Theor Appl Genet (2015) 128:1531–1540
DOI 10.1007/s00122-015-2528-7

ORIGINAL PAPER

Development of *Gossypium anomalum*-derived microsatellite markers and their use for genome-wide identification of recombination between the *G. anomalum* and *G. hirsutum* genomes

Caijiao Zhai[1,2] · Peng Xu[1,2] · Xia Zhang[1,2] · Qi Guo[1,2] · Xianggui Zhang[1,2] · Zhenzhen Xu[1,2] · Xinlian Shen[1,2]

Received: 5 February 2015 / Accepted: 27 April 2015 / Published online: 9 May 2015

Abstract

Key Message **We reported the first development of *Gossypium anomalum*-derived microsatellite markers and identification of recombination between sexually incompatible species by a synthesized hexaploid on genome level.**

Abstract To continue to develop improved cotton varieties, it is essential to transfer desired characters from diploid wild cotton species such as *Gossypium anomalum* to cultivated allotetraploid cotton species. However, interspecific reproductive barriers limit gene transfer between species. In a previous study, we used colchicine treatment to produce a synthesized hexaploid derived from an interspecific hybrid between *Gossypium hirsutum* and *G. anomalum* and demonstrated its hybridity and doubled status using morphological, cytological and molecular marker methods. In the current study, to effectively monitor *G. anomalum* genome components in the *G. hirsutum* background, we developed 5974 non-redundant *G. anomalum*-derived SSR primer pairs using RNA-Seq technology, which were combined with a publicly available physical map. Based on this combined map and segregation data from the BC_2F_1 population, we identified a set of 230 informative *G. anomalum*-specific SSR markers distributed on the chromosomes, which cover 95.72 % of the cotton genome. After analyzing BC_2F_1 segregation data, 50 recombination types from 357 recombination events were identified, which cover 81.48 % of the corresponding *G. anomalum* genome. A total of 203 recombination events occurred on chromosome 11, accounting for 56.86 % of the recombination events on all chromosomes. Recombination hotspots were observed at marker intervals JAAS1148-NAU5100 on chromosome 1 and JAAS0426-NAU998 on chromosome 2. Therefore, all *G. anomalum* chromosomes are capable of recombining with At chromosomes in *G. hirsutum*. This study represents an important step towards introgressing desirable traits into cultivated cotton from the wild cotton species *G. anomalum*.

Introduction

The genus *Gossypium* is composed of four cultivated species (*Gossypium hirsutum* L., *Gossypium barbadense*, *Gossypium herbaceum* and *Gossypium arboreum*) and 46 wild species representing AA, BB, CC, DD, EE, FF, GG, KK and AADD genome types (Fryxell 1992). As in other crops, the domestication of cotton has led to a dramatic loss of allelic diversity. The impoverished cotton gene pool has been further eroded by the over-exploitation of a few genetic backgrounds (Paterson et al. 2004). Wild species of *Gossypium* present an impressive range of variation in many characters, all of which can potentially be exploited in cotton improvement programs. Transferring desired characters from wild species into cultivated cotton is therefore important for the continued improvement

Communicated by P. Heslop-Harrison.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2528-7) contains supplementary material, which is available to authorized users.

✉ Xinlian Shen
xlshen68@126.com

[1] Key Laboratory of Cotton and Rapeseed (Nanjing), Ministry of Agriculture, Nanjing, People's Republic of China

[2] The Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, People's Republic of China

of cotton varieties. Several important traits of wild species have been successfully introduced into cultivated cotton varieties via wide hybridization. Some of these exotic sources have made significant contributions to cotton genetic improvement, including *Gossypium harknessii* as a source of cytoplasmic male sterility (Meyer 1973), *Gossypium thurberi* as a source of improved fiber quality traits (Culp and Harrell 1973; Culp et al. 1979) and *Gossypium sturtianum* as a source of low-gossypol and high-gossypol plant traits (Benbouza et al. 2010). Despite these successes, most genetic variations in wild *Gossypium* species remain untapped.

Gossypium anomalum ($2n = 2x = 26$, B_1) is a member of the subsection *Anomala Todaro*. This species grows in the southern borders of the Great Desert Belt and dry coastal belt in Angola and South-West Africa (Silow 1941). *G. anomalum* possesses several desirable characters, such as resistance to bacterial diseases, cotton wilt, aphids, black arm and jassids, as well as high lint quality and water deficit tolerance (as this species is endemic to arid steppe regions) (Ganesh et al. 2013). High fiber quality traits and immunity to bacterial blight were previously introgressed into *G. hirsutum* from *G. anomalum* (Qian et al. 1992; Mehetre 2010). However, the genomic differences between tetraploid Upland cotton $(AD)_1$ and diploid *G. anomalum* (B_1B_1) cause serious interspecific reproductive barriers, which limit gene transfer between the species. As a result, little is known about the introgression of DNA from this species into cultivated varieties at the genome-wide level (Zhang et al. 2014).

