Theor Appl Genet (2006) 112: 430–439
DOI 10.1007/s00122-005-0142-9

ORIGINAL PAPER

Zhiguo Han · Changbiao Wang · Xianliang Song
Wangzhen Guo · Jinying Gou · Chunhong Li
Xiaoya Chen · Tianzhen Zhang

Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton

Received: 5 August 2005 / Accepted: 24 October 2005 / Published online: 9 December 2005

Abstract In order to construct a saturated genetic map and facilitate marker-assisted selection (MAS) breeding, it is necessary to enhance the current reservoir of known molecular markers in *Gossypium*. Microsatellites or simple sequence repeats (SSRs) occur in expressed sequence tags (EST) in plants (Kantety et al., Plant Mol Biol 48:501–510, 2002). Many ESTs are publicly available now and represent a good tool in developing EST-SSRs. From 13,505 ESTs developed from our two cotton fiber/ovule cDNA libraries constructed for Upland cotton, 966 (7.15%) contained one or more SSRs and from them, 489 EST-SSR primer pairs were developed. Among the EST-SSRs, 59.1% are trinucleotides, followed by dinucleotides (30%), tetranucleotides (6.4%), pentanucleotides (1.8%), and hexanucleotides (2.7%). AT/TA (18.4%) is the most frequent repeat, followed by CTT/GAA (5.3%), AG/TC (5.1%), AGA/TCT (4.9%), AGT/TCA (4.5%), and AAG/TTC (4.5%). One hundred and thirty EST-SSR loci were produced from 114 informative EST-SSR primer pairs, which generated polymorphism between our two mapping parents. Of these, 123 were integrated on our allotetraploid cotton genetic map, based on the cross [(TM-1×Hai7124)TM-1]. EST-SSR markers were distributed over 20 chromosomes and 6 linkage groups in the map. These EST-SSR markers can be used in genetic mapping, identification of quantitative trait loci (QTLs), and comparative genomics studies of cotton.

Abbreviations EST: Expression sequence tag · PCR: Polymerase chain reaction · dpa: Day post-anthesis bp: Base pair · SSR: Simple sequence repeats

Introduction

Cotton (*Gossypium* spp.) is an important cash crop and the second largest source of textile fiber and edible oil throughout the world. In recent years, improvement in the quality of cotton fiber has been extremely important because of changes in spinning technology (Shen et al. 2005). With advancements in molecular marker technology, marker-assisted selection (MAS) combined with conventional breeding has been one way in which fiber quality can be improved. To increase the reliability and usability of MAS breeding, there is a need to develop more polymorphic molecular markers. As the cotton genome is relatively large, with a 1C content of 2,230 Mbp (Arumuganathan and Earle 1991), approximately 5,000 markers are needed to effectively understand and quantitatively interpret the cotton genome (Lacape et al. 2003). To date, many types of DNA markers, including restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNAs (RAPD), amplified fragment length polymorphisms (AFLP), simple sequence repeats (SSR) and sequence-tagged sites (STS) have been developed for cotton research (Reinisch et al. 1994; Jiang et al. 1998, 2000; Shappley et al. 1998; Ulloa and Meredith 2000; Reddy et al. 2001; Ulloa et al. 2002; Zhang et al. 2002; Lacape et al. 2003; Mei et al. 2004; Nguyen et al. 2004; Rong et al. 2004). However, to construct a saturated

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-005-0142-9 and is accessible for authorized users.

Communicated by C. Möllers

Zhiguo Han and Changbiao Wang contributed equally to this work.

Z. Han · C. Wang · X. Song · W. Guo · T. Zhang (✉)
National Key Laboratory of Crop Genetics & Germplasm Enhancement, Cotton Research Institute, Nanjing Agricultural University, 210095 Nanjing, Jiangsu Province, P. R. China
E-mail: cotton@njau.edu.cn
Fax: 86-25-84395307

J. Gou · C. Li · X. Chen
National Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200032 Shanghai, P. R. China

genetic map that would expedite genetic improvement in cotton, new sources of molecular markers are needed.

SSRs or microsatellites are tandemly repeated DNA motifs (1–6 bp long) which may vary in the number of repeats at a given locus. SSRs are easy to use and analyze (Morgante and Olivieri 1993). Recent studies have revealed that gene transcripts can also contain repeat motifs, and the abundance of expressed sequence tags (ESTs) makes this an attractive potential source of microsatellite markers (Kantety et al. 2002). EST-SSRs have been identified for many crops, including *Triticum aestivum* L. (Gupta et al. 2003; Gao et al. 2004; Nicot et al. 2004), *Medicago truncatula* (Eujayl et al. 2004) and *Vitis vinifera* (Decroocq et al. 2003). EST-SSRs have also been developed for *Gossypium* (Saha et al. 2003; Han et al. 2004; Qureshi et al. 2004).

Qureshi et al. (2004) reported that out of 9,948 ESTs belonging to *Gossypium hirsutum*, 84 primer pairs were designed for amplification of EST-SSR markers, and these primer pairs were tested to detect polymorphism among three lines of *G. hirsutum* and one line of *G. barbadense*. Their study showed 26% intraspecies polymorphism among *G. hirsutum* cotton cultivars and 52% interspecies polymorphism between *G. hirsutum* and *G. barbadense*. We have developed 544 *G. arboreum* derived SSR primer pairs from publicly available EST sequences; 99 of them were mapped in our interspecific BC_1 mapping population from cross [(TM-1×Hai7124)×TM-1] (Han et al. 2004). These *G. arboreum* derived SSR data are publicly available in two websites (http://algodon.tamu.edu/~mapbase/SSR-frame-page.htm and http://www.mainlab.clemson.edu/cmd/projects/nau).

In the present study, we developed 489 *G. hirsutum* derived EST-SSRs. Their sequence characteristics and putative functions of their predicted products were reported. Based on the polymorphism surveyed in interspecific cotton mapping parents, *G. hirsutum* cv. TM-1 and *G. barbadense* cv. Hai7124, 123 EST-SSR markers were integrated into our backbone genetic map in tetraploid cotton.

