GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Characterization and development of chloroplast microsatellite markers for *Gossypium hirsutum*, and cross-species amplification in other *Gossypium* species

X.Y. Cai, F. Liu, Z.L. Zhou, X.X. Wang, C.Y. Wang, Y.H. Wang and K.B. Wang

State Key Laboratory of Cotton Biology, Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, Anyang, China

Corresponding author: K.B. Wang
E-mail: wkbcri@163.com

Genet. Mol. Res. 14 (4): 11924-11932 (2015)
Received February 7, 2015
Accepted May 29, 2015
Published October 5, 2015
DOI http://dx.doi.org/10.4238/2015.October.5.6

ABSTRACT. Cotton is an important economic crop worldwide; its fiber, commonly known as cotton lint, is the main natural source for the textile industry. Sixty chloroplast microsatellites were identified and characterized from the complete sequence of the *Gossypium hirsutum* chloroplast genome using a bioinformatic approach. Twenty chloroplast microsatellite loci were polymorphic in the 66 *Gossypium* germplasm accessions. A total of 85 alleles were detected, with allele numbers varying from 2-7 per locus. Polymorphism information content varied from 0.02-0.66, with a mean of 0.48. Additionally, transferability of the 20 polymorphic chloroplast microsatellite primers was evaluated in other 31 *Gossypium* species. Sixteen markers were successfully amplified across all species tested, while the remaining 4 markers cross-amplified in most species tested. These polymorphic chloroplast microsatellite markers may be useful tool for studies of individual

identification, genetic diversity, evolution, conservation genetics, and molecular breeding in *Gossypium*.

Key words: Chloroplast microsatellite; *Gossypium*; Polymorphism; Transferability

INTRODUCTION

Cotton is one of the most important economic crops worldwide. Its fiber is the main natural source for the textile industry. The genus *Gossypium* consists of over 45 diploid (2N = 2x = 26) and 5 tetraploid (2N = 4x = 52) species distributed widely throughout the tropical and subtropical regions of the world (Wang et al., 2012). Diploid species have been classified into 8 genome groups designated A-G and K, while the 5 tetraploid species were designated as the AD genome group based on morphologic traits, cytogenetic pairing, and fertility of interspecific hybrids (Fryxell, 1992; Wendel and Albert, 1992). Four species are cultivated, including *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium herbaceum*, and *Gossypium arboretum*, while the others are wild species. Upland cotton (*G. hirsutum*) accounts for more than 90% of cotton fiber output, whereas sea cotton (*G. barbadense*) produces long-staple fibers and accounts for approximately 8% of cotton production. Wild cotton species have many valuable agronomic traits and abundant gene resources (Wang, 2007). However, many wild *Gossypium* species are at risk of extinction because of habitat deterioration, human activities, and low germination rates under natural conditions because of its hard episperm. Extensive field studies showed that wild *G. tomentosum* accessions, endemic to Hawaii, have sharply reduced in number.

Chloroplast microsatellites (cpSSRs) were developed by Powell et al. (1995) and successfully applied in population genetic studies of pines. Compared to the nuclear genome, the chloroplast genome is typically non-recombinant, uniparentally inherited, has a high copy number, and shows relatively conserved gene order, making it possible to detect relatively rapidly mutating sites in the chloroplast genome using cpSSR markers. To date, cpSSR markers have been developed for a variety of species and were shown to be invaluable in studying the evolution, differentiation, genetic diversity, and genetic structure of plant populations (Powell et al., 1995; Ishii and McCouch, 2000; Xu et al., 2002; Deguilloux et al., 2003; Dzialuk et al., 2009; Xue et al., 2012; Wang, 2013; Sugiura et al., 2014).

