TÜBİTAK

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2013) 37: 288-299

doi:10.3906/tar-1202-35

Development of short gSSRs in *G. arboreum* and their utilization in phylogenetic studies

Tayyaba SHAHEEN[1,2,*], Yusuf ZAFAR[1], James M. STEWART[3], Mehboob-ur RAHMAN[1]

[1]Plant Genomics and Molecular Breeding Laboratory, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

[2]Department of Bioinformatics and Biotechnology, GC University, Faisalabad, Pakistan

[3]Department of Crop, Soil, and Environmental Sciences, Division of Agriculture, University of Arkansas, Fayetteville, Arkansas 72701, USA

Received: 12.02.2012 • **Accepted:** 08.09.2012 • **Published Online:** 15.05.2013 • **Printed:** 05.06.2013

Abstract: Microsatellite regions in DNA fragments of *Gossypium arboreum* were explored and PCR products 500–800 bp long were analyzed for genomic simple sequence repeats (gSSRs). From 39 segments, 23 short gSSRs were identified; on average they occurred every 1.6 kb. Primers were used to amplify fragments from 18 diploid species representing 6 diploid genomes and 1 tetraploid species, *G. hirsutum* (AD). The phylogeny of diploid genomes was subsequently determined. No primer amplified the DNA from all genotypes; however, amplification occurred in 59% of 437 PCR events. All genomes were grouped into 2 clusters (a and b). The obtained phylogeny of the species parallels previous reports. The strategy adopted here is an efficient method of identifying new gSSRs from species not extensively explored.

Key words: *Gossypium*, phylogenetic assessment, RAPD, short gSSRs

1. Introduction

DNA markers offer an expedient way to study polymorphisms and their use in crop improvement. PCR-based markers are either sequence-specific or random (Gupta et al. 1999). Sequence specific markers are preferred over random markers because of the unreliability of random markers (Rahman et al. 2002; Guo et al. 2003). Microsatellites, also called simple sequence repeats (SSRs), have been utilized in genetic mapping, population genetic studies, gene tagging, and genetic diversity estimation (Kuleung et al. 2004; Cubry et al. 2008; Lancon et al. 2008; Ali et al. 2009; Arunita et al. 2010; Buriev et al. 2010; Kalivas et al. 2011). Expressed sequence tag (EST)-SSRs have been derived from public databases in rice (Cho et al. 2000), grape (Effie et al. 2000; Scott et al. 2000), and cotton (Qureshi et al. 2004). EST-SSRs are derived from transcribed genes and are usually present in the gene-rich regions of the genome (Karaca and İnce 2011). In contrast, gSSRs are highly polymorphic and they tend to be widely dispersed throughout the genome, resulting in better map coverage than EST-SSRs (La Rota et al. 2005). Development of gSSR markers is a time-consuming and laborious task (Saha et al. 2006; Karaca and İnce 2011).

Of more than 50 species of the genus *Gossypium* L., only 4 are cultivated: 2 diploids (2n = 2x = 26), *G. arboreum* L. (A_2A_2) and *G. herbaceum* L. (A_1A_1), and 2 allotetraploids (2n = 4x = 52), *G. hirsutum* L. (AADD) and *G. barbadense* L. (AADD). Numerous methods have been employed to classify cotton species (*Gossypium* spp.) based on morphology (Fryxell et al. 1992) including meiotic behavior (Menzel 1954) and karyotype (Beasley 1940); genetic and molecular methods have also been used (Wendel 1995; Iqbal et al. 1997). The common origin of *Gossypium* is estimated at 5–15 million years ago (Wendel 2000). The diploid species in *Gossypium* are morphologically diverse and are represented genetically by 8 different genome types including the A, B, C, D, E, F, G, and K genomes (Endrizzi et al. 1985; Stewart 1995). With the exception of the A genome, all the diploid genomes are composed of wild species. These wild species represent a rich genetic resource for cotton improvement (Stewart 1995).

In this study, short gSSRs were identified with a novel strategy. Random decamer primers were utilized to amplify genomic regions of *G. arboreum* from all chromosomes because they comprise only 10 bases

* Correspondence: tayaba_pgmb@yahoo.com

and can anneal anywhere in the genome (Welsh and McClelland 1990). These gSSRs primers were further utilized to amplify the genomic DNA of 18 diploid species and 1 tetraploid species, *G. hirsutum*. Results were analyzed to assess transferability of the gSSRs and the phylogenetic relations of the different diploid genomes.

2. Materials and methods

2.1. Plant material

Experimental materials consisted of 18 diploid species and 1 tetraploid species (Table 1). Leaves of the cotton species were collected from the cotton fields of the Central Cotton Research Institute (CCRI), Multan, Pakistan. in 2006, and genomic DNA of some of the species was obtained from Dr JM Stewart (Arkansas, USA).

2.2. Total genomic DNA isolation

DNA was extracted from 10 individual plants of each genotype according to the method used by Iqbal et al. (1997). After RNase treatment, the DNA concentration was measured by fluorometer (DyNA QuantTM 200®). The quality of the DNA was checked by running 25 ng of DNA through 0.8% agarose gel. DNA samples that did not show a discrete band were rejected. Total genomic DNA was diluted in double distilled water to a concentration of 15 ng μL^{-1} for PCR analysis.

2.3. Amplification of genomic DNA of *G. arboreum* cultivar Ravi with random decamer primers

Genomic DNA of *G. arboreum* 'Ravi' was amplified with 48 random decamer primers. PCR was performed in a 50-µL reaction volume containing 10 mM Tris-HCl (pH 8.3); 50 mM KCl; 3 mM $MgCl_2$; 100 µM each of dATP, dCTP, dGTP, and dTTP; 30 ng of primer; 0.001% gelatin; 37.50 ng of genomic DNA; and 2 units of Taq DNA polymerase. Taq DNA polymerase, buffer, $MgCl_2$, dNTPs, and gelatin were all obtained from MBI Fermentas. Amplification was performed in an Eppendorf Mastercycler. Amplified products ranged in size from 100 bp to 3 kb.