SSR markers, also known as microsatellites or short tandem repeats (STRs), are PCR-based markers. SSR markers are randomly distributed tandem repeats of short nucleotide motifs (usually 2–6 bp long). Due to their hypervariability, reproducibility, high abundance, multiallelic nature and codominant inheritance, SSR markers have become valuable, reliable tools for genetic diversity analysis, genetic mapping and gene tagging in cotton (reviewed by ur Rahman et al. 2009). To date, all SSR primers used in cotton studies have been based on sequences from *G. arboreum* (A_1 genome), G. *herbaceum* (A_2 genome), *Gossypium raimondii* (D_5 genome) and *G. hirsutum* (A_1D_1 genome) (http://www.cottonmarker.org). Due to their high levels of transferability among the *Gossypium* species, these markers have been used to identify exotic chromosome fragments in the *G. hirsutum* background (Benbouza et al. 2010; Chen et al. 2014). However, in a previous study involving the synthetic hexaploid derived from *G. hirsutum* and *G. anomalum* (Zhang et al. 2014), we identified some SSR primer pairs showing weak or failed amplification of *G. anomalum*-specific SSR fragments in hybrid plants and BC segregation populations but normal amplification in *G. anomalum*. We hypothesized that the sequence discrepancy between different species results in the preferential binding of SSR primers to genomic DNA from *G. hirsutum*. As a result, these specific SSR loci of *G. anomalum* would be undetectable in the hybrid plants. This phenomenon would make these polymorphic markers useless for monitoring *G. anomalum*-specific chromosome segments in backcross generations. Therefore, developing *G. anomalum*-derived SSR markers would greatly facilitate identification of *G. anomalum* genome components in a *G. hirsutum* background.

In a previous study, we obtained triploid hybrids with the genome composition $A_1D_1B_1$ by crossing G. *hirsutum* $(AD)_1$ with *G. anomalum* (B_1B_1), and we produced a putative fertile hexaploid ($A_1A_1D_1D_1B_1B_1$) via colchicine treatment. We confirmed the nature of the hexaploid plants using morphological, cytological and molecular methods (Zhang et al. 2014). The objectives of the current study were (1) to develop *G. anomalum*-derived SSR markers; (2) to identify a set of *G. anomalum*-specific SSR markers that are evenly distributed on the chromosomes; and (3) to verify the occurrence of recombination between the *G. anomalum* and *G. hirsutum* genomes at the genome-wide level.

Materials and methods

Plant materials

Plants of the wild *Gossypium* species *G. anomalum*, *G. hirsutum* var. 86-1 plants and synthesized hexaploid plants derived from an interspecific hybrid between *G. hirsutum* var. 86-1 and *G. anomalum* were grown at Lishui Plant Experiment Station, Jiangsu Academy of Agricultural Sciences, China. In 2010 and 2011, four hexaploid plants were backcrossed as females to *G. hirsutum* var. Su8289, yielding 503 BC_1F_1 (pentaploid) seeds, 97 of which produced viable plants in 2012. These 97 BC_1F_1 plants were again backcrossed as females to Su8289, and more than 2400 BC_2F_1 seeds were obtained, 384 of which produced viable plants in 2013.

Microsatellite mining of *G. anomalum* transcriptome data

Total RNA was isolated from frozen leaves of *G. anomalum* using the cold-acidic phenol method with a modified extraction buffer (10 mM Tris–HCl, pH 8.0, 25 mM EDTA, pH 8.0, 2 % CTAB, 2 % PVP). The RNA was then precipitated with ethanol, dissolved in DEPC-treated water and stored at −70 °C. The RNA samples were quantified and examined spectrophotometrically for protein contamination (A_{260}/A_{280} ratio) and reagent contamination

(A_{260}/A_{230} ratio) prior to library construction. The cDNA library construction and sequencing with an Illumina HiSeq™ 2000 were performed at NovoGene Company (Beijing, China). After removing sequence reads containing adapter sequences, low-quality sequences (reads with ambiguous bases 'N') and reads with more than 10 % Q\20 bases, de novo assembly of the clean reads was performed using the Trinity program (http://trinityrnaseq.sourceforge.net/).

The unigene dataset was mined for microsatellite repeats using a Perl script known as MISA (http://pgrc.ipk-gatersleben.de/misa). The *G. anomalum*-derived SSR markers in this work were designated "JAASXXXX", with "JAAS" representing Jiangsu Academy of Agricultural Sciences and "XXXX" replaced with the serial number of the SSR marker.

SSR primer design and redundancy analysis

SSR primers were designed using Primer 3.0. The major parameters for designing PCR primers were as follows: (1) primer length ranging from 18 to 27 bases; (2) PCR product size ranging from 100 to 280 bp; (3) melting temperature between 55 and 65 °C, with 60 °C being the optimum annealing temperature; and (4) a GC content of 45–65 %, with an optimum of 50 %.

Microsatellite sequences downloaded from the Cotton Marker Database (CMD) (http://www.cottonmarker.org/) served as templates. The program e-PCR (http://www.ncbi.nlm.nih.gov/projects/e-pcr/) was utilized for simulated PCR amplification. If there were PCR products, the newly designed primers were considered to be redundant with the original SSR primers in CMD, regardless of whether the products sizes were consistent.

Combining *G. anomalum*-derived markers with the whole-genome marker map

After the marker sequences were prepared, BLASTN was applied to anchor *G. anomalum*-derived markers to the cotton D-genome pseudo molecules as described by Wang et al. (2013). Markers with alignments of *E* value $\leq 1e^{-10}$ for SSR sequences were assembled into loci. The genetic distances of *G. anomalum*-derived SSR loci were estimated according to the genetic distance of the nearest marker in the whole-genome marker map (WGMM) (Wang et al. 2013).