Numerous studies have examined the ecological and evolution aspects of *Gossypium* based on allozymes (Wendel et al., 1994), restriction fragment length polymorphism (Wendel and Brubaker, 1993), random-amplified polymorphic DNA (Tatineni et al., 1996; Lu and Myers, 2002), amplified fragment length polymorphism (Pillay and Myers, 1999; Rana et al., 2005), and simple sequence repeats (SSRs) (Wu et al., 2007; Yu et al., 2012). The results of these studies were complementary and largely congruent with existing genome designations and geographical distributions. However, there are still some ambiguities in the systematic relationships of these species and which genomes were more ancestral among extant cotton species. Additional markers should be developed to study the genetic diversity and phylogeny of *Gossypium*. In this study, we developed and characterized a set of polymorphic cpSSR markers from the complete chloroplast genome of *G. hirsutum* and tested their transferability to other *Gossypium* species, which will be useful in addressing questions related to the genetic diversity and phylogeny of *Gossypium*, thus making them important for evolutionary and conservation purposes.

MATERIAL AND METHODS

Sixty-six germplasm accessions of *Gossypium* belonging to 32 species were used to evaluate the polymorphism and transferability of the cpSSR markers (Table 1). All voucher samples were maintained in the National Wild Cotton Nursery (Sanya, Hainan Province, China), and supervised by the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences.

DNA was extracted from fresh leaves using the CTAB method as described by Zhang and Stewart (2000), with some modifications. DNA concentration and quality were evaluated by 1.0% (w/v) agarose gel electrophoresis and by spectrophotometry at 260/280 nm. The DNA solutions were stored at -20°C.

The complete sequence of the *G. hirsutum* chloroplast genome (accession No. NC_007944) was downloaded from GenBank (www.ncbi.nlm.nih.gov). Mononucleotide cpSSRs were located using the program Tandem Repeats Finder 4.04 (Benson, 1999), and polynucleotide cpSSRs were searched in SSRhunter (Li and Wan, 2005). The labile length of the motif was considered to be 8 or more nucleotides (Rose and Falush, 1998; Raubeson et al., 2007). In this study, 10 nucleotides was the threshold for motifs used to search for mononucleotide repeats. Motifs of 5, 4, and 3 were used to search for dinucleotide, trinucleotide, and tetranucleotide repeats, respectively. In addition, analysis of the structure of chloroplast genome was performed using the DOGMA software (http//:dogma.ccbb.utexas.edu/).

Primers were designed from the sequences of cpSSRs using the Primer Premier 5.0 software (Premier Biosoft, Palo Alto, CA, USA), based on G+C content, Tm, and the lack of secondary structure. Primers were designed in the 18-22-bp range to yield amplification products of 100-400 bases.

DNA from 64 individuals including 32 species was amplified by polymerase chain reaction in a 10-μL reaction volume containing 1 μL 10X PCR Buffer, 1.75 mM $MgCl_2$, 175 μM dNTPs, 0.5 U TransStart Taq DNA polymerase (TransGen, Biotech, Beijing, China), and approximately 30 ng DNA template. The amplification was performed in a PT600 Programmable Thermal Controller (TaKaRa, Shiga, Japan) following a touchdown cycling program: 5 min denaturation at 95°C, followed by 10 cycles of 30 s denaturation at 95°C, decreasing temperatures from 60° to 50°C by 1°C every cycle for 30 s, 50 s elongation at 72°C, followed by 30 cycles of denaturation at 95°C, 30 s annealing at 50°C, and 50 s elongation at 72°C, and a final extension for 5 min at 72°C. The amplification products were separated on 8% denatured polyacrylamide gel. Gels were visualized by silver-staining. The molecular sizes of the amplicons were estimated by comparison with a 500-bp DNA ladder (Invitrogen, Carlsbad, CA, USA).