Table 1. List of 18 diploid and 1 tetraploid species of *Gossypium* included in study (Guo et al. 2006). NIBGE: National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan.

Sr #	Species name	Genome	Source	Distribution
1	*G. arboreum*	(A_2)	NIBGE	Old World
2	*G. herbaceum africanum*	(A_1)	CCRI (Multan)	Africa
3	*G. khlotzchianum*	($D_{3\text{-}k}$)	CCRI (Multan)	Galapagos Islands
4	*G. thurberi*	(D_1)	CCRI (Multan)	Mexico, Arizona
5	*G. herkensii*	($D_{2\text{-}2}$)	Dr JM Stewart (Arkansas)	Mexico
6	*G. davidsonii*	($D_{3\text{-}d}$)	Dr JM Stewart (Arkansas)	Mexico
7	*G. aridum*	(D_4)	CCRI (Multan)	Mexico
8	*G. captis viridis*	(B_3)	Dr JM Stewart (Arkansas)	Africa, Cape Verde
9	*G. raimondii*	(D_5)	CCRI (Multan)	Peru
10	*G. gossypoides*	(D_6)	CCRI (Multan)	Mexico
11	*G. lobatum*	(D_7)	Dr JM Stewart (Arkansas)	Mexico
12	*G. laxum*	(D_9)	CCRI (Multan)	Mexico
13	*G. trilobum*	(D_8)	CCRI (Multan)	Mexico
14	*G. stocksii*	(E_1)	Dr JM Stewart (Arkansas)	Arabian Peninsula
15	*G. incanum*	(E_4)	CCRI (Multan)	Arabian Peninsula
16	*G. longicalyx*	(F)	CCRI (Multan)	East Africa
17	*G. bickii*	(G_1)	Dr JM Stewart (Arkansas)	Australia
18	*G. nelsonii*	(G_3)	Dr JM Stewart (Arkansas)	Australia
19	*G. hirsutum*	(AD)	NIBGE	New World

2.4. Cloning and sequencing of PCR products of required size

From the gel, 72 PCR products amplified with 48 primers including OPC-1, OPC-6, OPC-8, OPC-9, OPC-11, OPC-13, OPD-16, OPG-17, OPH-13, OPI-1, OPI-3, OPJ-1, OPJ-5, OPJ-7, OPJ-10, OPJ-19, OPK-2, OPK-4, OPK-5, OPK-6, OPK-7, OPK-8, OPK-11, OPK-12, OPK-17, OPL-1 through 8, OPL-11 through 15, OPL-18, OPL-20, OPN-1, OPZ-1, OPZ-7, OPZ-10, OPZ-12, OPZ-14, OPZ-15, and OPZ-17 ranging in size from 500 to 800 bp were eluted and cloned into the ptz57R/T vector for sequencing purposes. Cloned fragments were sequenced on an ABI automated sequencer, and sequences were analyzed manually for detection of microsatellites.

2.5. Primer designing for flanking regions of repetitive regions

Repetitive regions or genomic SSRs were detected in sequences amplified with 21 primers (OPC-6, OPC-8, OPC-13, OPC-11, OPD-16, OPJ-1, OPJ-5, OPK-6, OPK-7, OPK-12, OPL-1, OPL-2, OPL-3, OPL-4, OPL-6, OPL-11, OPL-12, OPL-14, OPL-18, OPZ-1, and OPZ-7).

Primers ranging in size from 18 to 20 bp were designed using Primer 3 Plus software based on the flanking regions of repetitive regions and synthesized. These primers were named PGMB (an abbreviation for the Plant Genomics and Molecular Breeding Laboratory) (Table 2).

2.6. PCR amplification

Conditions for amplifying the *G. arboreum* genomic DNA with these primers were optimized. All primers were amplified at 55 °C, with the exception of PGMB-6, which was amplified at 50 °C. PCR was performed in a total volume of 20 µL using 2.5 µL (15 ng μL^{-1}) of cotton DNA; 10X 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; 0.5 units of Taq DNA polymerase; and 0.15 mM of each primer. Taq DNA polymerase, 10X PCR buffer, $MgCl_2$, and dNTPs were purchased from MBI Fermentas. Following 1 min at 94 °C, PCR consisted of 35 cycles of 94 °C for 30 s, 55 °C for 30 s, a 72 °C extension for 1 min, and a final extension at 72 °C for 10 min. PCR products were resolved on Metaphor agarose gels (4%).

2.7. Data scoring and statistical analysis

The number of amplified fragments with PGMB primers in each species was counted as dominant type without considering polymorphism. Genotypes without any amplification were considered null alleles. The size of the most intense bands was scored using a 50-bp DNA ladder. The data were used to estimate similarity based on Nei and Li's (1979) coefficient. Similarity coefficients were used to generate a dendrogram by the unweighted pair group method of arithmetic means (UPGMA) implemented in PAUP* Version 4.0. Statistics for phylogenetic associations among accessions consisted of the UPGMA based on similarity. Percentage amplification was estimated according to Kuleung et al. (2004) as:

% Amplification = (no. of amplified markers × 100/ total no. of markers).

Percentage of occurrence was the percentage of amplification of gSSRs markers amplified in *Gossypium* species other than *G. arboreum*.

2.8. Finding repetitive regions in *Arabidopsis thaliana*, *Oryza sativa*, and cotton ESTs and their comparison with *G. arboreum*

DNA sequences of *Arabidopsis thaliana*, *Oryza sativa*, and cotton ESTs of the same length covered by random primers in *G. arboreum* (A_2) were retrieved randomly from GenBank. Repeats were detected in these sequences, and their frequency of occurrence was compared to the frequency of their occurrence in the A_2 genome.