SSR analysis

Genomic DNA from hexaploid plants, the BC_2F_1 population and their parents was extracted as described by Paterson et al. (1993). SSR primers were selected using the following two steps: First, 707 SSR primers pairs covering the whole cotton genome were selected based on the cotton genetic map (Guo et al. 2007) and marker chromosome location information (Xiao et al. 2009) as previously described (Zhang et al. 2014). The sequences of these primers are available from the Cotton Marker Database (CMD) website (http://www.cottonmarker.org). Second, after combining the *G. anomalum*-derived markers with the whole-genome marker map, *G. anomalum*-derived SSR markers were selected at approximately 3 centiMorgan (cM) intervals. SSR analysis was conducted according to Zhang et al. (2000).

Results

Frequency and distribution of different types of SSR markers

The *G. anomalum* cDNA library was sequenced by the Illumina HiSeqTM 2000 system, yielding a total of 46.8 million raw reads. After cleaning and quality checks, approximately 45.6 million clean reads were obtained. These short reads were assembled, resulting in 117,845 transcripts. The mean transcript size was 1393 bp, with lengths ranging from 201 to 16,902 bp. A total of 53,094 unigenes were obtained, with a mean size of 864 bp.

Using a Perl script known as MISA (http://pgrc.ipk-gatersleben.de/misa), 13,963 microsatellite repeats were detected from 53,094 non-redundant unigenes. Of these unigenes harboring microsatellites, 11,047 unigene sequences contained one microsatellite repeat, while 2297 contained more than one microsatellite repeat, and 855 microsatellite repeats were present in compound form. The overall microsatellite density was found to be one microsatellite spanning every 3.3 kb.

The most abundant type of repeat motif was mononucleotides (9546, 68.37 %), followed by dinucleotides (2173, 19.43 %), trinucleotides (2079, 14.89 %), tetranucleotides (136, 0.9 %), pentanucleotides (14, 0.1 %) and hexanucleotides (15, 0.1 %). We calculated the frequencies of EST-SSRs with different numbers of repeat units (Supplementary Table S1). Within the detected SSRs, 125 motif sequence types were identified, including 28, 22, 43, 32 and 9 types of mono-, di-, tri-, tetra- and pentanucleotide repeats, respectively. Among these, A/T (64.68 %) represented the dominant type, followed by AT/AT (7.78 %), AG/CT (5.87 %) and AAG/CTT (4.66 %).

Combining *G. anomalum*-derived SSRs with the publicly available whole-genome marker map

A total of 6656 SSR primer pairs were designed based on the 13,963 SSR loci using Primer3 software

Table 1 Number and coverage of *G. anomalum*-derived SSR markers on the chromosomes of *G. raimondii* (D_5) genome

Chr.	SSR marker	Cover length (Mb)	Chr. length (Mb)[a]	Coverage (%)	Density (marker/Mb)
1	391	62.65	63.55	98.58	6.2
2	426	63.99	64.94	98.54	6.6
3	307	45.69	46.34	98.60	6.6
4	272	35.37	35.87	98.61	7.6
5	737	70.64	71.60	98.66	10.3
6	390	62.09	62.95	98.63	6.2
7	435	55.42	56.57	97.98	7.7
8	461	62.04	62.96	98.54	7.3
9	418	50.86	51.71	98.36	8.1
10	366	62.54	63.46	98.55	5.8
11	551	60.79	61.74	98.46	8.9
12	489	56.98	57.84	98.53	8.5
13	402	58.3	59.05	98.73	6.8
Total	5645	747.26	758.58	98.50	7.4

[a] Chr. length in Mb based on Paterson et al. (2012)

(Supplementary Table S2). Subsequently, all publicly available microsatellite sequences (17,343 items) were downloaded from the Cotton Marker Database (CMD) (http://www.cottonmarker.org/). We used the program e-PCR (http://www.ncbi.nlm.nih.gov/projects/e-pcr/) to simulate PCR amplification. The results show that 682 primer pairs were redundant with SSR primers in CMD. Ultimately, a total of 5974 non-redundant *G. anomalum*-derived SSR primer pairs were developed. These 5974 EST sequences were mapped to the D-genome with BLASTN, and 5557 primer sequence corresponding 5645 loci were placed on this reference sequence. A combined marker map was then generated (Supplementary Table S3).

The number of *G. anomalum*-derived SSR markers on the chromosomes of the combined marker map ranged from 272 to 737 per chromosome, which covered approximately 98.5 % of the physical length of the whole cotton D genome. The marker density along each chromosome ranged from 5.8 to 10.3 markers per Mb, with an average of 7.4 markers per Mb (Table 1).

Genome-wide screening of a set of informative *G. anomalum*-specific SSR loci based on BC_2F_1 segregation data

We first selected 707 public SSR primer pairs covering the cotton genome to amplify SSR markers from three parents and four hexaploid hybrid plants. Although all selected SSR markers exhibited high levels of transferability and polymorphism, almost half of the markers (47.24 %) were dominant in *G. hirsutum*. Since *G. hirsutum* was used as the recurrent parent in the backcrossing population, these dominant markers in *G. hirsutum* cannot be used to monitor the introgression of *G. anomalum*-specific segments in the backcross population (Zhang et al. 2014). Moreover, among the 349 informative SSR markers (333 codominant loci and 16 dominant loci in *G. anomalum*), a considerable proportion of SSR primers failed to produce clear amplification products in the BC_2F_1 population. Only 93 well-amplified microsatellite markers were selected to genotype the entire BC_2F_1 population.