RESULTS

A total of 60 cpSSRs were developed from the whole chloroplast genome sequence of *G. hirsutum* (Table 2). We found 1 long single-copy sequence, 1 short single-copy sequence, and 2 inverted repeat in the *G. hirsutum* chloroplast genome. cpSSRs were mainly distributed randomly across the single-copy region with both the long single-copy sequence (49) and short single-copy sequence (9) of the *G. hirsutum* chloroplast genome, but only 2 cpSSRs in the inverted repeat region. There were 43 mononucleotide cpSSRs (mainly A/T), 15 dinucleotide cpSSRs, and 2 trinucleotide cpSSRs, with lengths of at least 10 nucleotides. Although cpSSRs were present throughout the whole chloroplast genome, they were not evenly distributed, appearing to be concentrated in several areas.

Table 1. Sixty-six cotton germplasm accessions used in this study.

Sample No.	Material	Genome	Geographic origin
1	*G. hirsutum* L. 'Pengpengmian'	AD_1	Hainan, China
2	*G. hirsutum* L. 'zhong12'	AD_1	China
3	*G. hirsutum* L. 'zhong16'	AD_1	China
4	*G. hirsutum* L. 'TM-1'	AD_1	America
5	*G. hirsutum* L. 'T582'	AD_1	America
6	*G. barbadense* L. 'H7124'	AD_2	South America
7	*G. barbadense* L. 'Yuanmo'	AD_2	Yunnan, China
8	*G. barbadense* L. 'Mile'	AD_2	Yunnan, China
9	*G. barbadense* L. 'Kaiyuanlihe'	AD_2	Yunnan, China
10	*G. barbadense* L. 'cheli'	AD_2	Yunnan, China
11	*G. barbadense* L. 'Kaiyuanlianhe'	AD_2	Yunnan, China
12	*G. barbadense* L. 'Arablianhe'	AD_2	Yunnan, China
13	*G. tomentosum* Nuttall ex Seemann	AD_3	Hawaii Islands
14	*G. tomentos*um Nuttall ex Seemann	AD_3	Hawaii Islands
15	*G. mustelinum* Miers ex Watt	AD_4	Brazil
16	*G. mustelinum* Miers ex Watt	AD_4	Brazil
17	*G. darwinii* Watt	AD_5	Galapagos Islands
18	*G. darwinii* Watt	AD_5	Galapagos Islands
19	*G. darwinii* Watt	AD_5	Galapagos Islands
20	*G. herbaceum* L. 'jinta'	A_1	Asia
21	*G. herbaceum* L. 'Africanum'	A_1	Africa
22	*G. arboreum* L. 'Rozi'	A_2	Asia
23	*G. arboreum* L. 'zhongya1'	A_2	Asia and Africa
24	*G.arboreum* L. 'AK235'	A_2	Asia
25	*G. anomalum* Wawra & Peyritsch	B_1	Africa
26	*G. anomalum* Wawra & Peyritsch	B_1	Africa
27	*G. capitis-viridis* Mauer	B_2	Cape Verde Islands
28	*G. capitis-viridis* Mauer	B_2	Cape Verde Islands
29	*G. sturtianum* Willis	C_1	Central Australia
30	*G. sturtianum* Willis	C_1	Central Australia
31	*G. nandewarense* Derera	$C_{1\text{-}n}$	Southeast Australia
32	*G. robinsonii* Mueller	C_2	West Australia
33	*G. thurberi* Todaro	D_1	Arizona
34	*G. armourianum* Kearney	$D_{2\text{-}1}$	Mexica and California
35	*G. harknessii* Brandegee	$D_{2\text{-}2}$	Mexica and California
36	*G. davidsonii* Kellogg	$D_{3\text{-}d}$	Mexica and California
37	*G. davidsonii* Kellogg	$D_{3\text{-}d}$	Mexica and California
38	*G. davidsonii* Kellogg	$D_{3\text{-}d}$	Mexica and California
39	*G. klotzschianum* Andersson	$D_{3\text{-}k}$	Galapagos Islands
40	*G. aridum* (Rose & Standley) Skovsted	D_4	Mexica and California
41	*G. raimondii* Ulbrich	D_5	Peru
42	*G. raimondii* Ulbrich	D_5	Peru
43	*G. raimondii* Ulbrich	D_5	Peru
44	*G. gossypioides* (Ulbrich) Standley	D_6	Mexica
45	*G. gossypioides* (Ulbrich) Standley	D_6	Mexica
46	*G. gossypioides* (Ulbrich) Standley	D_6	Mexica
47	*G. lobatum* Gentry	D_7	Michoacan
48	*G. trilobum* (DC.) Skovsted	D_8	West Mexica
49	*G. laxum* Phillips	D_9	Guerrero
50	*G. schwendimanii* Fryxell	D_{11}	Mexica
51	*G. stocksii* Masters in Hooker	E_1	Arabia
52	*G. stocksii* Masters in Hooker	E_1	Arabia
53	*G. stocksii* Masters in Hooker	E_1	Arabia
54	*G. somalense* (Gurke) Hutchinson	E_2	North Africa
55	*G. somalense* (Gurke) Hutchinson	E_2	North Africa
56	*G. areysianum* Deflers	E_3	South Yemen
57	*G. incanum* (Schwartz) Hillcoat	E_4	South Africa
58	*G. incanum* (Schwartz) Hillcoat	E_4	South Africa
59	*G. longicalyx* Hutchinson & Lee	F_1	East Africa
60	*G. bickii* Prokhanov	G_1	Central Australia
61	*G. bickii* Prokhanov	G_1	Central Australia
62	*G. bickii* Prokhanov	G_2	Central Australia
63	*G. australe* Mueller	G_2	Australia
64	*G. nelsonii* Fryxell	G_3	Central Australia
65	*G. nelsonii* Fryxell	G_3	Central Australia
66	*G. nelsonii* Fryxell	G_3	Central Australia