3. Results

A total of 48 random decamer primers were used to amplify genomic DNA of *G. arboreum* cultivar Ravi. Amplification products ranging in size from 500 to 800 bp were cloned and sequenced to find repetitive regions. Repetitive regions were detected in sequences of 23 DNA fragments. In total, 39 kb of the cotton genome was covered by these random primers, in which 23 gSSRs were identified, indicating their occurrence every 1.7 kb on average. Similarly, the 39-kb genome of *Arabidopsis thaliana* contained 24 repeats, indicating an average of 1 repeat every 1.6 kb. In *Oryza sativa*, 25 repeats were detected, an average of 1 repeat every 1.6 kb. However, the 39-kb ESTs derived from *G. hirsutum* contained 21 repeats, for an average of 1.85 kb between repeats.

Among the repeats found in the genomic DNA of *G. arboreum*, most repeats were dinucleotide (34%), followed by tri- (27.58%), mono- (20.6%), tetra- (13.79%), and heptanucleotide (3.44%) (Table 2). In 8 regions, more than 1 type of repetitive motif occurred. In the case of *Arabidopsis* most repeats were dinucleotides (55.55%), followed by tri- (33.33%), mono- (7.4%), and tetranucleotides (3.7%). In the case of *Oryza sativa* most repeats were dinucleotide (59%), followed by trinucleotides (27.27%), tetranucleotides (9.9%), and hexanucleotides (4.5%); in cotton ESTs the greatest number of repeats were trinucleotides (52.38%) followed by dinucleotides (47.6%).

In the present study, 6 out of 23 sequences of *G. arboreum* did not exhibit homology with any previously reported sequence. Ten sequences showed homology with reported *G. hirsutum* sequences [alcohol dehydrogenase A gene (2), *G. hirsutum* BAC clones (2), *G. hirsutum* restorer (Rf1) gene (1), *G. hirsutum* putative calcium binding protein (2), and *G. hirsutum* retrotransposon genes (3)], and 4 sequences showed homology with *Arabidopsis*

Table 2. SSR primer sequences, number of repeats and their status of polymorphism interspecifically, homology of the sequence amplified by these primers, and transferability of these regions in other genomes. In the fifth column, polymorphism (P) or monomorphism (M) observed in these primers across the diploid genome is indicated; in the sixth column, the percentage of transferability of the genomic regions amplified by these primers is given.

Primer name	Sequence	Repeat unit	Homology of the sequences amplified by these primers	Diploid species	Transferability (%)
PGMB-1	F 5′TTCCAGATACTATGGGCC3′ R 5′CTGGCGAACTAAATCACC3′	$(GAC)_4$	No significant homology	P	21
PGMB-2	F 5′GCGGTACTTAGTCCTTAG3′ R 5′GGACAGCTACATGCTAAC3′	$(AAT)_4$ $(ATTTT)_4$ $(TTC)_4$	Partial homology with ribosomal protein gene	M	31
PGMB-3	F 5′CTTTTTGTCCCTACCACC3′ R 5′ACAAGTGCTTTTGCCACC3′	$(TTTTAAA)_2$	73% homology with *G. arboreum* alcohol dehydrogenase A gene, partial cds	P	26.31
PGMB-4	F 5′GAGGGAAGAGTGAAGAAG3′ R 5′AGAAAAGGCTTGAGGACC3′	$(TAA)_n$ $(CAT)_3$	69% homology with *G. hirsutum* BAC clone	P	83.00
PGMB-5	F 5′GAAAGGGAAGGGAGTAAGC3′ R 5′GCGGTACTTTACCCATGA3′	$(T)_n$ $(A)_n$	Homology with *G. hirsutum* clone UBC169-800 restorer (Rf1) gene	P	63.15
PGMB-6	F 5′AGGACTTACTTACCCGAC3′ R 5′CTATTCTATCTAGCCTCGC3′	$(A)_{12}$ $(CA)_3$	No significant homology	P	15.78
PGMB-7	F 5′GCTGCGGACCAACAAAAC3′ R 5′GCTTAGGTATGGTACCCC3′	$(T)_{10}$ $(TTA)_2$	78% homology of partial sequence with *Solanum lycopersicum* genomic DNA	P	89.47
PGMB-8	F 5′GCACAGAGGACAAATGGT3′ R 5′GCGAGCAAGAGCAATTAC3′	$(ATT)_4$	No significant homology	P	89.47
PGMB-9	F 5′TTGGGCCGATTACTCCAT3′ R 5′CCATAAGGAACTCAACAAC3′	$(TAA)_3$ $(TTTA)_2$ $(TA)_3$ $(CAA)_2$	No significant homology	P	63
PGMB-10	F 5′CACCATCTCAGGATTCTC3′ R 5′TCTCACTCCGCCATTCTG3′	$(CAAA)_n$	Homology of partial sequence with *Z. mays* Gypsy/Ty3-, *Z. mays* (adh1) gene, adh1-F allele	P	31
PGMB-10B	F 5′TCCTACAGATAGAGTTC3′ R 5′CCCACCAATCATACAAGG3′	$(CA)_3$ $(GA)_3$ $(CCT)_2$	Homology with *Arabidopsis thaliana* transporter-related gene	P	21

Table 2. (Continued).