After developing *G. anomalum*-derived SSR primers and combining them with the whole-genome marker map (Supplementary Table S3), we selected 777 SSR markers (approximately one marker/3 cM), including 642 *G. anomalum*-derived SSR primer pairs and 135 primer pairs from other publicly available markers, to analyze the three parents and four hexaploid hybrid plants. Among the 642 *G. anomalum*-derived SSR primers, 549 produced polymorphic bands between *G. hirsutum* and *G. anomalum*, including 344 codominant markers, 85 dominant markers in *G. anomalum* and 120 dominant markers in *G. hirsutum*. Of the 135 remaining pairs of publicly available markers, 60 informative markers (55 codominant loci and 5 dominant loci in *G. anomalum*) were identified. Based on the combined map information (Supplementary Table S3), 155 evenly distributed, well-amplified, informative microsatellite markers (including 133 *G. anomalum*-derived SSR primers) were selected to genotype the entire BC_2F_1 population.

Segregation data from the above 248 SSR loci were used to conduct linkage analysis via MapMaker/EXP 3.0b (Lander et al. 1987), revealing 243 SSR loci in 13 linkage groups; 13 SSR loci were not mapped to the corresponding chromosomes, as they were detected in the combined marker map, while 5 were not linked to any of the 13

Table 2 Distribution of *G. anomalum*-specific SSR markers on chromosomes

Chr.	SSR markers	Range of genetic distance between neighbor markers (cM)[a]	Average genetic distance between neighbor markers (cM)	Cover length (cM)[b]	Chr. length (cM)[c]	Coverage (%)
1	20	0–26.4	9.2	175.5	176.4	99.49
2	16	3.7–20.5	11.5	172.6	191.5	90.13
3	13	4.3–19.3	9.7	116.8	118.7	98.40
4	12	0–18.8	9.2	100.9	101.6	99.31
5	24	0.8–35.8	11.2	256.4	263.9	97.16
6	18	0.3–26.0	9.4	160.1	167.8	95.41
7	19	0.5–23.0	10.6	185.0	195.4	94.68
8	18	3.3–23.1	10.7	181.8	183.5	99.07
9	15	4.1–30.1	12.2	170.6	172.6	98.84
10	16	0.4–19.4	9.5	142.5	165.5	86.10
11	27	0.1–24.6	7.8	201.5	216.2	93.20
12	15	3.6–21.6	11.3	157.5	159.5	98.75
13	17	4.4–29.5	11.9	190.6	198.1	96.21

[a] If two neighboring loci have different genotypes, half of the intervals between them are considered to be derived from one parent and the other half from the other parent (Young and Tanksley 1989)

[b] Covered length of *G. anomalum*-specific markers on chromosomes based on the combined marker map (Supplementary Table S3)

[c] Chr. length in cM cited from the whole-genome marker map (WGMM) (Wang et al. 2013)

linkage groups. In order to determine genome attribution of these 13 linkage groups, we selected 18 SSR markers from genetic map (Guo et al. 2007) with known fragment size on particular At and Dt chromosomes to genotype parents (*G. anomalum*, 86-1 and Su8289), their hexaploid progenies and TM-1 as bridge cultivar. By comparison of molecular weight of polymorphic loci in above materials, we found that markers from At chromosomes amplified the same size of the fragments in TM-1 as Guo et al. (2007) reported; however, markers from Dt chromosome amplified different size of fragment from TM-1(data not shown). In this way, we assumed that these 13 linkage groups were belonged to At subgenome and assigned to chromosomes 1–13. The results indicated that recombination between the At subgenome of *G. hirsutum* and the B_1 genome is favored. However, we cannot rule out the possibility of recombination between the Dt subgenome of *G. hirsutum* and the B_1 genome because there are five loci unlinked to any linkage group.

To better evaluate the distribution of the marker loci on chromosomes, only 230 marker loci with positional information on chromosomes were included in Fig. 1 and Table 2. Mapped marker loci were not assayed based on BC_2F_1 segregation data since severe recombination suppression was observed. These loci were therefore placed at their approximate positions based on the combined marker map (Supplementary Table S3).

The number of *G. anomalum*-specific marker loci per chromosome varied from 12 to 27. The mean distance between the adjacent marker loci varied from 0–35.8 cM, with an average of 10.5 cM. The coverage of *G. anomalum*-specific SSR marker loci on every chromosome varied from 86.1 % (chromosome 10) to 99.49 % (chromosome 1), with an average of 95.72 %. Although we tried to obtain evenly distributed *G. anomalum*-specific marker loci, two gaps of more than 30 cM were still present on chromosomes 5 and 9 (Fig. 1).

Identifying recombination events in the BC_2F_1 population

A total of 357 recombination events (50 recombination types) were identified among the 384 BC_2F_1 individuals. Although recombination suppression was generally observed, the frequency of recombination events varied greatly between individual chromosomes, ranging from one (chromosome 3 and chromosome 4) to 203 (chromosome 11). A high frequency of recombination events was observed in the marker regions JAAS1148-NAU5100 on chromosome 1 and JAAS0426-NAU998 on chromosome 2 (Table 3).