Table 2. Distribution of 60 cpSSRs in the *Gossypium hirsutum* chloroplast genome.

No.	Motif	SSR start (bp)	SSR end (bp)	Distribution area
1	A_{10}	1906	1916	LSC
2	T_{11}	2419	2430	LSC
3	C_{10}	5505	5515	LSC
4	$(T_8)CG(A_{10})$	5664	5684	LSC
5	A_{10}	6657	6667	LSC
6	A_{10}	7620	7630	LSC
7	A_{10}	8897	8907	LSC
8	T_{10}	9429	9439	LSC
9	$(C_{10})(T_7)\ldots(T_9)$	14715	14749	LSC
10	A_{12}	14924	14936	LSC
11	T_{11}	19341	19352	LSC
12	T_{11}	21770	21781	LSC
13	T_{12}	27055	27067	LSC
14	$A_{10}\ldots G_8$	38033	38057	LSC
15	A_{12}	52682	52694	LSC
16	T_{11}	54773	54784	LSC
17	T_{12}	66007	66019	LSC
18	$(T_{10})C(T_6)$	67266	67283	LSC
19	G_{13}	67304	67317	LSC
20	T_{10}	69432	69442	LSC
21	T_{12}	70335	70347	LSC
22	$(T_{14})AG(T_9)$	71286	71311	LSC
23	T_{10}	71629	71639	LSC
24	T_{10}	74663	74673	LSC
25	$A_{10}\ldots A_8$	74965	74988	LSC
26	A_{10}	79410	79420	LSC
27	G_{10}	80014	80024	LSC
28	A_{10}	80784	80794	LSC
29	T_{10}	81140	81150	LSC
30	A_{10}	82150	82160	LSC
31	$(T_6)C(T_{10})$	84443	84460	LSC
32	T_{10}	84959	84969	LSC
33	A_{11}	85490	85501	LSC
34	T_{12}	103834	103846	IR
35	A_{12}	115241	115253	SSC
36	A_{10}	115362	115373	SSC
37	A_{11}	115885	115896	SSC
38	$(A_{11})T(A_7)$	117135	117154	SSC
39	A_{10}	118295	118305	SSC
40	A_{11}	120246	120257	SSC
41	T_{10}	125787	125797	SSC
42	A_{10}	131300	131310	SSC
43	A_{12}	145199	145211	IR
44	$(CT)_7T_8$	17110	17132	LSC
45	$(AT)_5$	20722	20732	LSC
46	$(AT)_5$	28125	28134	LSC
47	$(TG)_5T_9$	32721	32740	LSC
48	$(TA)_6$	33024	33035	LSC
49	$(AT)_5$	44519	44528	LSC
50	$(AT)_5$	46839	46848	LSC
51	$(AT)_5$	48436	48445	LSC
52	$(TA)_5$	49137	49146	LSC
53	$(AT)_5$	51060	51069	LSC
54	$(TA)_5$	58355	58364	LSC
55	$(AT)_5$	64971	64980	LSC
56	$(TC)_5$	65263	65272	LSC
57	$(AT)_5$	86188	86197	LSC
58	$(TA)_5$	131174	131183	SSC
59	$(AAT)_4$	13592	13603	LSC
60	$(ATA)_4$	48444	48455	LSC