Materials and methods

EST sequencing and EST-SSR identification

A cDNA library was constructed from 5 to 25 day post-anthesis (dpa), developing fiber cells of the Upland cotton germplasm "7235", at Nanjing Agricultural University (NAU) in China. This introgression germplasm line was developed by crossing *G. anomalum* and *G. hirsutum*, then backcrossing to cultivars and strains with high fiber strength such as Acala 3080 and PD4381 (Qian et al. 1992). It is characterized by the properties of its very high quality fiber (Zhang et al. 2003). Another cDNA library was constructed with the 0–5 dpa ovules and 3–22 dpa fibers from *G. hirsutum* cv. Xuzhou 142, at the Institute of Plant Physiology and Ecology, Shanghai, China. Random sequencing of 5,767 ESTs at the 5′-end from the "7235" library and 7,738 ESTs at the 3′-end from Xuzhou 142 was conducted in Bioasia Biotech, Shanghai, China. The simple sequence repeat identification tool (SSRIT) (http://www.gramene.org/db/searches/ssrtool) was used to identify SSRs from these cotton ESTs.

The EST-SSRs were found to contain motifs of 2–6 nucleotides in size. The minimum repeat unit was defined as five for dinucleotides and four for the higher order motifs including tri-, tetra-, penta-, and hexanucleotides. The EST-derived SSR markers in this work were designated using "NAU" (Nanjing Agricultural University) as a short prefix (Han et al. 2004).

Development of EST-SSR primer pairs

The software program, Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), was utilized to design primer pairs flanking SSRs. The major parameters for primer design were set as follows: primer length 18–24 bp with 20 bp as the optimum; PCR product size 100–300 bp; optimum annealing temperature 57°C; GC content 35–60% with 50% as the optimum. The primers were synthesized by Bioasia Biotech, Shanghai, China.

DNA extraction, PCR amplification, and electrophoresis

Cotton genomic DNA was isolated from young leaves as described by Paterson et al. (1993). PCR amplifications on a Peltier Thermal Cycler (MJ Research) and product electrophoresis were performed as previously described (Zhang et al. 2000, 2002).

Linkage mapping

The mapping population was comprised of 140 BC_1 individuals which were generated from the cross [(TM-1×Hai7124)×TM-1] (Song et al. 2005). TM-1 was developed as a genetic standard accession of *G. hirsutum* in the United States (Kohel et al. 2001), whereas Hai7124, developed in China, is a commercial *Verticillium*-resistant cultivar of *G. barbadense*. Based on this backcross population, a backbone genetic map including 482 loci was constructed in allotetraploid cotton using SSR markers and two morphology markers (Song et al. 2005). Additional SSR and sequence-related amplified polymorphism (SRAP) markers have subsequently been integrated into this map (Han et al. 2004; Song et al., unpublished data). SSR primers were obtained from DNA sequences from the following sources: BNL primers from Research Genetics Co. (Huntsville, AL, USA, http://www.resgen.com); JESPR from Reddy et al. (2001); TM from Dr John Yu, USDA-ARS, Crops

Germplasm Research Unit, Texas, USA; EST from Dr S. Saha, USDA-ARS, Crop Science Research Laboratory, Mississippi, USA; and CIR from Nguyen et al. (2004). These primers can be downloaded at http://www.mainlab.clemson.edu/cmd/projects.

Linkage groups were assigned to the subgenomes and chromosomes based on our backbone linkage maps (Zhang et al. 2002; Song et al. 2005) and published maps (Lacape et al. 2003; Rong et al. 2004) after aligning groups with common SSR loci.

The EST-SSR primer combinations (489 in total) were employed to screen interspecific polymorphisms between TM-1 and Hai7124. When an SSR was found to be polymorphic between TM-1 and Hai7124, it was used to survey 140 individuals of the BC_1 mapping population. The maternal (TM-1) genotype and the heterozygous F_1 genotype of the BC_1 population were scored as 1 and 3, respectively. Missing data were designated zero (0).

MapMaker version 3.0b (Lander et al. 1987) was employed to construct the linkage map. First, for all with "error detection on", the command "Group" (LOD 8, recombination fraction 30 cM) was used to identify all the markers. Then the commands "Try" and "Map" were used. Marker order was confirmed with the "Ripple" command. Recombination frequencies were converted into map distances (centi-Morgans) using the Kosambi mapping function (Kosambi 1944).

Putative function analysis

Using BLASTX algorithms, 489 ESTs containing microsatellites were used to search the GenBank nonredundant database; the E-value threshold was set to $\leq 10^{-7}$ (http://www.ncbi.nlm.nih.gov/BLAST). They were then allocated to the corresponding functional categories by referring to the SRB embryonic EST project (http://www.mcdb.ucla.edu/Research/Goldberg).

Results

Characteristics of EST-SSR derived from the tetraploid cotton

A total of 13,505 ESTs of *Gossypium* fiber cells and ovules were employed in this investigation to develop SSR markers. These ESTs were generated from two cultivars and lines: germplasm 7235 (5,767 ESTs) and Xuzhou 142 (7,738 ESTs) in *G. hirsutum*. We detected 966 ESTs, 223 from the "7235" and 743 from Xuzhou 142, which contain one to several SSRs. Based on DNA sequences, 489 EST-SSR primer pairs were developed. Among them, 61 ESTs (12.5%) contain two or three adjacent repeats. The primer sequence, GenBank accession number, repeat motif and number, polymorphism detected between TM-1 and Hai7124 and putative function (BlastX) for these 489 EST-SSR primer pairs were presented in our Electronic Supplementary Material. Other EST sequences were unsuitable for designing PCR primers. Among these 489 EST-SSR primer pairs, the most common repeat types were trinucleotides (59.1%), followed by dinucleotides (30%), tetranucleotides (6.4%), hexanucleotides (2.7%), and pentanucleotides (1.8%) (Fig. 1). The motif AT/TA had the highest frequency of 18.4% followed by the motifs CTT/GAA (5.3%), AG/TC (5.1%), AGA/TCT (4.9%), AGT/TCA (4.5%), and AAG/TTC (4.5%) (Fig. 2).

Genetic mapping of EST-SSRs

Among the 489 EST-SSR primer pairs developed in this study, 114 (23.3%) were informative when used to screen TM-1 and Hai7124, the parents of our interspecific BC_1 mapping population (Song et al. 2005). Only two primer pairs (NAU2000 and NAU2040) from 489 markers amplified unexpected fragment sizes. Altogether, 114 primer pairs amplified 130 loci in allotetraploid cotton. Of these, a total of 123 microsatellite loci, including 9 deviated loci (7%), were integrated into our backbone genetic map. They were anchored on all 20 chromosomes and 6 linkage groups (Figs. 3, 4). Out of these 123 EST-SSR markers, 62 were assigned to the At subgenome and 61 to the Dt subgenome in the allotetraploid cotton. But 72 loci were anchored to the At and 37 to the Dt subgenome of allotetraploid cotton based on linkage tests in our mapping EST-SSR markers derived from the diploid *G. arboreum* (Han et al. 2004). If we include all new integrated loci, our genetic map now consists of 907 loci and 5,060 cM, with an average between-loci distance of 5.6 cM.

