sorrells, M. Baum, P. Wolters, and W. Powell. 2002. Assessment of genotypic varation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Theor. Appl. Genet. 104:399-407.

Gutiérrez, O.A., S. Basu, S., S. Saha, J.N. Jenkins, D.B. Shoemaker, C. Cheatham, and Jr. J.C. McCarty, Jr. 2002. Genetic distance among selected cotton genotypes and its relationship with F_2 performance. Crop Sci. 42:1841-1847.

Gupta, P.K., H.S. Balyan, P.C. Sharma, and B. Ramesh. 1996. Microsatellites in plants: a new class of molecular markers. Curr. Sci. 70:45-54.

Herron, B.J., G.H. Silva, and L. Flaherty. 1998. Putative assignment of ESTs to the genetic map by use of SSLP database. Mammalian Genome 9:1072-1074.

Jiang C., R. Wright, S. Woo, T. Delmonte, and A. Paterson. 2000. QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor. Appl. Genet. 100:409-418.

Kantety, R.V., M.L. Rota, D.E. Mathews, and M.E. Sorrells. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant. Mol. Biol. Rep. 48:501-510.

Liu, S., S. Saha, D. Stelly, B. Burr, and R.G. Cantrell. 2000a. Chromosomal assignment of microsatellite loci in cotton. J. Hered. 91:326-332.

Liu, S., R. G. Cantrell, J.C. McCarty, Jr., and J. McD. Stewart. 2000b. Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. Crop Sci. 40:1459-1469.

Liu L., W. Guo., X. Zhu, T. Zhang. 2003. Inheritance and fine mapping of fertility restoration for cytoplasmic male sterility in *Gossypium hirsutum* L. Theor. Appl. Genet. 106:461-469.

Miller R.T., A.G. Christoffels, C. Gopalakrishnan, J. Burke, A.A. Ptitsyn, T.R. Broveak, and W.A. Hide. 1999. A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. Genome Res. 9: 1143-55.

Morgante, M., and A.M. Oliver. 1993. PCR-amplified microsatellite markers in plant genetics. Plant J. 3:175–182.

Paterson A.H., Y. Saranga, M. Menz, and C.X. Ziang. 2003. QTL analysis of genotype X environment interactions affecting cotton fiber quality. Theor. Appl. Genet. 106:384-396.

Paterson, A.H., and R.H. Smith. 1999. Future horizons: biotechnology for cotton improvement. p. 415-432. *In* C.W. Smith and J.T. Cothren (ed.) Cotton. John Wiley & Sons, Inc., NY.

Qureshi, S.N., M. Altaf-Khan, J.N. Jenkins, C.L. Brubaker, U.K. Reddy, and S. Saha. 2001. A genetic linkage map of SSR markers based on *Gossypium nelsonii* and *G. australe* F_2 population. p. 219. *In* Plant and Animal Genome Meeting IX Abstract, San Diego, CA. 13-17 Jan 2001. Available online at http://www.intl-pag.org/pag/9/abstracts/P5q_22.html (verified 7 June 2004).

Reinisch, J. A., J. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendell, and A. H. Paterson. 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* X *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829-847.

Reddy, O.U.K., T.D. Brooks, K.M. El-Zik, and A.E. Pepper. 2000. Development and use of PCR-based technologies for cotton mapping. p. 483. *In* Proc. Beltwide Cotton Res. Conf., San Antonio, TX. 4-8 Jan. 2000. Natl. Cotton Council Am., Memphis, TN.

Reddy, O.U., A.E. Pepper, I. Abdurakmonov, S. Saha, J.N. Jenkins, T. Brook, and K.M. El-Zik. 2001. New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. J. Cotton Sci. 5:103-113. [Online]. Available at http://journal.cotton.org/2001/issue02/html/page103.html (verified 7 June 2004).

Ren, H.-L., W.-Z. Guo and T.-Z. Zhang. 2002. Identification of quantitative trait loci (QTLs) affecting yield and fiber properties in chromosome 16 in cotton using substitution line. Acta Botanica Sinica 44(7): 815-820.

Rozen, S. and H. J. Skaletsky. 1996. Primer3: A software component for picking PCR primers. Source code available online at http://www-genome.wi.mit.edu/genome_software/other/primer3.html (verified 7 June 2004).

Saha, S., J.N. Jenkins, J.C. McCarty, and D.M. Stelly. 1998. Chromosomal location of RFLP markers linked to QTL in cotton. p. 204-205. *In* Proc. World Cotton Res. Conf., Athens, Greece. 6-12 Sept. 1998. Int. Cotton Advisory Committee, Washington, DC.