LSC = long single-copy sequence in the chloroplast genome; SSC = short single-copy sequence in the chloroplast genome; IR = 2 inverted repeats in the chloroplast genome.

Of the 60 cpSSRs, 40 were amenable to primer design. Primers could not be designed for the remaining sequences because of the unavailability of a sufficient number of bases to meet the primer designing criteria or because of biased base compositions in the flanking nucleotide sequences. A set of 40 primer pairs were synthesized and checked for 5 *G. hirsutum* accessions (Sample No. 1-5). Of the 40 cpSSR primers, 20 primers producing repeatable banding patterns were polymorphic (Table 3). Other primers that showed no amplification, no multiple bands, or no pronounced stutters were excluded from further study.

Table 3. Characteristics of 20 polymorphic cpSSRs developed in *G. hirsutum*.

Markers	Motif	Primer sequences (5'-3')	Ta (°C)	Size (bp)	N_A	PIC
cp01	A_{10}	F: AGTCGGATGGAGTAGATAA R: TCAGGCAGTAAGAATAAGAG	TD	207	5	0.61
cp02	T_{11}	F: TGTTCTTCTTCCGTTTCC R: TTATTGACCGATTTGTGC	TD	211	2	0.02
cp03	$(T_8)CG(A_{10})$	F: ATCCATAAACCAGCAAATC R: TTCTAAGCGAGACAAACA	TD	184	5	0.62
cp04	A_{10}	F: GTTTGGCAAGCTGCTGTA R: GACTCTAATCTAAATAGGGTTCT	TD	233	2	0.06
cp05	T_{11}	F: AAGCGACCAGTTCCAATC R: ATCTAGGCCAGAAGCAGA	TD	498	2	0.25
cp06	T_{12}	F: CTTTACTCTTTCTACCCATCA R: AATACGCCCAGTTGCTTC	TD	326	3	0.51
cp07	$(T_{10})C(T_6)$	F: TGTTCTTCTTGGCATCTA R: ATACTCCCTGTTTGTTCC	TD	481	6	0.63
cp08	T_{10}	F: GGCCTAGAAACGGAAGA R: AATAGAGCGGCAGGACAA	TD	396	4	0.56
cp09	T_{12}	F: CAATACAGACGTGGTGAT R: GAGCAGTAACGAGGGAGT	TD	439	5	0.62
cp10	$(T_{14})AG(T_9)$	F: GGTTCTTTAGCGGGTTTA R: GAGTTTCGGTGAGGTTGA	TD	254	7	0.58
cp11	T_{10}	F: GAGAAATGTGGCTGGAAC R: CTGGAAATACCCGTTGAT	TD	380	6	0.66
cp12	T_{10}	F: CACATAATGAGGCGAGAA R: GCTGAATCACAGACGAAA	TD	285	5	0.61
cp13	A_{10}	F: AAGCAAGGCATTTCTGGT R: TCTTACGGAGCCCTATTT	TD	433	5	0.58
cp14	T_{10}	F: ATAGGCTGGTTCGCTTGA R: TGCTTTATGGGTCGTCTT	TD	335	4	0.56
cp15	T_{10}	F: ACCGTAGGGCAGGAGGAGT R: AATGACAGACCGGGAGGC	TD	162	2	0.11
cp16	T_{10}	F: GCATTTCGTATAGAGGTT R: TAAGACGAGACCCAAGTA	TD	203	3	0.45
cp17	A_{11}	F: CAATGTAACAAAGGACGAA R: TTATGAAATGAGCGGAGT	TD	214	3	0.40
cp18	T_{12}	F: CCGAAGAGTAACTAGGACA R: GAATGATGGAAGGGAATA	TD	333	4	0.54
cp19	A_{13}	F: ACCGTCACCCATTCTAAC R: ACTTTCTTGATTCCCATC	TD	322	5	0.67
cp20	A_{10}	F: GGGAAAATCGCATAAAAT R: AAAACTGGGCGTAGATAG	TD	253	6	0.65