To date, 13 homeologous pairs have been identified (Endrizzi et al. 1985; Crane et al. 1994; Reinisch et al. 1994; Lacape et al. 2003; Rong et al. 2004). Seven of 10 duplicated loci surveyed by *G. arboreum* derived EST-SSRs were mapped on their corresponding homeologous chromosomes or linkage groups (Han et al. 2004). In this study, 18 EST-SSR markers generated by 8 primer pairs were also found to be distributed on their corresponding homeologous chromosomes or linkage groups: NAU2016-250 in LGA03 and NAU2016-200 in LGD02; NAU2170-250 in Chr. 12 and NAU2170-230 in

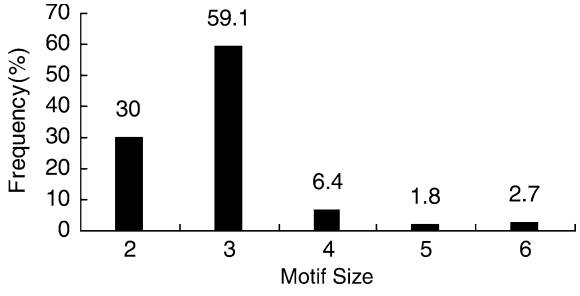
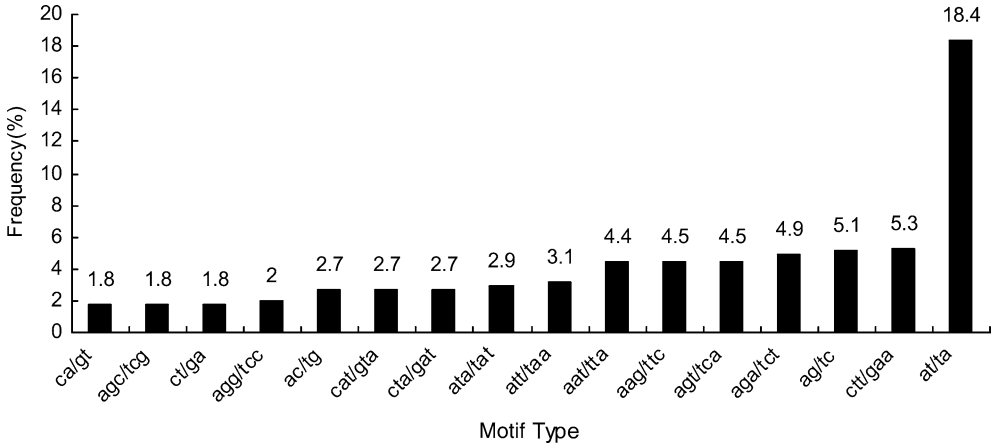


Fig. 1 Distribution of EST-derived simple sequence repeats (SSRs) based on motif size



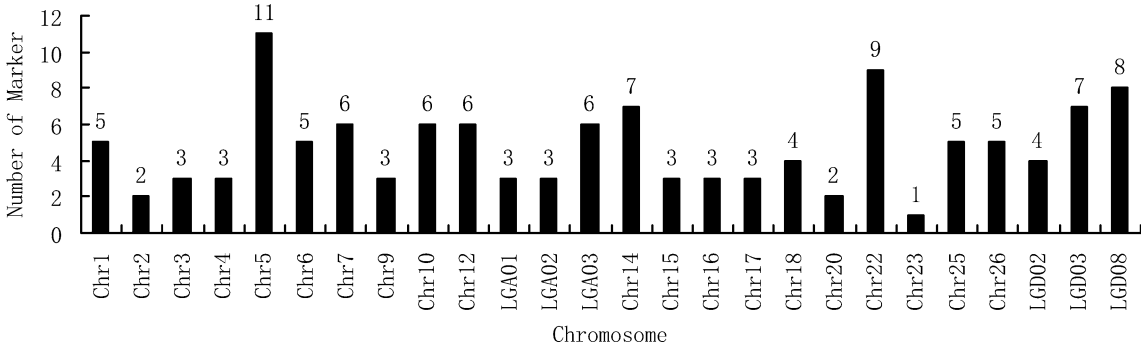


Fig. 2 Distribution of EST-derived SSRs according to motif sequence type

Fig. 3 Distribution of EST-derived SSRs based on motif sequence type on the chromosomes and/or linkage groups

Chr. 26; NAU2186-300 and NAU2186-360 in Chr. 7 and NAU2186-150 in Chr. 16; NAU2238-120 in Chr. 6 and NAU2238-150 in Chr. 25; NAU2274-110 in Chr. 5 and NAU2274-105 in LGD08; NAU2363-170 in Chr. 4 and NAU2363-175 in Chr. 22; NAU2432-320 and NAU2432-460 in Chr. 7 and NAU2432-170 in Chr. 16; and NAU2477-200 in Chr. 4 and NAU2477-205 in Chr. 22 (Fig. 4).

Putative functions of the products of ESTs containing SSR

To explore the potential utility of the EST-SSR markers for use in research of the cotton structural genome, the 489 EST-SSRs were compared to those in the GenBank database using BLASTX with an E-value $\leq 10^{-7}$. In reference to the SRB embryonic EST project (http://www.mcdb.ucla.edu/Research/Goldberg), putative functions of these ESTs were classified into 14 main categories: transcription and post-transcription (4.1%), protein destination and storage (3.1%), metabolism (5.3%), protein synthesis (3.7%), signal transduction (1.6%), energy (1.6%), intracellular trafficking (1.2%), transport (0.8%), cell structure (5.1%), disease and defense (1.8%), cell growth and division (0.6%), secondary metabolism (0.2%), uncategorized (29.4%), and unknown function (42.3%). Some important functional genes were identified, including acyl-coenzyme A-binding protein, adenylate kinase, and anthocyanidin reductase (Ji et al. 2003; Arpat et al. 2004). Information of their GenBank accession numbers, primer number, sequence, motif and the repeats, and putative functions are available in the Electronic Supplementary Material.

Discussion

Many microsatellite markers (SSR) have been developed in *Gossypium* (Connell et al. 1998; Reddy et al. 2001; Kumpatla et al. 2002; Saha et al. 2003; Nguyen et al. 2004; Han et al. 2004). These markers were generated from cotton genomic or EST sequences and most have been used successfully in the construction of genetic map and molecular tagging.