Primer name	Sequence	Repeat unit	Homology of the sequences amplified by these primers	Diploid species	Transferability (%)
PGMB-11	F 5′CTTGGACTTCAGCAGGAC3′ R 5′GATGTTTCGAGCAGGATC3′	$(CAA)_2(GAAA)_2$	Homology with *A. thaliana* transporter-related gene	P	73.68
PGMB-13	F 5′GCTAGTGATTCTCGAAGG3′ R 5′GCAAACAACGATGTTGCTCC3′	$(TA)_3(CAA)_2(CAT)_2$	No significant homology	M	57.89
PGMB-14	F 5′CAATCACACGGCATCCA3′ R 5′CCGGCATAATGGGGTTTATG3′	$(AC)_5(AT)_{11}$	Homology of partial sequence with *G. hirsutum* clone	P	68.42
PGMB-15	F 5′CGTTTATTGGGAGGCAACTC3′ R 5′AGGAATGTCTCCCCCTACTT3′	$(ATTT)_5$	Homology with *Arabidopsis thaliana* genome	P	68.42
PGMB-16	F 5′GAGGGGTATTTGACATGC3′ R 5′GAGGTTTCGAGACCACAA3′	$(CA)_3(CA)_4$	Homology with *G. arboreum* adh A gene and *G. hirsutum* putative calcium-binding protein gene	P	83.21
PGMB-17	F 5′CCTTCCCACTACTACACA3′ R 5′TAGAGTTGTGACGCCTCA3′	$(CA)_4(AC)_3(GT)_3$	Homology with *G. hirsutum* putative calcium-binding protein gene	P	68.42
PGMB-18	F 5′CAGATCCGGTCATAACGT3′ R 5′TGGCCGATACTTTCCT3′	$(GA)_6$	Homology with *G. hirsutum* retrotransposon putative copia and *G. arboreum* adh A gene	P	15
PGMB-18B	F 5′AGGAAAGTATCGGCCA3′ R 5′GGTGGTGTACCGAATTTG3′	$(GAA)_n(GA)_n(A)_n$	Homology with *G. hirsutum* retrotransposon putative copia and *G. arboreum* adh A gene	P	15
PGMB-19	F 5′CAGTAGGCCACTAATCCA3′ R 5′CACCTTTCCCACGATT3′	$(TA)_4$	Homology with *Arabidopsis thaliana* glycosyl hydrolase family	P	42.1
PGMB-20	F 5′TAGTCCCGCCTCTATCTT3′ R 5′TCTACGTCACCACTGCAA3′	$(T)_n(A)_n$	No significant homology	P	68.42
PGMB-21	F 5′GGGGTTGAAGATGGATAC3′ R 5′CCTAGTCGAAATGGTGGT3′	$(T)_n(A)_n$	Homology with *G. hirsutum* clone Cott-1 retrotransposon Ty3-Gypsy-like nonfunctional gag and pol genes, complete sequence	P	36.84
PGMB-22	F 5′ACATAAAAGCCCCCTAGC3′ R 5′ACAAAACACCTGAGCAGG3′	$(T)_n(A)_n$	Homology with *G. hirsutum* retrotransposon putative copia, transposon, with *Gossypium hirsutum* clone and *G. arboreum* adh A gene	P	21.05

thaliana sequences, 1 sequence with *Solanum lycopersicum*, 1 sequence with a ribosomal protein, and 1 sequence with a *Zea mays* retrotransposon (Table 2).

Primers were designed for the 23 gSSRs containing repeats. These primer pairs were surveyed on 6 of the 8 diploid cotton genomes (Figure 1) and amplified 54 alleles with an average of 2.39 per primer pair. The most alleles (6) were amplified with gSSR primer pair PGMB-21. Of the 23 primer pairs, only PGMB-2 and PGMB-13 produced a monomorphic pattern. Among the gSSR primer pairs, 12 were genome- or species-specific (Tables 3 and 4). PGMB-22, -6, -18, and -18b could amplify only in the A and AD genomes. PGMB-1, -2, and -3 could not amplify any of the B-, D-, E-, or F-genome species. As a result, these primers could be used as B-, D-, E-, and F-genome negative primers. PGMB-21 could not amplify any D- or B-genome species. PGMB-14 did not amplify in D_9 or D_4 genomes or B, E, or F genomes. PGMB-7 could not amplify only in *G. laxum* (D_9) and *G. raimondii* (D_5). PGMB-8 did not amplify in the D_9 genome. PGMB-10 could not amplify B-, D-, E-, or F-genome species (Table 4). These primers can be used as genome-specific primers.

3.1. Transferability of gSSRs among diploid genomes

All the primer pairs yielded amplification products in type A- and AD-genome genotypes (Table 3). None of the primers amplified the genome of all of the species. No clear correlation

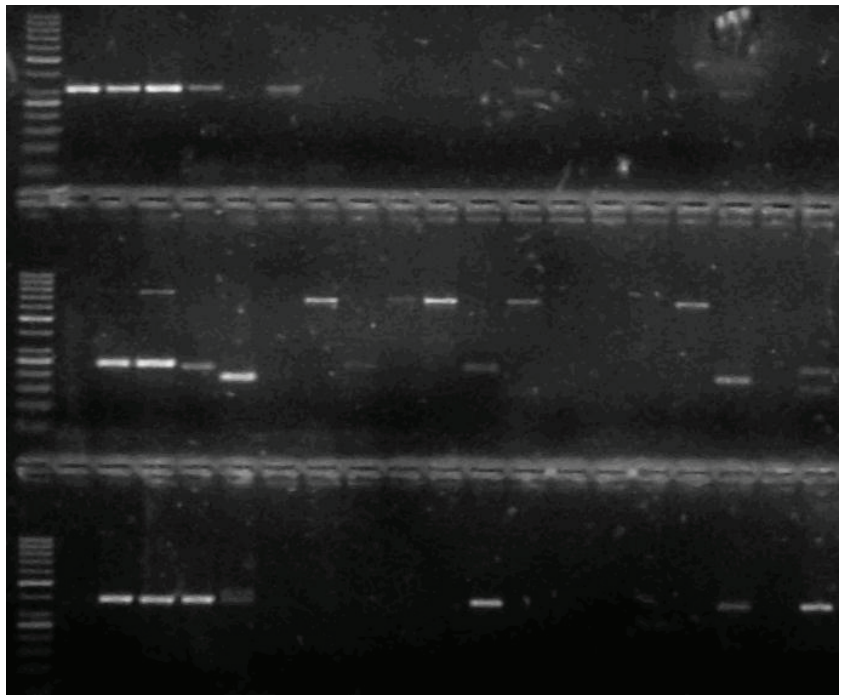

Figure 1. Gel picture of PGMB-19, -20, and -21 amplifications in 18 diploid subgenomes and *G. hirsutum*; M = marker (50-bp DNA ladder).