Based on the combined marker map (Supplementary Table S3), we estimated the size of each introgressed segment and the ratios they accounted for compared to the whole donor genome (Table 3). The estimated donor segment length ranged from 15.4 to 235.05 cM. The donor genome was represented on between 43.31 % (chromosome 4) and 99.37 % (chromosome 12) of each chromosome. The total length of the introgressed *G. anomalum* genome was 1882.7 cM, covering 81.48 % of the donor genome.

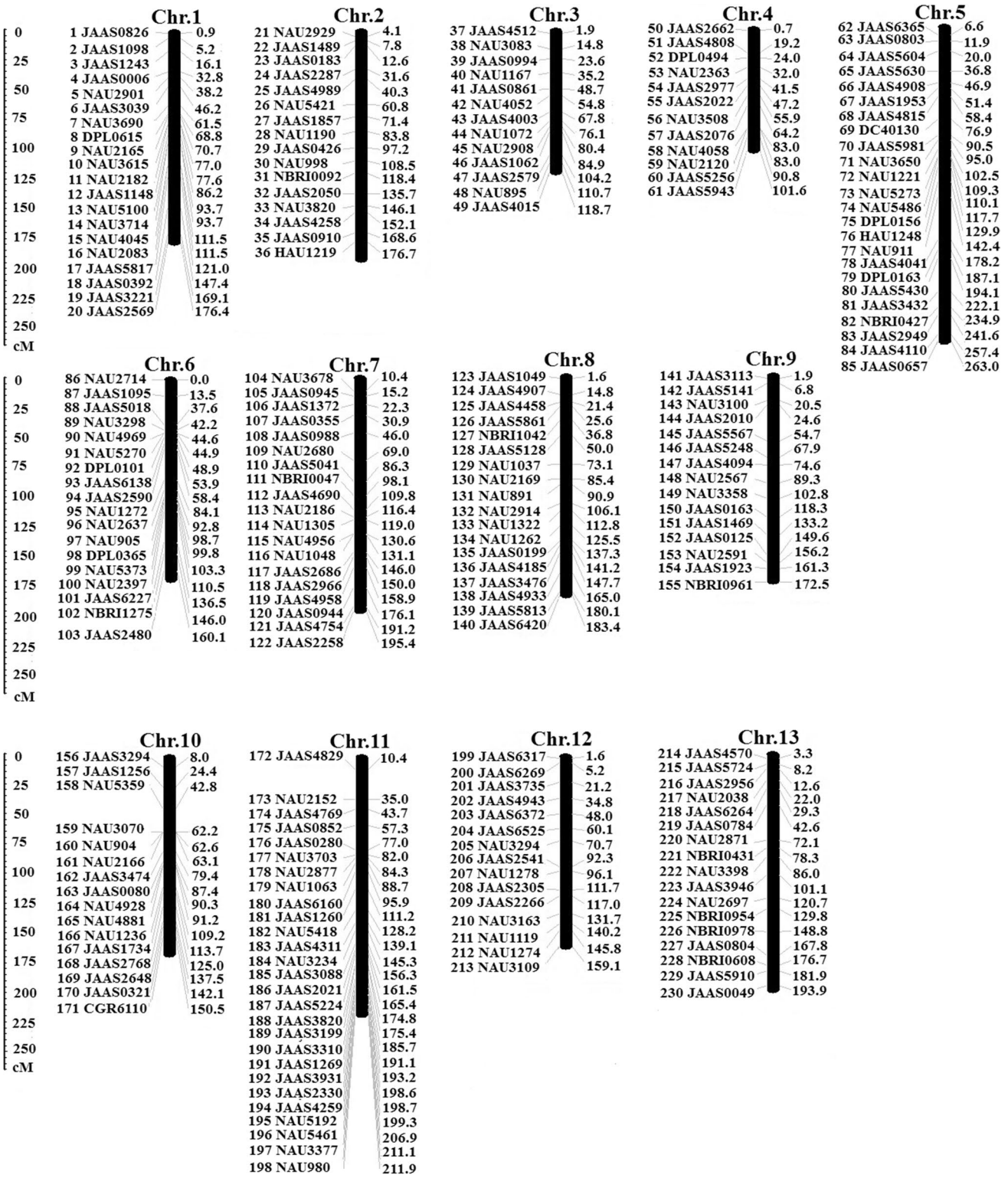

Fig. 1 Map of *G. anomalum*-specific SSR loci based on the combined map information. Mapping of *G. anomalum*-specific SSR loci was performed using MapInspect (http://www.plantbreeding.wur.nl/UK/software_mapinspect.html). The *scales* represent centiMorgans (cM)

Table 3 Estimated size and coverage of *G. anomalum*-introgressed segments in the BC_2F_1 generation