Saha et al. (2003) identified 34% EST-SSR sequences in *G. hirsutum* containing trinucleotide repeat motifs. Our previous investigation indicated that hexanucleotide motifs were the most frequent (40.1%) in the 544 EST-SSRs from 931 ESTs of *G. arboreum* (Han et al. 2004). In the present study, 489 EST-SSR primer pairs were developed from two cotton cDNA libraries in *G. hirsutum*. Among 550 SSRs in these ESTs, 59.1% also contain trinucleotides. Using the end-sequencing data of BAC/BIBAC clones (http://algodon.tamu.edu/~mapbase/SSR-frame-page.htm), 265 repeat motifs were detected,

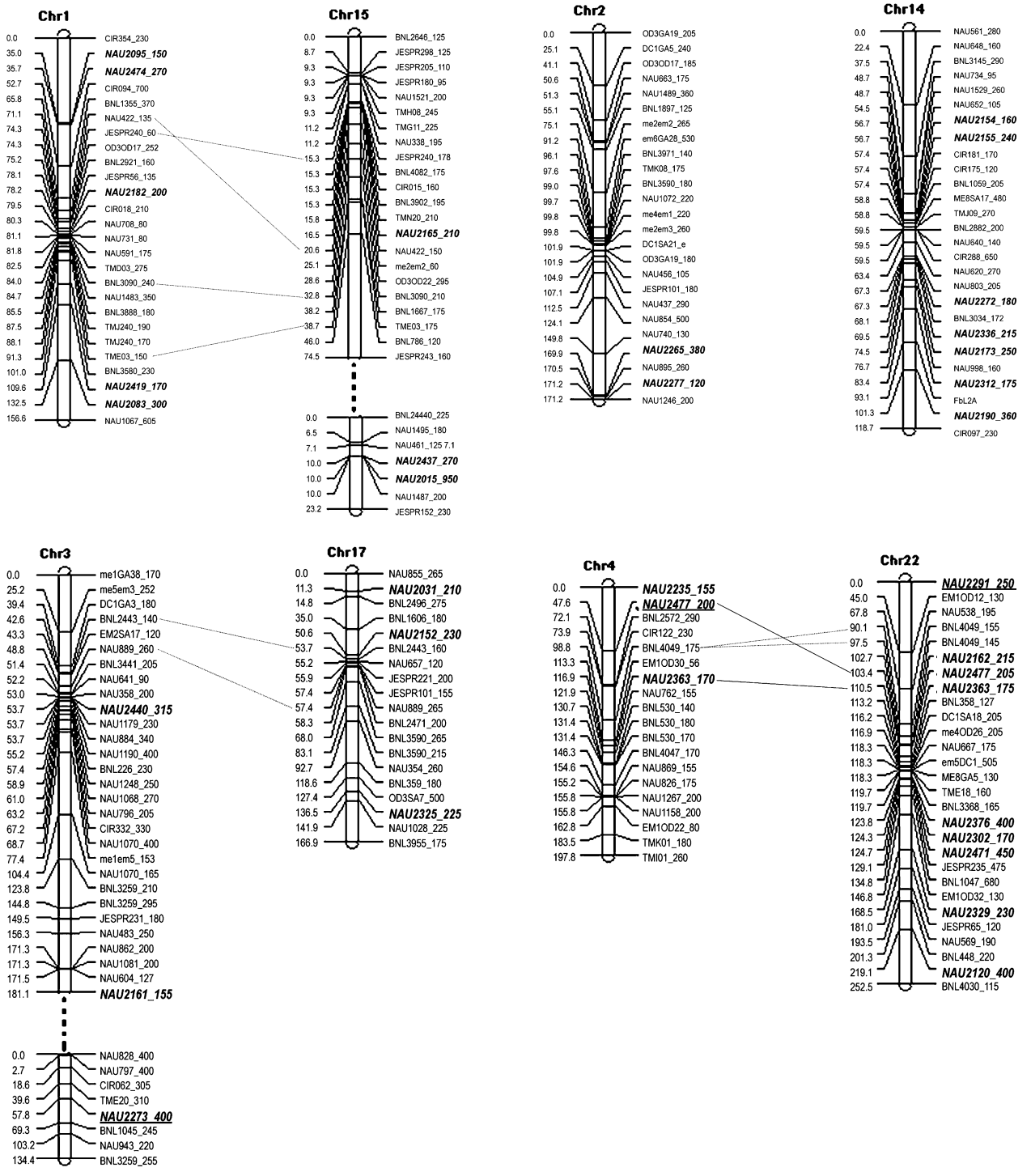


Fig. 4 New genetic map constructed using a BC_1 population obtained from the interspecific cross: *G. hirsutum* L. cv. TM-1 × *G. barbadense* L. cv. Hai7124. Chromosomes and linkage groups are arranged by 13 homeologous pairs. Positions of loci are given in centiMorgans (Kosambi 1944). Fragment sizes (in base pair) are given next to the maker name. EST-SSR markers developed in this study are indicated in *bold italics*. Deviated loci are *underlined*. Homeologous loci identified in this study were connected by a *solid bar*. Homeologous loci identified before were connected with a *broken line*

including 213 (80%) dinucleotides, 49 (18.5%) trinucleotides, 1 tetranucleotide, 1 pentanucleotide, and 1 hexanucleotide. Thereafter, they developed 192 genome SSR primers in Upland cotton. The differences in Upland cotton repeat motif type detected in various studies may be partially due to different materials or analyzing methods employed. Additionally, Reddy et al. (2001) found the most frequently occurring microsatellite motif in *G. hirsutum* genome was AAG, which constituted 32% percent of all 10 possible types of di- to hexanucleotide