Ta = annealing temperature; TD = touchdown PCR program with a temperature from 60°C, decreasing by 1°C per cycle until 50°C; N_A = number of alleles; PIC = polymorphism information content.

Polymorphism of the 20 cpSSR primer pairs was tested across all samples. A total of 85 alleles were detected in 66 *Gossypium* germplasm accessions. The number of alleles per locus ranged from 2-7, with an average of 4.20 per locus. Polymorphism information content (PIC) varied from 0.02-0.66, with a mean of 0.48 (Table 3).

In this study, the transferability of the 20 polymorphic cpSSR primers was tested in other 31 *Gossypium* species. Sixteen cpSSR primer pairs successfully amplified across all species tested, while the remaining 4 primer pairs (cp07, cp11, cp12, and cp20) produced amplification products across most species (Table 4). The results showed that all cpSSR markers had good transferability in the *Gossypium* species. All of these polymorphic cpSSR markers from *G. hirsutum* were transferable to other *Gossypium* species, suggesting that these cpSSR markers could be applied to evaluating the genetic diversity of *Gossypium*.

Table 4. Cross-amplification information in 32 *Gossypium* species for 20 polymorphic cpSSR markers developed in *Gossypium hirsutum*.

Species	cp01	cp02	cp03	cp04	cp05	cp06	cp07	cp08	cp09	cp10	cp11	cp12	cp13	cp14	cp15	cp16	cp17	cp18	cp19	cp20
G. hirsutum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. barbadense	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. tomentosum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. mustelinum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. darwinii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. herbaceum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. arboreum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. anomalum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. capitis-viridis	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. sturtianum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. nandewarense	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. robinsonii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. thurberi	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. armourianum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. harknessii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. davidsonii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. klotzschianum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. aridum	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+	+
G. raimondii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. gossypioides	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
G. lobatum	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
G. trilobum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. laxum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. schwendimanii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. stocksii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
G. somalense	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+	-
G. areysianum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. incanum	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. longicalyx	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. bickii	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
G. australe	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
G. nelsonii	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-

(+) PCR amplification; (-) no PCR amplification.