Table 3. Cross-species transferability of *G. arboreum*-derived genomic SSRs among different *Gossypium* genomes.

Genome	No. of SSRs amplified wholly	% amplified wholly	No. of SSRs amplified partially	No. null amplified
A	23	100	0	0
B	13	56	0	10
D	0	0	15	8
E	9	39	5	9
F	11	47	0	12
G	12	52	4	7
AD	23	100	0	0

was observed in the type of repeat motif or in transferability in the genomes. When all amplification events (23 gSSRs × 19 species = 437) were examined, 59% of the gSSRs occurred in more than 1 genome group. The lowest number of species providing amplification products occurred in PGMB-6 and PGMB-18 primer pairs, while PGMB-8 amplified fragments from the most species. Among the genomes, species belonging to F, B, G, and E genomes showed high transferability; D-genome species exhibited low transferability. The gSSRs PGMB-7 and PGMB-8, which consisted of trinucleotide repeats, exhibited transferability (89%) to the most genotypes. After PGMB-8, the most common gSSRs were PGMB-4 and PGMB-16 (83%). These are tri- and dinucleotide gSSRs, respectively. Transferability of primers based on single nucleotide repetitions was also high (Table 4).

3.2. Genetic similarity and phylogenetic study of diploid genomes with genomic SSRs

The genetic similarity of A_2 and AD genomes with the A_1 genome was 71.9%, while genetic similarity among the other genomes ranged from 39.6% for A_1 and D_4 to 92.5% for D_9 and D_5 (Table 5). A high degree of genetic similarity between E_1 and E_4 species (89.0%) and for each combination $D_5 \times D_6$ and $D_6 \times D_9$ (87.5 %) was observed. For the combinations $D_1 \times D_{2\text{-}2}$ and $D_8 \times D_{3\text{-}k}$ 85.4% and 85%, respectively, and for each combination $D_1 \times D_{3\text{-}d}$ and $D_{2\text{-}2} \times D_{3\text{-}d}$ 84.4% genetic similarity was estimated (Table 5).

Cluster analysis grouped the species into 2 main clusters, 'a' and 'b'. A-genome species *G. arboreum, G. herbaceum africanum* (A_1), and *G. hirsutum* (AD) were grouped into the a cluster (Figure 2). The second cluster, b, comprised all other species. This cluster consisted of 4 subclusters: b_1 comprised *G. thurberi* (D_1), *G. trilobum* (D_8), *G. harknessii* ($D_{2\text{-}2}$), *G. davidsonii* ($D_{3\text{-}d}$), and *G. klotzschianum* ($D_{3\text{-}k}$); b_2 included *G. aridum* (D_4), *G. raimondii* (D_5), *G. laxum* (D_9), and *G. gossypoides* (D_6); *G. lobatum* (D_7) was intermediate between these 2 clusters. In subcluster b_3 *G. stocksii* (E_1) and *G. incanum* (E_4) were closely related; other members of the cluster are *G. captis viridis* (B_3) and *G. longicalyx* (F). Subcluster b_4 included *G. bickii* (G_1) and *G. nelsonii* (G_3) (Figure 2).

4. Discussion

Many studies have been conducted to find SSRs in large genomes like cotton because of their robustness (Reddy et al. 2001; Nguyen et al. 2004; Qureshi et al. 2004). However, to find genomic SSRs in a large genome is an uphill task that involves construction and screening of genomic libraries (Nguyen et al. 2004). Here, we adopted a different strategy for detecting gSSRs in the cotton genome using random decamer primers, which have successfully delineated the phylogeny of cotton species. This strategy is time-friendly and cost-effective because it involves amplification using random decamer primers followed by cloning into a T/A

Table 4. Amplification features of genome-specific and partial species-specific genomic SSRs.

Primer	Genomes						
	A	B	D	E	F	G	AD
PGMB-1	Ö	X	X	X	X	* 5	Ö
PGMB-2	Ö	X	X	X	X	Ö	Ö
PGMB-3	Ö	X	X	X	X	Ö	Ö
PGMB-6	Ö	X	X	X	X	X	Ö
PGMB-7	Ö	Ö	* 1	Ö	Ö	Ö	Ö
PGMB-8	Ö	Ö	* 2	Ö	Ö	Ö	Ö
PGMB-10	Ö	X	X	X	X	Ö	Ö
PGMB-14	Ö	X	* 3	X	X	Ö	Ö
PGMB-18	Ö	X	X	X	X	X	Ö
PGMB-18b	Ö	X	X	X	X	X	Ö
PGMB-21	Ö	X	X	* 4	Ö	Ö	Ö
PGMB-22	Ö	X	X	X	X	X	Ö

X indicates no amplification in any species belonging to this genome; * indicates partial amplification except for the following genomes: * 1 = null in *G. laxum* and *G. raimondii*, * 2 = null in *G. laxum*, * 3 = null in *G. laxum* and *G. aridum*, * 4 = null in *G. incanum*, * 5 = null in *G. nelsonii*.

Table 5. Similarity matrix of all species and varieties assessed by PGMB primers.