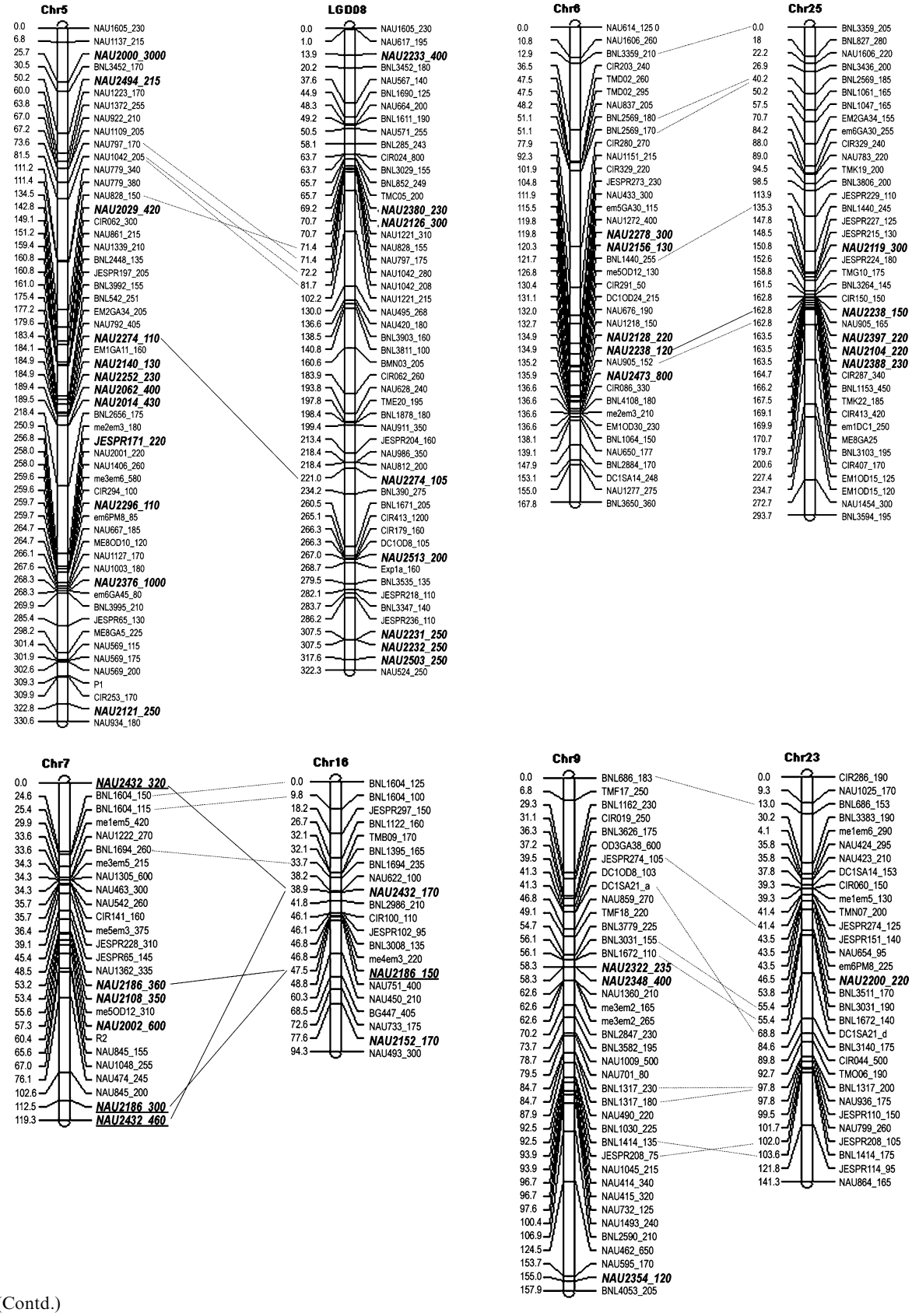

Chr5
0.0 NAU1605_230
6.8 NAU1137_215
25.7 NAU2000_3000
30.5 BNL3452_170
50.2 NAU2494_215
60.0 NAU1223_170
63.8 NAU1372_255
67.0 NAU922_210
67.2 NAU1109_205
73.6 NAU797_170
81.5 NAU1042_205
111.2 NAU779_340
111.4 NAU779_380
134.5 NAU828_150
142.8 NAU2029_420
149.1 CIR062_300
151.2 NAU861_215
159.4 NAU1339_210
160.8 BNL2448_135
160.8 JESPR197_205
161.0 BNL3992_155
175.4 BNL542_251
177.2 EM2GA34_205
179.6 NAU792_405
183.4 NAU2274_110
184.1 EM1GA11_160
184.9 NAU2140_130
184.9 NAU2252_230
189.4 NAU2062_400
189.5 NAU2014_430
218.4 BNL2656_175
250.9 me2em3_180
256.8 JESPR171_220
258.0 NAU2001_220
258.0 NAU1406_260
259.6 me3em6_580
259.6 CIR294_100
259.7 NAU2296_110
259.7 em6PM8_85
264.7 NAU667_185
264.7 ME8OD10_120
266.1 NAU1127_170
267.6 NAU1003_180
268.3 NAU2376_1000
268.3 em6GA45_80
269.9 BNL3995_210
285.4 JESPR65_130
298.2 ME8GA5_225
301.4 NAU569_115
301.9 NAU569_175
302.6 NAU569_200
309.3 P1
309.9 CIR253_170
322.8 NAU2121_250
330.6 NAU934_180
LGD08
0.0 NAU1605_230
1.0 NAU617_195
13.9 NAU2233_400
20.2 BNL3452_180
37.6 NAU567_140
44.9 BNL1690_125
48.3 NAU664_200
49.2 BNL1611_190
50.5 NAU571_255
58.1 BNL285_243
63.7 CIR024_800
63.7 BNL3029_155
65.7 BNL852_249
65.7 TMC05_200
69.2 NAU2380_230
70.7 NAU2126_300
70.7 NAU1221_310
71.4 NAU828_155
71.4 NAU797_175
72.2 NAU1042_280
81.7 NAU1042_208
102.2 NAU1221_215
130.0 NAU495_268
136.6 NAU420_180
138.5 BNL3903_160
140.8 BNL3811_100
160.6 BMN03_205
183.9 CIR062_260
193.8 NAU628_240
197.8 TME20_195
198.4 BNL1878_180
199.4 NAU911_350
213.4 JESPR204_160
218.4 NAU986_350
218.4 NAU812_200
221.0 NAU2274_105
234.2 BNL390_275
260.5 BNL1671_205
265.1 CIR413_1200
266.3 CIR179_160
266.3 DC1OD8_105
267.0 NAU2513_200
268.7 Exp1a_160
279.5 BNL3535_135
282.1 JESPR218_110
283.7 BNL3347_140
286.2 JESPR236_110
307.5 NAU2231_250
307.5 NAU2232_250
317.6 NAU2503_250
322.3 NAU524_250
Chr6
0.0 NAU614_125 0
10.8 NAU1606_260
12.9 BNL3359_210
36.5 CIR203_240
47.5 TMD02_260
47.5 TMD02_295
48.2 NAU837_205
51.1 BNL2569_180
51.1 BNL2569_170
77.9 CIR280_270
92.3 NAU1151_215
101.9 CIR329_220
104.8 JESPR273_230
111.9 NAU433_300
115.5 em5GA30_115
119.8 NAU1272_400
119.8 NAU2278_300
120.3 NAU2156_130
121.7 BNL1440_255
126.8 me5OD12_130
130.4 CIR291_50
131.1 DC1OD24_215
132.0 NAU676_190
132.7 NAU1218_150
134.9 NAU2128_220
134.9 NAU2238_120
135.2 NAU905_152
135.9 NAU2473_800
136.6 CIR086_330
136.6 BNL4108_180
136.6 me2em3_210
136.6 EM1OD30_230
138.1 BNL1064_150
139.1 NAU650_177
147.9 BNL2884_170
153.1 DC1SA14_248
155.0 NAU1277_275
167.8 BNL3650_360
Chr25
0.0 BNL3359_205
18 BNL827_280
22.2 NAU1606_220