Chr.	Recombination type	Introgression segments	No. of recombination events	Size of introgression segments (cM)	Cover length (cM)	Coverage/ chromosome (%)
Chr1	1	JAAS0392-JAAS2569 (18–20) [a]	1	42.20	134.20	76.08
	2	NAU5100-JAAS2569 (13–20)	45	86.45		
	3	NAU3615-JAAS2569 (10–20)	1	102.55		
	4	JAAS3039-JAAS2569 (6–20)	2	134.20		
Chr2	5	NAU2929-JAAS1857 (21–27)	2	75.55	182.05	95.07
	6	NAU2929-JAAS0426 (21–29)	60	100.80		
	7	NBRI0092-HAU1219 (31–36)	3	70.65		
	8	NAU1190-JAAS4258 (28–34)	1	82.75		
Chr3	9	JAAS4512-JAAS4003 (37–43)	1	71.00	71.00	59.81
Chr4	10	JAAS2662-JAAS2977 (50–54)	1	44.00	44.00	43.31
Chr5	11	JAAS6365-JAAS1953 (62–67)	2	51.60	260.15	98.58
	12	JAAS6365-DC40130 (62–69)	1	80.40		
	13	JAAS6365-HAU1248 (62–76)	1	132.85		
	14	JAAS5630-JAAS0657 (65–85)	1	235.05		
	15	DC40130-JAAS0657 (69–85)	1	195.80		
	16	JAAS5981-JAAS0657 (70–85)	1	179.75		
	17	NAU911-JAAS0657 (77–85)	3	127.30		
Chr6	18	NAU1272-JAAS2480 (95–103)	1	92.70	92.70	55.24
	19	DPL0365-JAAS2480 (98–103)	1	64.70		
Chr7	20	NAU3678-JAAS5041(104–110)	1	87.00	190.20	97.34
	21	NAU3678-NAU1305 (104–114)	1	119.60		
	22	JAAS2966-JAAS2258 (118–122)	1	47.40		
Chr8	23	JAAS1049-JAAS4907 (123–124)	1	17.30	94.65	51.58
	24	JAAS1049-NBRI1042 (123–127)	1×2^b	42.60		
		JAAS0199-JAAS6420 (135–140)		52.05		
Chr9	25	NAU3100-JAAS4094 (143–147)	1×2^b	68.30	124.80	72.31
		NAU3358-JAAS0163 (149–150)		29.70		
	26	JAAS3113-JAAS4094 (141–147)	2	81.00		
	27	JAAS3113-JAAS0163 (141–150	1	124.80		
Chr10	28	NAU5359-NAU4881(158–165)	2	66.60	127.25	76.89
	29	JAAS3294-JAAS2768 (156–168)	1	127.25		
Chr11	30	JAAS4259-NAU980 (194–198)	3	15.40	208.85	96.60
	31	JAAS3310-NAU980 (190–198)	14	33.50		
	32	JAAS4829-JAAS4769 (172–174)	1×2^b	45.30		
		JAAS3310-NAU980 (190–198)		33.50		
	33	JAAS3088-NAU980 (185–198)	6	63.25		
	34	JAAS4829-JAAS4769 (172–174)	5×2^b	45.30		
		JAAS3088-NAU980 (185–198)		63.25		
	35	JAAS0280-NAU980 (176–198)	6	146.90		
	36	JAAS4829-JAAS3199 (172–189)	30	175.35		
	37	JAAS4829-NAU3234 (172–184)	23	145.60		
	38	JAAS4829-JAAS4769 (172–174)	24	45.30		
	39	JAAS3088-JAAS3199 (185–189)	33	29.75		
	40	JAAS4829-JAAS4769 (172–174)	8×2^b	45.30		
		JAAS3088-JAAS3199 (185–189)		29.75		
	41	JAAS4829-NAU3234 (172–184)	18×2^b	145.60		
		JAAS3310-NAU980 (190–198)		33.50		

Table 3 continued

Chr.	Recombination type	Introgression segments	No. of recombination events	Size of introgression segments (cM)	Cover length (cM)	Coverage/ chromosome (%)
Chr12	42	JAAS6317-JAAS3735 (199–201)	1	27.20	158.50	99.37
	43	JAAS3735-JAAS6372 (201–203)	1	40.85		
	44	JAAS6317-JAAS6372 (199–203)	1	53.25		
	45	NAU1274-NAU3109 (212–213)	1	16.30		
	46	JAAS6525-NAU3109 (204–213)	1	105.25		
	47	JAAS6317-NAU1274 (199–212)	1	151.65		
Chr13	48	JAAS3946-JAAS0049 (223–230)	2	102.45	194.35	98.11
	49	JAAS0784-JAAS0049 (219–230)	2	160.05		
	50	JAAS4570-JAAS6264 (214–218)	1	34.30		

[a] Numbers in parentheses indicate introgression segments in markers (serial numbers shown in Fig. 1)

[b] The recombination type had two introgression segments on the target chromosome

Discussion

EST-SSR markers have high transferability among *Gossypium* species (Guo et al. 2006). The transferability rate from the three species (A_2, D_5, A_1D_1) to *G. anomalum* (B_1) is also high (as high as 98.0 %) (Zhang et al. 2014). However, in our previous study of the synthesized hexaploid derived from *G. hirsutum* and *G. anomalum*, some SSR primer pairs produced little or no *G. anomalum*-specific SSR amplification products in hybrid plants and the BC segregation population but normal amplification products in *G. anomalum*. We proposed that sequence discrepancy between different species resulted in the preferential binding of SSR primers to genomic DNA from *G. hirsutum* (Zhang et al. 2014). In the current study, we developed 5974 non-redundant *G. anomalum*-derived SSR primer pairs using RNA-Seq technology. Some *G. anomalum*-derived primers (such as JAAS0861, JAAS5817 and others) failed to produce *G. hirsutum*-specific bands in hexaploid plants, further indicating that SSR primers preferentially bind to the genomic DNA from which the SSR primers was developed.