Sr #	A_2, AD	A_1	D_7	D_1	D_{2-2}	D_8	D_{3-d}	D_4	D_{3-k}	D_9	D_5	D_6	B_3	F	G_1	G_3	E_1	E_4
A2, AD	1																	
A1	0.719	1																
D7	0.525	0.500	1															
D1	0.584	0.459	0.792	1														
D2-2	0.525	0.542	0.804	0.854	1													
D8	0.521	0.521	0.803	0.870	0.855	1												
D3-d	0.525	0.584	0.792	0.844	0.844	0.855	1											
D4	0.521	0.396	0.688	0.771	0.730	0.730	0.688	1										
D3-k	0.542	0.542	0.792	0.834	0.822	0.850	0.834	0.688	1									
D9	0.605	0.480	0.713	0.755	0.835	0.792	0.771	0.834	0.771	1								
D5	0.521	0.438	0.771	0.743	0.813	0.792	0.730	0.834	0.783	0.925	1							
D6	0.563	0.438	0.713	0.771	0.813	0.792	0.730	0.812	0.778	0.875	0.875	1						
B3	0.563	0.521	0.771	0.713	0.771	0.750	0.713	0.709	0.771	0.750	0.750	0.750	1					
F	0.625	0.500	0.709	0.750	0.667	0.605	0.750	0.730	0.667	0.730	0.646	0.646	0.730	1				
G1	0.667	0.667	0.667	0.709	0.75	0.730	0.750	0.605	0.709	0.646	0.646	0.646	0.730	0.584	1			
G3	0.625	0.542	0.542	0.667	0.667	0.605	0.667	0.563	0.584	0.646	0.605	0.563	0.688	0.625	0.765	1		
E1	0.563	0.563	0.730	0.688	0.771	0.709	0.771	0.584	0.771	0.667	0.667	0.667	0.800	0.771	0.688	0.730	1	
E4	0.521	0.563	0.730	0.771	0.713	0.750	0.855	0.667	0.771	0.750	0.750	0.750	0.814	0.771	0.730	0.730	0.890	1

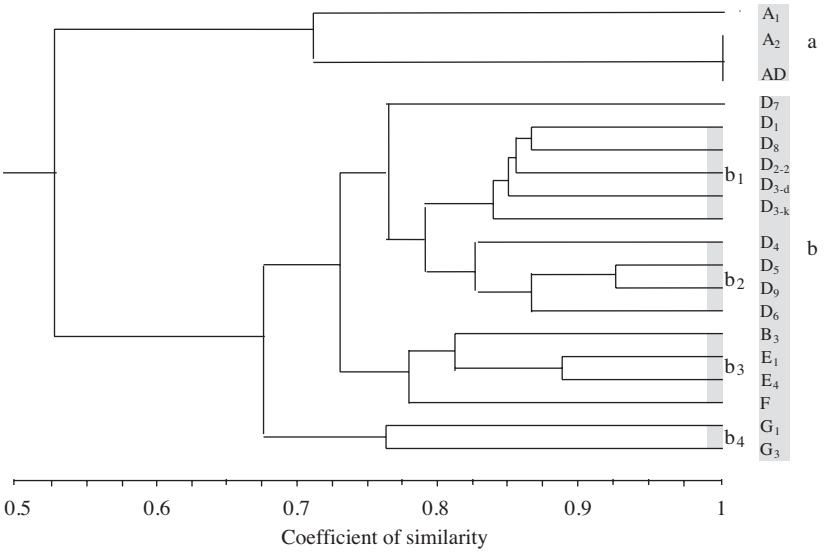

Figure 2. Dendrogram of 19 *Gossypium* species, including *G. hirsutum*, developed from randomly amplified polymorphic DNA (RAPD)-derived gSSRs using UPGMA. Scale is based on Nei and Li's coefficients of similarity.

cloning vector. Direct sequencing of PCR products can further simplify the process.

The frequency of dinucleotide repeats was highest, followed by other types including tri-, mono-, tetra-, penta-, hexa-, and heptanucleotide repeats. Generally, the occurrence, relative abundance, and relative density of SSRs decrease as the repeat unit increases (Lee et al. 2004; Ince et al. 2011). In the case of cotton ESTs, trinucleotide repeats were more prevalent than dinucleotide repeats, which corresponds well with the earlier findings of Lawson and Zhang (2006). This dominance of trimeric SSRs over di-, tetra-, and pentameric SSRs may be explained by the suppression of nontrimeric SSRs in coding regions due to the risk of frameshift mutations that may occur when those microsatellites alternate in size by 1 unit (La Rota et al. 2005; Hong et al. 2007).

In the present study, homology of the sequences with other genomes was not associated with their transferability and polymorphism, a result confirmed in an earlier report (Guo et al. 2006). Most sequences (43%) had homology with previously reported sequences derived from *G. hirsutum* and *G. arboreum*. Four sequences (17%) had homology with *A. thaliana* sequences. Such homologies have been reported in earlier investigations. For example, at least 59% of cotton and 52% of *Arabidopsis* transcriptomes showed correspondence in multilocus gene arrangement (Rong et al. 2005); this is indicative of a shared common ancestor and divergence between 83 and 86 million years ago (Benton 1993). Homology searches of sequences showed that cotton (Malvaceae) has homology with members of the families Solanaceae (dicot), Brassicaceae (dicot), and Poaceae (monocot). This also indicates the genetic relatedness of cotton with other plant families (Shaheen et al. 2006; Ahmad et al. 2007).

At the interspecific level, allelic polymorphism was 91%, which might be the result of accumulation of mutations during evolution across the *Gossypium* genomes (Nei et al. 2007). These gSSRs are informative interspecifically, which is consistent with previous studies (Qureshi et al. 2004). The short length gSSRs, due to their high frequency, can play a vital role in constructing high resolution genetic maps in interspecific populations.

Genomic SSRs have a high rate of polymorphism and a low rate of transferability across species (Peakall et al. 1998; Kuleung et al. 2004). In contrast, EST-SSRs are less informative (Decroocq et al. 2003). Guo et al. (2006) found a high rate (96.5%) of cross-species transferability in the diploid genomes of cotton using EST-SSRs. In the present study we observed 59% cross-species transferability, while none of the primer pairs could amplify all the species. This showed a comparatively low rate of cross-species amplification and transferability in gSSRs. EST-SSRs are derived from transcribed regions of DNA and are more conserved; this limits their polymorphism and, thus, they exhibit high transferability (Cho et al. 2000; Thiel et al. 2003).