26.9 BNL3436_200
40.2 BNL2569_185
50.2 BNL1061_165
57.5 BNL1047_165
70.7 EM2GA34_155
84.2 em6GA30_255
88.0 CIR329_240
89.0 NAU783_220
94.5 TMK19_200
98.5 BNL3806_200
113.9 JESPR229_110
135.3 BNL1440_245
147.8 JESPR227_125
148.5 JESPR215_130
150.8 NAU2119_300
152.6 JESPR224_180
158.8 TMG10_175
161.5 BNL3264_145
162.8 CIR150_150
162.8 NAU2238_150
162.8 NAU905_165
163.5 NAU2397_220
163.5 NAU2104_220
163.5 NAU2388_230
164.7 CIR287_340
166.2 BNL1153_450
167.5 TMK22_185
169.1 CIR413_420
169.9 em1DC1_250
170.7 ME8GA25
179.7 BNL3103_195
200.6 CIR407_170
227.4 EM1OD15_125
234.7 EM1OD15_120
272.7 NAU1454_300
293.7 BNL3594_195
Chr7
0.0 NAU2432_320
24.6 BNL1604_150
25.4 BNL1604_115
29.9 me1em5_420
33.6 NAU1222_270
33.6 BNL1694_260
34.3 me3em5_215
34.3 NAU1305_600
34.3 NAU463_300
35.7 NAU542_260
35.7 CIR141_160
36.4 me5em3_375
39.1 JESPR228_310
45.4 JESPR65_145
48.5 NAU1362_335
53.2 NAU2186_360
53.4 NAU2108_350
55.6 me5OD12_310
57.3 NAU2002_600
60.4 R2
65.6 NAU845_155
67.0 NAU1048_255
76.1 NAU474_245
102.6 NAU845_200
112.5 NAU2186_300
119.3 NAU2432_460
Chr16
0.0 BNL1604_125
9.8 BNL1604_100
18.2 JESPR297_150
26.7 BNL1122_160
32.1 TMB09_170
32.1 BNL1395_165
33.7 BNL1694_235
38.2 NAU622_100
38.9 NAU2432_170
41.8 BNL2986_210
46.1 CIR100_110
46.1 JESPR102_95
46.8 BNL3008_135
46.8 me4em3_220
47.5 NAU2186_150
48.8 NAU751_400
60.3 NAU450_210
68.5 BG447_405
72.6 NAU733_175
77.6 NAU2152_170
94.3 NAU493_300
Chr9
0.0 BNL686_183
6.8 TMF17_250
29.3 BNL1162_230
31.1 CIR019_250
36.3 BNL3626_175
37.2 OD3GA38_600
39.5 JESPR274_105
41.3 DC1OD8_103
41.3 DC1SA21_a
46.8 NAU859_270
49.1 TMF18_220
54.7 BNL3779_225
56.1 BNL3031_155
56.1 BNL1672_110
58.3 NAU2322_235
58.3 NAU2348_400
62.6 NAU1360_210
62.6 me3em2_165
62.6 me3em2_265
70.2 BNL2847_230
73.7 BNL3582_195
78.7 NAU1009_500
79.5 NAU701_80
84.7 BNL1317_230
84.7 BNL1317_180
87.9 NAU490_220
92.5 BNL1030_225
92.5 BNL1414_135
93.9 JESPR208_75
93.9 NAU1045_215
96.7 NAU414_340
96.7 NAU415_320
97.6 NAU732_125
100.4 NAU1493_240
106.9 BNL2590_210
124.5 NAU462_650
153.7 NAU595_170
155.0 NAU2354_120
157.9 BNL4053_205
Chr23
0.0 CIR286_190
9.3 NAU1025_170
13.0 BNL686_153
30.2 BNL3383_190
4.1 me1em6_290
35.8 NAU424_295
35.8 NAU423_210
37.8 DC1SA14_153
39.3 CIR060_150
39.3 me1em5_130
41.4 TMN07_200
41.4 JESPR274_125
43.5 JESPR151_140
43.5 NAU654_95
43.5 em6PM8_225
46.5 NAU2200_220
53.8 BNL3511_170
55.4 BNL3031_190
55.4 BNL1672_140
68.8 DC1SA21_d
84.6 BNL3140_175
89.8 CIR044_500
92.7 TMO06_190
97.8 BNL1317_200
97.8 NAU936_175
99.5 JESPR110_150
101.7 NAU799_260
102.0 JESPR208_105
103.6 BNL1414_175
121.8 JESPR114_95
141.3 NAU864_165

Fig. 4 (Contd.)