Saha S., M. Karaca, J.N. Jenkins, A.E. Zipf, O.U. Ramesh, and R.V. Kantety. 2003. Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130: 355-364.

Sasaki, T., J.Y. Song, Y. Kogaban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. Murayamakayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamamoto, Y. Mukai, I. Ohta, N. Miyadera, I. Havukkala, and Y. Minobe. 1994. Toward cataloguing all rice genes: Large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J. 6:615-624.

Scott, K.D., P. Eggler, G. Seaton, M. Rossetto, E.M. Ablett, L.S. Lee, and R.J. Henry. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. 100:723-726.

Shappley, Z.W., J.N. Jenkins, W.R. Meredith, and J.C. McCarty Jr. 1998. An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. Theor. Appl. Genet. 97:756–761.

Shappley, Z. W., J. N. Jenkins, C.E. Watson Jr, A .L. Kahler, and W. R. Meredith. 1996. Establishment of molecular markers and linkage groups in two F2 populations of Upland cotton. Theor. Appl. Genet. 92:915-919.

Smulders, M.J.M., G. Bredemeijer, W. Rus-Koretekaas, and P. Arens. 1997. Use of short microsatellites from database sequences to generate polymorphism among Lycopersicon species. Theor. Appl. Genet. 97:264-272.

Soltis, P.S., D.E. Soltis, and M.W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402:402–404.

Sorrells, M.E. 2000a. The evolution of comparative plant genetics. *In* J.P. Gustafson (ed.) Genomes. Proc. 22nd Stadler Symp., Columbia, MO. 6-8 Jun. 1998.

Sorrells, M.E. 2000b. Comparative genomics for plant improvement. *In* H. Tefera (ed.) Proc. of Intl. Workshop for Tef Improvement., Addis Ababa, Ethiopia. 13-16 Oct. 2000.

Sorrells ME, M. La Rota, C.E. Bermudez-Kandianis, R.A. Greene, R. Kantety, J.D. Munkvold, Miftahudin, A. Mahmoud, X. Ma, P.J. Gustafson, L.L. Qi, B. Echalier, B.S. Gill, D.E. Matthews, G.R. Lazo, S. Chao, O.D. Anderson, H. Edwards, A.M. Linkiewicz, J. Dubcovsky, E.D. Akhunov, J. Dvorak, D. Zhang, H.T. Nguyen, J. Peng, N.L. Lapitan, J.L. Gonzalez-Hernandez, J.A. Anderson, K. Hossain, V. Kalavacharla, S.F. Kianian, D.W. Choi, T.J. Close, M. Dilbirligi, K.S. Gill, C. Steber, M.K. Walker-Simmons, P.E. McGuire, C.O. Qualset. 2003. Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. 8:1818-1827.

The Arabidopsis Information Resource (TAIR). 2004. TAIR Blast 2.2.8. Available online at http://www.arabidopsis.org/Blast/

Ulloa, M., and W.R. Meredith, Jr. 2000. Genetic linkage map and QTL analysis of agronomic and fiber quality traits in intraspecific populations J. Cotton Sci. 4:161-170. [Online]. Available at http://journal.cotton.org/2000/issue03/html/page161.html (verified 7 June 2004).

Weber J.L. 1990. Informativeness of human (cC-dA)n (dG-dT)n polymorphisms. Genomics 7:524-530.

Wright, R., P. Thaxton, A. Paterson, and K. El-Zik. 1998. D-subgenome bias of Xcm resistance genes in tetraploid *Gossypium* (Cotton) suggests that polyploidy formation has created novel avenues for evolution. Genetics 149:1987-1996.

Wright R., P. Thaxton, A. Paterson, and K. El-Zik. 1999. Molecular mapping of genes affecting pubescence of cotton. J. Hered. 90:215-219.

Wu, K., and S.D. Tanksley. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. 241:225–235.

Yamamoto, K., and T. Sasaki. 1997. Large scale EST sequencing in rice. Plant Mol. Biol. Rep. 35:135-144.

Yu, J., and R.J. Kohel. 1999. Update of the cotton genome mapping. p. 439. *In* Proc. Beltwide Cotton Res. Conf., Orlando, FL. 3-7 Jan. 1999. Natl. Cotton Council Am., Memphis, TN.

Zhang J., W. Guo, T. Zhang. 2002. Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* X *Gossypium barbadense* L.) with a haploid population. Theor. Appl. Genet. 105:1166-1174.

Zhang T., Y. Yuan., J. Yu, W. Guo, RJ Kohel. 2003. Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theor. Appl. Genet. 106:262-268.