In the present study, 12 out of 23 gSSRs were genome-specific, a result which can be used to distinguish cotton genomes. The advantage of genome-specific SSRs lies in their utility as codominant markers for identifying DNA

fragments introgressed from other species in *Gossypium* (Guo et al. 2006). It is thought that the diploid genomes of *Gossypium* diverged from a common ancestor (Stewart 1995). Results obtained in this study indicate that all genomes, to a large extent, share genomic footprints with the A genome. Wu et al. (2007) concluded that there are ancient genetic backgrounds among diploid cotton accessions. The gSSRs derived from A genome species showed a high level of transferability in F-, B-, G-, and E-genome species compared to the D-genome species, suggesting that D-genome species share a minimum of genetic material with the A genome. Similar results have been reported in earlier studies (Guo et al. 2006; Wu et al. 2007).

The A-genome species have been domesticated and have gone through fewer evolutionary changes than the D-genome species. In addition, 2 independent phylogenetic analyses of the D genome found that it has a faster evolutionary rate than the A genome (Adams and Wendel 2004). Comparative mapping studies of the A and D genomes have also demonstrated the differences in these genomes (Brubaker et al. 1999).

In the cluster analysis, the A genome and the tetraploid (AD) species were grouped into 1 cluster. In a second cluster, D-genome species were grouped into 2 subclusters: D_1, D_8, $D_{2\text{-}2}$, $D_{3\text{-}d}$, and $D_{3\text{-}k}$ species were grouped into 1 subcluster; the D_4-, D_5-, D_9-, and D_6-genome species were grouped into the second subcluster; and D_7 joined these 2 clusters separately. Numerous molecular and phylogenetic analyses have demonstrated the monophyletic origin of subgenus D (for example, Wendel and Cronn 2003). The D genome represents a morphological and cytogenetic New World assemblage of diploid *Gossypium* species (Endrizzi et al. 1985; Wendel 1995; Wendel and Cronn 2003). None of the D-genome species are important in terms of producing commercially important fiber; however, their contribution as a parental lineage of allotetraploid cultivated cotton (*G. hirsutum* L. and *G. barbadense* L.) is well established (Cronn et al. 1999; Endrizzi et al. 1985; Small et al. 1998; Small and Wendel 2000), and this gives special significance to the genome's systematics and evolutionary relationships. Both E-genome species were closely related to each other, and B_3 (*G. captis viridis*) and F (*G. longicalyx*) were also grouped in the same cluster while G-genome species were grouped in subcluster b_4.

Gossypium consists of 8 genomes that comprise 4 major lineages. The diploid species of these genomes are distributed in Australia (C-, G-, and K-genome species), the Americas (D-genome species), and Africa/Arabia (first lineage of the A-, B-, and F-genome species and second lineage of the E-genome species) (Fryxell 1979; Fryxell et al. 1992). The grouping of the B_3 and F genomes into 1 subcluster can be justified on the basis of their shared common lineage. In addition, extensive molecular and biological evidence from nuclear and/or chloroplast genome research is consistent with this classification (Wu et al. 2007). The findings of the present study suggest that this strategy is successful for identification of gSSRs from complex genomes, and its utility in establishing phylogenetic relationships at the interspecific level is comparable to the evolutionary relationships of diploid genomes of cotton reported in previous studies (Wendel and Albert 1992; Seelanan et al. 1997; Small et al. 1998, 1999).

The high density of short gSSRs can be extremely useful in establishing high-resolution genetic maps, especially in interspecific populations. Additionally, new gSSRs may be identified from less frequently surveyed genomes in order to increase the density of genetic maps.

Acknowledgments

Partial funding for the present research was provided through a project entitled "DNA-based characterization of cotton" (ALP-PARC, Islamabad). We are also grateful to the Higher Education Commission, Islamabad, Pakistan, for partial funding of this study through a PhD student grant.

References

Adams KL, Wendel JF (2004) Exploring the genomic mysteries of polyploidy in cotton. Biol J Linn Soc 82: 573–581.

Ahmad S, Ashraf M, Zhang T, Islam N, Shaheen T, Rahman M (2007) Identifying genetic variation in *Gossypium* L. based on single nucleotide polymorphism. Pak J Bot 39: 1245–1250.

Ali I, Ashraf M, Rehman M, Zafar Y, Asif M, Kausar A, Riaz S, Niaz M, Wahid A, Abbas SQ (2009) Development of genetic linkage map of leaf red colour in cotton (*Gossypium hirsutum*) using DNA markers. Pak J Bot 41: 1127–1136.

Arunita R, Rakshit S, Santhy V, Gotmare VP, Mohan P, Singh VV, Singh S, Singh J, Balyan HS, Gupta PK, Bhat SR (2010) Evaluation of SSR markers for the assessment of genetic diversity and fingerprinting of *Gossypium hirsutum* accessions. J Plant Biochem Biotech 19: 153–160.

Beasley JO (1940) The origin of American tetraploid Gossypium species. Am Nat 74: 285–286.

Benton MJ (1993) The Fossil Record 2. Chapman and Hall, New York.

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42: 184–203.

Buriev ZT, Saha S, Abdurakhmonov IY, Jenkins JN, Abdukarimov A, Scheffler BE, Stelly DM (2010) Clustering, haplotype diversity and locations of MIC-3: a unique root-specific defense-related gene family in upland cotton (*Gossypium hirsutum* L.). Theor Appl Gen 120: 587–606.

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Gene 156: 847–857.

Cho YG, Ishii T, Temnykh S, Chen X, Lopovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Gen 100: 713–722.