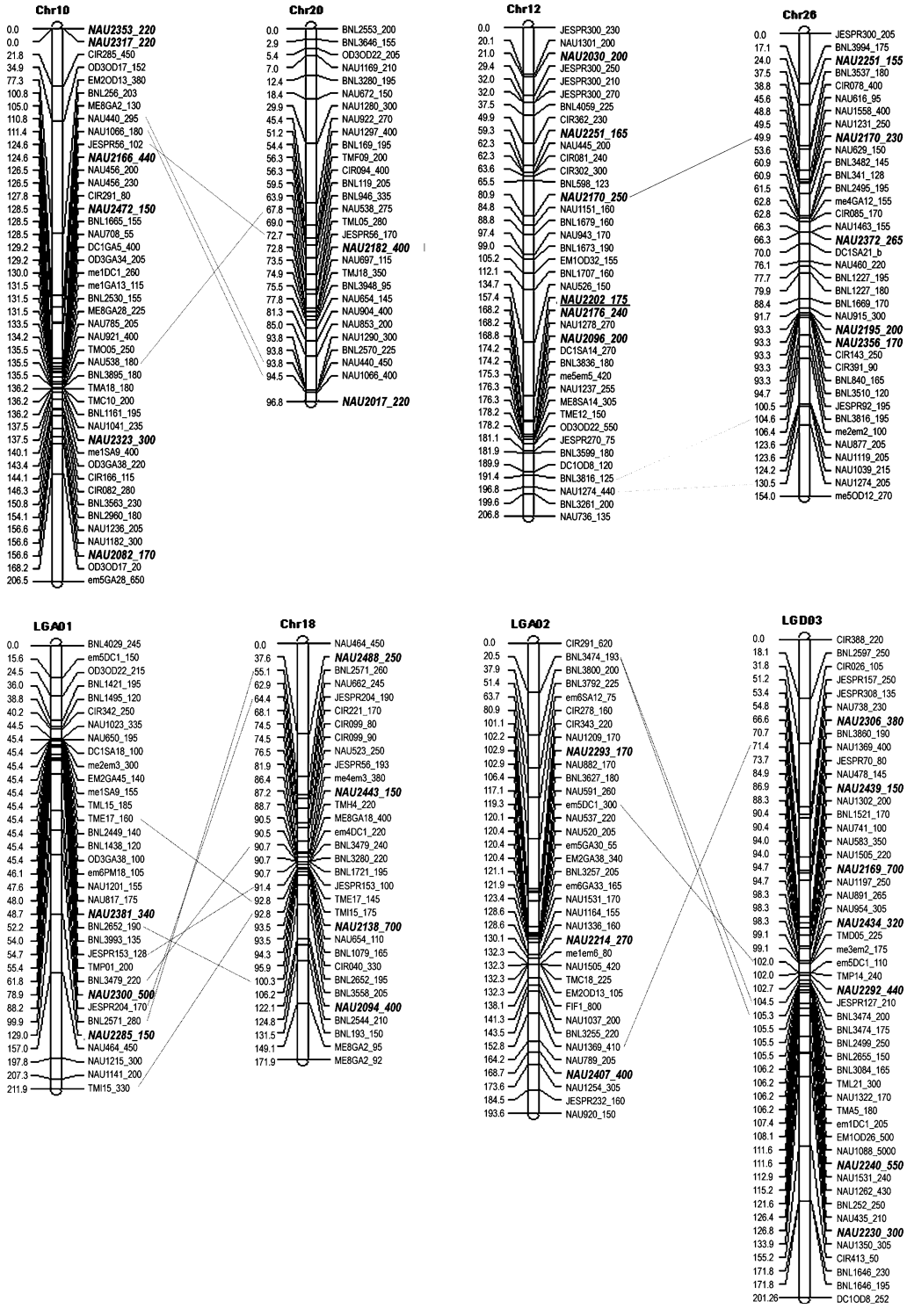
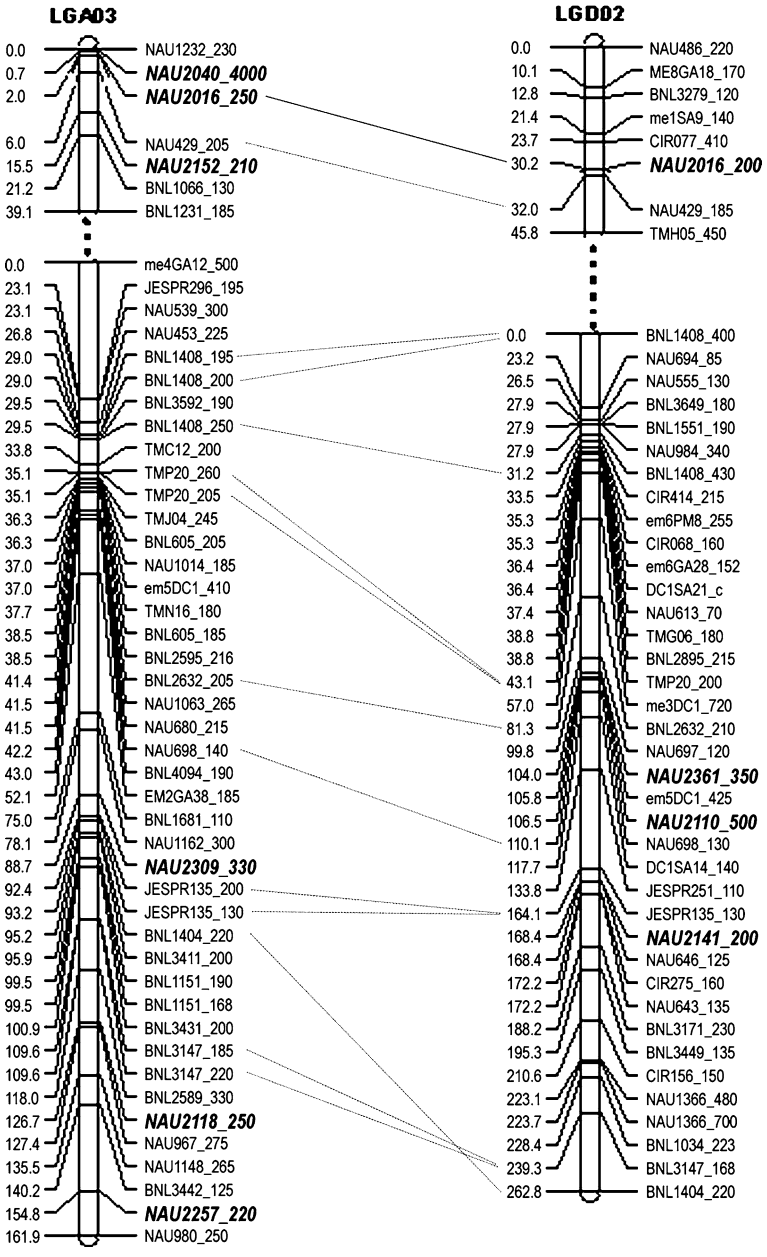


Fig. 4 (Contd.)

repeats. Saha et al. (2003) reported the GA/CT motif was the most abundant (24%) from 133 EST-SSRs of 9,447 ESTs in *G. hirsutum*, while the AAG/TTC repeat was the most common (18%) in 348 EST-SSRs from 26,630 ESTs of *G. arboreum*. Our study showed that AT/TA was the most frequent repeat followed by CTT/GAA, AG/TC, and AGA/TCT. All of these molecular markers are based on a small fraction of cotton genome information, which may account for SSR markers developed for amplified short products (see Electronic Supplementary Material). Further investigation of the interspecies variability in repeat motif frequency between *G. hirsutum* and *G. arboreum* might shed light onto the evolutionary events of the At and Dt subgenomes in tetraploid cotton.

Fig. 4 (Contd.)