G. anomalum belongs to the B_1 subgenome group. The general distribution of this species considerably overlaps with that of *G. arboreum* and *G. herbaceum*, two cultivated Asiatic cotton species (Silow 1941). The B genome is assumed to be the progenitor of the A genome. A hybrid between *G. anomalum* and *G. herbaceum* has 13 bivalents at metaphase I, whereas a hybrid between *G. anomalum* and *G. arboreum* has 11 bivalents and a quadrivalent, suggesting that the A genome and B_1 genome are closely related (Phillips 1966). This close affinity makes it possible to enhance genetic recombination with the At genome of *G. hirsutum*, as was observed in the present study. A total of 357 recombination events were identified between the B_1 and At genomes in the first segregation generation (BC_2F_1 population). Our data suggest that all *G. anomalum* chromosomes are capable of recombining with their counterpart chromosomes in *G. hirsutum*. However, the frequency of recombination events varied widely within and between chromosomes. A total of 203 recombination events occurred on chromosome 11, accounting for 56.86 % of the total number of recombination events for all chromosomes. Moreover, a high frequency of recombination events was observed on marker intervals JAAS1148-NAU5100 on chromosome 1 and JAAS0426-NAU998 on chromosome 2, corresponding to 5.74 and 18.93 recombination events per 1 million base pairs (Mb) of DNA, respectively. These frequencies are 7.36- and 19.72-fold higher than the average frequency for chromosomes 1 and 2, respectively. These two marker intervals are therefore considered to be recombination hotspots (defined as having recombination frequencies at least five times higher than the average frequency) (Katzer et al. 2011).

We observed severe recombination suppression on all chromosomes except chromosome 11. A significant reduction in genetic recombination of interspecific hybrids was also observed in classical genetic study of *Gossypium* based on phenotypic markers (Menzel and Brown 1955; Rhyne 1958, 1960). Both chromosomal differentiation and genetic factors that regulate recombination frequency throughout the genome are assumed to lead to changes in recombination frequencies in cotton (Rhyne 1958) and other crops (Sears and Okamoto 1958; Liharska et al. 1996). With the aid of species-specific molecular markers on a genome-wide scale, we were able to evaluate genetic recombination region-by-region for interspecific hybrids. This process, together with additional *Gossypium* sequencing efforts, will make it possible to further explore the molecular basis of the observed variability in recombination frequencies in the future.

To assess the potential for inter-genomic recombination on a genome-wide level, we analyzed 384 BC_2F_1 progeny derived from 97 BC_1 plants using a set of *G. anomalum*-specific SSR loci covering the whole *G. anomalum* genome. We obtained 50 recombination types from 357 recombination events, which covered 81.48 % of the *G. anomalum* genome. The occurrence of homeologous recombination provides evidence that the introgression of *G. anomalum* traits into *G. hirsutum* is possible between such sexually incompatible species in a synthesized hexaploid. However, approximately 47.1 % of the individuals in the first segregation generation (BC_2F_1) were sterile; no viable seeds were produced from 14 recombination types. Therefore, additional recombinants must be identified in later generations in order to obtain more recombination types to cover the *G. anomalum* genome. Nonetheless, the present study represents an important step towards introgressing desirable traits from wild cotton species into cultivated cotton.

Author Contribution statement Zhai CJ contributed to genotyping, data analysis and writing; Xu P contributed to SSR development and data analysis; Zhang X contributed to genotyping; Guo Q contributed to population construction; Zhang XG contributed to population construction; Xu ZZ contributed to data analysis; Shen XL contributed to experiment design and writing.

Acknowledgments This work was supported by grants from the National Natural Science Foundation of China (NSFC) (Grant Numbers 31471545, 31171595), the Jiangsu Independent Innovation Funds of Agricultural Technology [Grant Number CX (14)2065] and Jiangsu Collaborative Innovation Center for Modern Crop Production.

Conflict of interest The authors declare that there are no conflicts of interest in the reported research.

Ethical standard The authors note that this research was performed and reported in accordance with the ethical standards of scientific conduct.

References

Benbouza H, Lacape JM, Jacquemin JM, Courtois B, Diouf FBH, Sarr D, Konan N, Baudoin JP, Mergeai G (2010) Introgression of the *low-gossypol seed* and *high-gossypol plant* trait in upland cotton: analysis of [(*Gossypium hirsutum* × *G. raimondii*)^2 × *G. sturtianum*] trispecific hybrid and selected derivatives using mapped SSRs. Mol Breeding 25:273–286

Chen Y, Wang YY, Wang K, Zhu XF, Guo WZ, Zhang TZ, Zhou BL (2014) Construction of a complete set of alien chromosome addition lines from *Gossypium austral* in *Gossypium hirsutum*: morphological, cytological, and genotypic characterization. Theor Appl Genet 127:1105–1121

Culp TW, Harrell DC (1973) Breeding methods for improving yield and fiber quality of upland cotton (*Gossypium hirsutum* L). Crop Sci 13:686–689