Cronn RC, Small RL, Wendel JF, (1999) Duplicated genes evolve independently after polyploid formation in cotton. Proc Natl Acad Sci USA 96: 14406–14411.

Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2008) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome 51: 50–63.

Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Gen 106: 912–922.

Effie A, Seaton G, Scott K, Shelton D, Graham MW, Baverstock P, Lee LS, Henry R (2000) Analysis of grape ESTs: global gene expression patterns in leaf and berry. Plant Sci 159: 87–95.

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. Advan Gene 23: 271–375.

Fryxell PA (1979) The Natural History of the Cotton Tribe. Texas A & M University Press, College Station, TX, USA.

Fryxell PA, Craven LA, Stewart JM (1992) A revision of *Gossypium* sect. Grandicalyx (Malvaceae), including the description of six new species. Sys Bot 17: 91–114.

Guo W, Wang W, Zhou B, Zhang T (2006) Cross species transferability of *G. arboreum*-derived EST-SSRs in the diploid species of *Gossypium*. Theor Appl Gen 112: 1573–1581.

Guo WZ, Zhang TZ, Shen XL, Yu JZ, Kohel RJ (2003) Development of SCAR marker linked to a major QTL for high fiber strength and its usage in molecular-marker assisted selection in upland cotton. Crop Sci 43: 2252–2256.

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118: 369–390.

Hong CP, Piao ZY, Kang TW, Batley J, Yang T, Hur YK, Bhak J, Park BS, Edwards D, Lim YP (2007) Genomic distribution of simple sequence repeats in *Brassica rapa*. Mol Cell 23: 349–356.

Ince AG, Karaca M, Onus AN (2011) Exact microsatellite density differences among *Capsicum* tissues and development stages. J Agri Sci 17: 291–299.

Iqbal MJ, Aziz N, Saeed NA, Zafar Y, Malik KA (1997) Genetic diversity of some elite cotton varieties by RAPD analysis. Theor Appl Gen 94: 139–144.

Kalivas A, Xanthopoulos K, Kehagia O, Tsaftaris AS (2011) Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers. Gene Mol Res 10: 208–217.

Karaca M, İnce AG (2011) New non-redundant microsatellite and CAPS-microsatellite markers for cotton (*Gossypium* L.). Turk J Field Crops 16: 172–178.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Gene 95: 10774–10778.

Kuleung C, Baenziger PS, Dweikat I (2004) Transferability of SSR markers among wheat, rye and triticale. Theor Appl Gen 108: 1147–1150.

La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice wheat and barley. BMC Genom 6: 23–35.

Lancon J, Pichaut JP, Djaboutou M, Lewicki Dhainaut S, Viot C, Lacape JM (2008) Use of molecular markers in participatory plant breeding: assessing the genetic variability in cotton populations bred by farmers. Annal Appl Biol 152: 113–119.

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genom Bio 7: R14.

Lee JM, Nahm SH, Kim YM, Kim BD (2004) Characterization and molecular genetic mapping of microsatellite loci in pepper. Theor Appl Gen 108: 619–627.

Menzel MY (1954) A cytological method for genome analysis in *Gossypium*. Genet 40: 214–223.

Nei M (2007) The new mutation theory of phenotypic evolution. Proc Natl Acad Sci USA 104: 12235–12242.

Nei N, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonuclease. Proc Natl Acad Sci USA 76: 5269–5273.

Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Gen 109: 167–175.

Peakall R, Gilmore S, Keys W, Morgante M, Rafaske A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeat (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15: 1275–1287.

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8: 112–123.

Rahman M, Malik TA, Aslam N, Asif M, Ahmad R, Khan IA, Zafar Y (2002) Optimisation of PCR conditions to amplify microsatellite loci in cotton (*Gossypium hirsutum* L.). Int J Agri Biol 2: 282–284.

Reddy OUK, Pepper AE, Abdurakhmonov I, Saha S, Jenkins JN, Brooks T, Bolek Y, El-Zik KM (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. J Cotton Sci 5: 103–113.

Rong J, Bowers JE, Schulze SR, Waghmare VN, Rogers CJ, Pierce GJ, Zhang H, Estill JC, Paterson AH (2005) Comparative genomics of *Gossypium* and *Arabidopsis*: unraveling the consequences of both ancient and recent polyploidy. Genom Res 15: 1198–1210.

Saha MC, Cooper JD, Mian MAR, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Gen 113: 1449–1458.

Scott KD, Eggler P, Seaton G, Rosetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grapes ESTs. Theor Appl Gen 100: 723–726.

Seelanan T, Schnabel A, Wendel JF (1997) Congruence and consensus in the cotton tribe. Sys Bot 22: 259–290.

Shaheen T, Rahman M, Zafar Y (2006) Chloroplast RPS8 gene of cotton reveals the conserved nature throughout plant taxa. Pak J Bot 38: 1467–1476.

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85: 1301–1315.

Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol 16: 491–501.

Small RL, Wendel JF (2000) Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (*Gossypium*). Gene 155: 1913–1926.

Stewart JM (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In: Challenging the Future: Proceedings of the First World Cotton Research Conference (Eds. GA Constable, NW Forrester), Brisbane, Australia, 13–17 February 1995. CSIRO, Canberra, Australia, pp. 317–327.

Thiel T, Michalek V, Graner A (2003) Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Gen 106: 411–422.

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nuc Acids Res 18: 7213–7218.

Wendel JF (1995) Cotton. In: Evolution of Crop Plants (Ed. NW Simmonds). Longman Scientific & Technical, Essex, UK, pp. 358–366.

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42: 225–249.

Wendel JF, Albert VA (1992) Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Sys Bot 17: 115–143.

Wendel JF, Cronn R (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78: 139–186.

Wu YX, Daud MK, Chen L, Zhu SJ (2007) Phylogenetic diversity and relationship among *Gossypium germplasm* using SSRs markers. Plant Sys Evol 268: 199–208.