Several investigations have suggested that there is a higher frequency of SSR polymorphisms within *G. hirsutum* and/or between *Gossypium* species (Reddy et al. 2001; Qureshi et al. 2004; Nguyen et al. 2004). Qureshi et al. (2004) detected 26% polymorphism for EST-SSRs within *G. hirsutum* and 52% between *G. hirsutum* and *G. barbadense* among 84 EST-SSR primer pairs. Reddy et al. (2001) observed 21% polymorphism within *G. hirsutum* and 49% polymorphism between *G. hirsutum* and *G. barbadense* for SSR markers derived from genomic DNA. Similarly, Nguyen et al. (2004) showed 56% polymorphism between *G. hirsutum* and *G. barbadense*. In this study, we identified a 23% polymorphic rate between TM-1 and Hai7124, and TM-1 and 3–79. This is roughly the same as found in our previous study in which 99 (18.2%) of 544 *G. arboreum* derived EST-SSRs were polymorphic and segregated in the interspecific BC_1 mapping population [(TM-1×Hai7124)×TM-1] (Han et al. 2004). The discrepancy in polymorphic rates may be due to different plant materials, different number of ESTs from the different tissues or number of EST-SSR primers used.

It is interesting to note that of 111 *G. arboreum* derived EST-SSR markers, 72 were anchored to the At and 37 to the Dt subgenome of allotetraploid cotton, a nearly 2:1 ratio (Han et al. 2004). In this study, however, 62 and 61 EST-SSR markers developed from *G. hirsutum* were mapped on the At and Dt subgenomes, respectively, essentially a 1:1 ratio. EST-SSR markers, developed from cotton-developing fiber and ovule cells which are mapped on the D-subgenome, revealed that there are some important genes for fiber

development in the D-genome chromosome. This may partially explain why quantitative trait loci (QTLs) for fiber-related traits were mapped in Dt genome chromosomes in tetraploid cotton, even though D-genome species do not produce spinnable fibers (Jiang et al. 1998; Kohel et al. 2001; Nguyen et al. 2004; Mei et al. 2004). These results suggest there are complex evolutionary relationships between A, At, D, and Dt genomes.

To date, very few cotton genes have been identified experimentally, thus making bioinformatics analysis of ESTs, generated from various cDNA libraries, a valuable method of functional classification. Using the subtractive PCR of cDNA prepared from 10 dpa wild-type cotton fiber as tester and cDNA from a fuzzless-lintless mutant as driver, 280 independent cDNA fragments and 172 genes were significantly upregulated in elongating cotton fibers as confirmed by in situ hybridization in representative cases by cDNA macroarrays (Ji et al. 2003). Arpat et al. (2004) identified more than 2,500 stage-specific ''expansion-associated'' genes that are downregulated coincident with the termination of fiber elongation and 81 novel genes newly identified that are preferentially expressed during secondary cell wall synthesis from comparison of 10 versus 24 dpa fiber transcripts. In this study, BLASTX was employed to survey the 489 SSR-containing ESTs, and some of these, with apparently important functions, were mapped. This, together with other ESTs or genes mapped, makes a substantial contribution to our understanding of the structure and function of the cotton genome, which can lead to improvements in cotton production and quality. EST-SSR mapping is one way to achieve a saturated genetic map, which is invaluable not only to map genes or QTLs for cotton yield, fiber quality, and disease resistance, but also for integrating physical and genetic maps.

Acknowledgments This research was supported by grants from the National High-tech Program (2004AA211172, 2003AA222050), the State Key Basic Research and Development Plan of China (2002CB111301), Program for Changjiang Scholars and Innovative Research Team in University, the Key Project of Chinese Ministry of Education (10418) and Jiangsu High-tech Project (BG2002306).

References

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Rep 9:208–218

Connell JP, Pammi S, Iqbal MJ, Huizinga T, Reddy AS (1998) A high throughput procedure for capturing microsatellites from complex plant genomes. Plant Mol Biol Rep 16:341–349

Crane CF, Price HJ, Stelly DM, Czeschin DJ Jr (1994) Identification of a homeologous chromosome pair by in situ DNA hybridization to ribosomal RNA loci in meiotic chromosomes of cotton (*Gossypium hirsutum*). Genome 36:1015–1022

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 Genet 106:912–922

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

Eujayl I, Sledge MK, Wang L, May GD, Chekovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Gao LF, Jing RL, Huo NX, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270:315–323

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Gen Genomics 272:308–327

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC, Zhu YX (2003) Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31:2534–2543

Jiang C, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci USA 95:4419–4424

Jiang C, Wright RJ, Woo SS, Delmonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409–418

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barely, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kohel RJ, Yu J, Park YH (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Kumpatla SP, Horne EC, Shah MR, Gupta M, Thompson SA (2002) Development of SSR markers: towards genetic mapping in cotton (*Gossypium hirsutum* L) Abstract communication at the 3rd international cotton genome initiative workshop, Nanjing, China. Cotton Sci 14(Supp l):28

Lacape JM, Nguyen TB, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG, Burr B, Hau B (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population. Genome 46:612–626

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

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

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 Genet 109:167–175

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P. (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Paterson AH, Brubaker C, 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

Qian SY, Huang JQ, Peng YT, Zhou BL, Ying MC, Shen DZ, Liu GL, Hu TX, Xu YJ, Gu LM, Ni WC, Chen S (1992) Studies on the hybrid of *G. hirsutum* L. and *G. anomalum* Wawr. & Peyr. and application in breeding (in Chinese). Sci Agric Sin 25:44–51

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

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

Reinisch A, Dong JM, Brubaker CL, Stelly DM, Wendel JE, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*; chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rong JK et al (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy UK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Shappley ZW, Jenkins JN, Meredith WR, McCarty JC (1998) An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. Theor Appl Genet 97:756–761

Shen XL, Guo WZ, Zhu XF, Yuan YL, Yu J, Kohel RJ, Zhang TZ (2005) Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. Mol Breed 15:169–181

Song XL, Wang K, Guo WZ, Zhang J, Zhang TZ. (2005) A comparison of genetic maps constructed from haploid and BC_1 mapping populations from the same crossing between *Gossypium hirsutum L.* × *G. barbadense L.* Genome 48:378–390

Ulloa M, Meredith WR Jr (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. J Cotton Sci 4:161–170

Ulloa M, Meredith WR Jr, Shappley ZW, Kahler AL (2002) RFLP genetic linkage maps from four $F_{2:3}$ populations and a joinmap of *Gossypium hirsutum* L. Theor Appl Genet 104:2000–2008

Zhang J, Wu YT, Guo WZ, Zhang TZ (2000) Fast screening of microsatellite markers in cotton with PAGE/silver staining. Cotton Sci Sin 12:267–269

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

Zhang TZ, Yuan YL, Yu J, Guo WZ, Kohel RJ (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theor Appl Genet 106:262–268