DISCUSSION

In the present study, we found 60 cpSSRs in the *G. hirsutum* chloroplast genome and analyzed the characteristics of their distribution. We identified cpSSRs throughout the whole chloroplast genome, but they were not evenly distributed, appearing to be concentrated in several areas. This result was consistent with those found in other plants (Wang et al., 2006; Pian et al., 2008; Hu et al., 2009; Pan et al., 2014). In addition, there were clear differences in the cpSSR number in different plants. The plant chloroplast genome size is generally 120-160 kb. The chloroplast genome sizes in *A. thaliana*, *C. sativus*, *M. truncatula*, *V. unguiculata*, and *G. hirsutum* are 154, 155, 124, 152, and 160 kb, and they contain 84, 227, 241, 52, and

60 cpSSRs, respectively. Therefore, no significant positive correlation was found between chloroplast genome size and cpSSR number.

The PIC value of a marker provides an estimation of the discrimination power in accessions by considering the number of alleles and the relative frequency of each allele (Smith et al., 2000). This helps in choosing markers for evaluating germplasms. In this study, 14 (70% of 20) cpSSR markers were highly informative (PIC > 0.5), 3 (15% of 20) cpSSR markers were moderately informative (0.25 < PIC < 0.5), and 3 (15% of 20) cpSSR markers were lowly informative (PIC < 0.25) (Vigouroux et al., 2002). The number of alleles per locus in this study ranged from 2-7 (mean = 4.25), with a mean PIC value of 0.48. Compared with the alleles and PIC values in other plants described in previous studies, such as those by Yang et al., 2008 (alleles = 2-6); Ginwal et al., 2011 (alleles = 2-3); Wang, 2013 (alleles = 1-3), this higher number of alleles and PIC value were not unexpected, as the *Gossypium* accessions were natural populations, and the genetic bottleneck caused by human-mediated selection pressure to accumulate a particular type of alleles governing various traits has not yet been applied (Iqbal et al., 2001; Vigouroux et al., 2002).

Because the chloroplast genome is highly conservative, cpSSR primers show good transferability between different species (Diekmann et al., 2012; Xue et al., 2012). In this study, transferability of cpSSR primers was tested using 32 cotton species. Twenty polymorphic cpSSR markers developed from *G. hirsutum* were transferable to all or most of the 32 *Gossypium* species because the flanking regions of cpSSR loci are conserved. These results suggest that all novel cpSSR markers can be used to analyze genetic diversity and population structure in *Gossypium*.

ACKNOWLEDGMENTS

Research supported by the National High Technology Research and Development of China (#2013AA102601) and Projects in the National Science and Technology Pillar Program (#2013BAD01B03).

REFERENCES

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27: 573-580.

Deguilloux MF, Dumolin-Lapègue S, Gielly L, Grivet D, et al. (2003). A set of primers for the amplification of chloroplast microsatellites in *Quercus*. *Mol. Ecol. Notes* 3: 24-27.

Diekmann K, Hodkinson TR and Barth S (2012). New chloroplast microsatellite markers suitable for assessing genetic diversity of *Lolium perenne* and other related grass species. *Ann. Bot.* 110: 1327-1339.

Dzialuk A, Muchewicz E, Boratyński A, Montserrat JM, et al. (2009). Genetic variation of *Pinus uncinata* (Pinaceae) in the Pyrenees determined with cpSSR markers. *Plant Syst. Evol*. 277: 197-205.

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

Ginwal HS, Mittal N, Tomar A and Varshney VK (2011). Population genetic structure and diversity of high value vulnerable medicinal plant *Acorus calamus* in India using RAPD and chloroplast microsatellite markers. *J. Forest. Res*. 22: 367-377.

Hu JB, Zhou XY and Li JW (2009). Development of novel chloroplast microsatellite markers for *Cucumis* from sequence database. *Biol. Plantarum* 53: 793-796.

Iqbal MJ, Reddy OUK, El-Zik KM and Pepper AE (2001). A genetic bottleneck in the 'evolution under domestication' of Upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet*. 103: 547-554.

Ishii T and McCouch SR (2000). Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. *Theor. Appl. Genet*. 100: 1257-1266.

Li Q and Wan JM (2005). SSRHunter: Development of a local searching software for SSR sites. *Yi Chuan* 27: 808-810.