Culp TW, Harrell DC, Kerr T (1979) Some genetic implications in the transfer of high fiber strength genes to upland cotton. Crop Sci 19:481–484

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

Ganesh SN, Vivek PC, Subhash SM, Ashok SJ (2013) Interspecific hybridization in *Gossypium* L.: characterization of progenies with different ploidy-confirmed multigenomic backgrounds. Plant Breeding 132:211–216

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

Guo WZ, Cai CP, Wang CB, Han ZG, Song XL, Wang K, Niu XW, Wang C, Lu KY, Shi B, Zhang TZ (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. Genetics 176:527–541

Katzer F, Lizundia R, Ngugi D, Blake D, McKeever D (2011) Construction of a genetic map for *Theileria parva*: identification of hotspots of recombination. Int J Parasitol 41:669–675

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liharska T, Koornneef M, Wordragen MV, Kammen AV, Zabel P (1996) Tomato chromosome6: effect of alien chromosomal segments on recombination frequencies. Genome 39:485–491

Mehetre SS (2010) Wild *Gossypium anomalum*: a unique source of fibre fineness and strength-overview and achievements of prebreeding efforts. Curr Sci 99:58–71

Menzel MY, Brown MS (1955) Isolating mechanism in hybrids of *Gossypium gossypioides*. Am J Bot 42:49–57

Meyer VG (1973) Fertility restorer genes for cytoplasmic male sterility from *Gossypium harknessii*.In: Proceedings of beltwide cotton production research conference, p 65

Paterson AH, Brubaker CL, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11:122–127

Paterson AH, Boman RK, Brown SM, Chee P, Gannaway JR, Gingle AR, May OL, Smith CW (2004) Reducing the genetic vulnerability of cotton. Crop Sci 44:1900–1901

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin DC, Llewellyn D, Showmaker KC, Shu SQ, Udall J, Yoo MJ, Byers R, Chen W, Doron-Faigenboim A, Duke MV, Gong L, Grimwood J, Grover C, Grupp K, Hu GJ, Lee T, Li JP, Lin LF, Liu T, Marler BS, Page JT, Roberts AW, Romanel E, Sanders WS, Szadkowski E, Tan X, Tang HB, Xu CM, Wang JP, Wang ZN, Zhang D, Zhang L, Ashrafi H, Bedon F, Bowers JE, Brubaker CL, Chee PW, Das S, Gingle AR, Haigler CH, Harker D, Hoffmann LV, Hovav R, Jones DC, Lemke C, Mansoor S, ur Rahman M, Rainville LN, Rambani A, Reddy UK, Rong JK, Saranga Y, Scheffler BE, Scheffler JA, Stelly DM, Triplett BA, Van Deynze A, Vaslin MF, Waghmare VN, Walford SA, Wright RJ, Zaki EA, Zhang TZ, Dennis ES, Mayer KF, Peterson DG, Rokhsar DS, Wang XY, Schmutz J (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 492:423–427

Phillips LL (1966) The cytology and phylogenetics of the diploid species of *Gossypium*. Amer J Bot 53:328–335

Qian SY, Huang JQ, Peng YJ, Zhou BL, Ying MC, Shen DZ, Hu TX, Xu YJ, Gu LM, Ni WC, Cheng S (1992) Studies on the hybrid of *Gossypium hirsutum* L. and *G. anomalum* and application in breeding. Sci Agricul Sin 25(6):44–51

Rhyne CL (1958) Linkage studies in *Gossypium*.I. Altered recombination in allotetraploid *G. hirsutum* L. following linkage group transference from related diploid species. Genetics 43:822–883

Rhyne CL (1960) Linkage studies in *Gossypium*. II. Altered recombination values in a linkage group of allotetraploid *G. hirsutum* L. as a result of transferred diploid species genes. Genetics 45:673–681

Sears ER, Okamoto M (1958) Intergenomic chromosome relationships in hexaploid wheat. Genetics 10:258–259

Silow RA (1941) The comparative genetics of *Gossypium anomalum* and the cultivated Asiatic cottons. J Genet 42:259–358

ur Rahman M, Zafar Y, Paterson AH (2009) *Gossypium* DNA markers: types, numbers, and uses. In: Paterson AH (ed) Genetics and genomics of cotton, plant genetics and genomics: crops and models 3. Springer science, LLC, New York

Wang ZN, Zhang D, Wang XY, Tan X, Guo H, Paterson AH (2013) A whole-genome DNA marker map for cotton based on the D-genome sequence of *Gossypium raimondii* L. G3 3:1759–1767

Xiao J, Wu K, Fang DD, Stelly DM, Yu J, Cantrell RG (2009) New SSR markers for use in cotton (*Gossypium spp*.) improvement. J Cotton Sci 13:75–157

Young ND, Tanksley SD (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor Appl Genet 77:95–101

Zhang J, Wu YT, GuoWZ ZhangTZ (2000) Fast screening of SSR markers in cotton with PAGE/silver staining. Cotton Sci Sinica 12:267–269

Zhang X, Zhai CJ, He LC, Guo Q, Zhang XG, Xu P, Su HM, Gong YY, Ni WC, Shen XL (2014) Morphological, cytological and molecular analyses of a synthesized hexaploid derived from an interspecific hybrid between *Gossypium hirsutum* and *G. anomalum*. Crop J 2:272–277