Lu HJ and Myers GO (2002). Genetic relationships and discrimination of ten influential upland cotton varieties using RAPD markers. *Theor. Appl. Genet*. 105: 325-331.

Pan L, Li Y, Guo R, Zhang FY, et al. (2014). Development of cpSSR markers in *Vigna unguiculata* and their transferability in related species. *J. ChangJiang Vegetables* 6: 9-15.

Pian RQ, Li W, Li N, Chen XY, et al. (2008). Development of microsatellite from complete sequence of *Medicago truncatula* chloroplast DNA. *J. Anhui Agric. Sci*. 36: 3531-3534.

Pillay M and Myers GO (1999). Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. *Crop Sci*. 39: 1881-1886.

Powell W, Morgante M, McDevitt R, Vendramin GG, et al. (1995). Polymorphic simple-sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc. Natl. Acad. Sci. U. S. A*. 92: 7759-7763.

Rana MK, Singh VP and Bhat KV (2005). Assessment of genetic diversity in upland cotton (*Gossypium hirsutum* L.) breeding lines by using amplified fragment length polymorphism (AFLP) markers and morphological characteristics. *Genet. Resour. Crop Evol*. 52: 989-997.

Raubeson LA, Peery R, Chumley TW, Dziubek C, et al. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8: 174.

Rose O and Falush D (1998). A threshold size for microsatellite expansion. *Mol. Biol. Evol*. 15: 613-615.

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, et al. (2000). Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Sci*. 40: 226-232.

Sugiura N, Kurokochi H, Tan E, Asakawa S, et al. (2014). Development of 13 polymorphic chloroplast DNA markers in *Quercus gilva*, a regionally endemic species in Japan. *Conserv. Genet. Resour*. 6: 961-965.

Tatineni V, Cantrell RG and Davis DD (1996). Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. *Crop Sci*. 36: 186-192.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, et al. (2002). Rate and pattern of mutation at microsatellite loci in maize. *Mol. Biol. Evol*. 19: 1251-1260.

Wang HK, Qiao YS, Lou XM, Xue HB, et al. (2006). Distribution of microsatellite from complete sequence of *Arabidopsis thaliana* chloroplast DNA. *Chin. J. Biochem. Mol. Biol*. 22: 845-850.

Wang KB (2007). Introduction and conservation of wild cotton in China. *Cotton Sci*. 19: 354-361.

Wang KB, Wang ZW, Li FG, Ye WW, et al. (2012). The draft genome of a diploid cotton *Gossypium raimondii. Nat. Genet*. 44: 1098-1104.

Wang YL (2013). Chloroplast microsatellite diversity of *Opisthopappus Shih* (Asteraceae) endemic to China. *Plant Syst. Evol*. 299: 1849-1858.

Wendel JF and 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. *Syst. Bot*. 17: 115-143.

Wendel JF and Brubaker CL (1993). RFLP diversity in *Gossypium hirsutum* L. and new insights into the domestication of cotton. *Am. J. Bot*. 80: 71.

Wendel JF, Rowley R and Stewart J McD (1994). Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst. Evol*. 192: 49-59.

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

Xu DH, Abe J, Gai JY and Shimamoto Y (2002). Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theor. Appl. Genet*. 105: 645-653.

Xue J, Wang S and Zhou SL (2012). Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am. J. Bot*. 99: 240-244.

Yang J, Li S, Sun G, Yuan Y, et al. (2008). Population structure and genetic variation in the genus *Dipteronia* Oliv. (Aceraceae) endemic to China as revealed by cpSSR analysis. *Plant Syst. Evol*. 272: 97-106.

Yu JZ, Fang DD, Kohel RJ, Ulloa M, et al. (2012). Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. *Euphytica* 187: 203-213.

Zhang JF and Stewart JM (2000). Economical and rapid method for extracting cotton genomic DNA. *J. Cotton Sci*. 4: 193-201.