Theor Appl Genet (2004) 109:167–175
DOI 10.1007/s00122-004-1612-1

ORIGINAL PAPER

T.-B. Nguyen · M. Giband · P. Brottier · A.-M. Risterucci · J.-M. Lacape

Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers

Received: 18 August 2003 / Accepted: 19 January 2004 / Published online: 2 March 2004

Abstract Microsatellite [simple-sequence repeat (SSR)] markers were developed and positioned on the genetic map of tetraploid cotton. Three hundred and ninety-two unique microsatellite sequences, all but two containing a (CA/GT) repeat, were isolated, and the deduced primers were used to screen for polymorphism between the *Gossypium hirsutum* and *G. barbadense* parents of the mapping population analyzed in our laboratory. The observed rate of polymorphism was 56%. The 204 polymorphic SSRs revealed 261 segregating bands, which ultimately gave rise to 233 mapped loci. The updated status of our genetic map is now of 1,160 loci and 5,519 cM, with an average distance between two loci of 4.8 cM. The presence of a total of 466 microsatellite loci, with an average distance of 12 cM between two SSR loci, now provides wide coverage of the genome of tetraploid cotton and thus represents a powerful means for the production of a consensus map and for the effective tracking of QTLs.

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-004-1612-1

Communicated by C. Möllers

T.-B. Nguyen · M. Giband · A.-M. Risterucci · J.-M. Lacape (✉)
UMR1096, TA 70/03,
Centre International en Recherche Agronomique
pour le Développement (CIRAD),
Avenue Agropolis, 34398 Montpellier, Cedex 5, France
e-mail: marc.lacape@cirad.fr
Tel.: +33-4-67615968
Fax: +33-4-67615605

P. Brottier
Genoscope,
Centre National de Séquençage,
2 rue Gaston Crémieux CP 5706, 91057 Evry, Cedex, France

Introduction

Cotton (*Gossypium hirsutum* L.) is the most important source of natural fibers and, despite its economical importance for many countries, the application of molecular tools for the analysis of the genome of this crop has lagged behind efforts devoted to other important species. However, recent years have seen the development of efforts aimed at a better understanding of the cotton genome and its organization. The first detailed RFLP map of tetraploid ($2n=4x=52$) cotton was derived from the analysis of an interspecific *G. hirsutum* × *G. barbadense* F_2 population (Reinish et al. 1994). This map has now been further completed and is available from the University of Georgia Web site (http://www.plantgenome.uga.edu/cotton/GeneticMap.htm). Though less comprehensive in genome coverage, other maps had been reported both from intraspecific (Shappley et al. 1998; Ulloa et al. 2002) and interspecific populations (Yu and Kohel 2001; Zhang et al. 2002). Our team recently published another map stemming from an interspecific cross (Lacape et al. 2003). This map, which combines RFLP, AFLP, and SSR markers, comprises 888 loci mapped on 26 long and 11 short linkage groups and spans a genetic length of 4,400 cM.

Due to the facts that the cotton genome is relatively large, with a 1C content of 2,250 Mb (Arumuganathan and Earle 1991), and that intraspecific molecular polymorphism in this species is low (Tatineni et al. 1996; Brubaker and Wendel 2001), there is an unquestionable need for more highly polymorphic genetic markers if developments in molecular genetics are to be of practical use for marker-assisted breeding programs (Brubaker et al. 2000).

Microsatellite [or simple-sequence repeat, (SSR)] markers detect variation in short tandem-repeat motifs. The high level of variability and the relatively low cost for the user have made SSRs the molecular markers of choice for genetic mapping (Tautz 1989). In cotton, the first public microsatellite markers were developed at Brookhaven National Laboratory (usual prefix "BNL").

The set of 216 BNL microsatellites was derived from *G. hirsutum* genomic libraries enriched in $(GA)_n$ and $(CA)_n$ inserts. Other SSRs isolated from cotton genomic DNA libraries include the 500 CM SSRs reported by Connell et al. (1998), the set of 307 JESPR microsatellites of Reddy et al. (2001), as well as the more than 1,200 SSRs recently reported by the Dow Agroscience Seed Company (Kumpatla et al. 2002). To date, only the BNL SSRs have been used for mapping purposes (Zhang et al. 2002; Lacape et al. 2003; http://algodon.tamu.edu/cgi-bin/ace/searches/browser; http://www.plantgenome.uga.edu/cotton/GeneticMap.htm).

In addition to the generation of repeat-enriched libraries for the development of SSR markers, sequence databases represent valuable resources for the identification of microsatellite DNA. Such an approach has allowed the identification and the mapping of numerous SSRs in rice (McCouch et al. 2002) as well as in barley and wheat (Holton et al. 2002). The rapid increase in cotton sequences (ESTs, BAC end sequences, or STCs) deposited in public databases has also allowed the identification of microsatellite DNA-containing regions (Saha et al. 2003; http://www.genome.clemson.edu/projects/cotton/ssr/) from *G. hirsutum* and *G. arboreum*, which could be used to develop markers.

In this paper, we report on the development of a new set of microsatellite markers and the integration of additional SSR loci on the combined RFLP/SSR/AFLP genetic map of tetraploid cotton, thus providing a wider coverage of the cotton genome with this type of markers.

Materials and methods

Plant material and DNA isolation

The plant material used in this study has been fully described in Lacape et al. (2003). It consists of 75 BC_1 plants stemming from an interspecific cross between *G. hirsutum* cv. Guazuncho 2 and *G. barbadense* cv. VH8-4602, using *G. hirsutum* as the recurrent parent. The Guazuncho 2 × VH8-4602 population will be further referred to as the "GV" population.

DNA was isolated from young, fully expanded leaves using the MATAB (mixed alkyltrimethylammonium bromide) method, followed by further purification by ultracentrifugation on a cesium chloride gradient as previously described (Lacape et al. 2003).

Microsatellite-enriched library construction and marker development

Library construction and positive clone selection

The method used is based on the hybridization of biotin-labeled oligoprobes on digested genomic DNA, followed by a capture of selected sequences with streptavidin-coated magnetic beads (Kijas et al. 1994) according to the protocol described in Billotte et al. (1999) with slight modifications. In brief, total genomic DNA (10 µg) from *G. hirsutum* cv. Guazuncho 2 was restricted to completion with *Sau*3AI, and the protruding ends were polished using T4 DNA polymerase. The self-complementary 5′-phosphorylated adaptors Rsa21 (5′ CTCTTGCTTACGCGTGGACTA 3′) and Rsa25 (5′ TAGTCCACGCGTAAGCAAGAGCACA 3′) (Edwards et al. 1996) were then ligated to the blunt-ended DNA, and the ligated fragments were preamplified using oligonucleotide Rsa21 as a primer. SSR-containing fragments were selected by hybridization to biotinylated $I_5(CA)_8$ oligonucleotides and captured using streptavidin-coated paramagnetic beads. The selected fragments were further amplified using primer Rsa21, and the PCR products were cloned into plasmid pGEM-T (Promega, France). Aliquots of the ligation reaction were then used to transform *Escherichia coli* XL1-Blue MRF′ competent cells (Stratagene, The Netherlands).

In order to identify repeat-containing clones and to select for inserts above 300 bp, the insert DNA of randomly picked (white) colonies was amplified using M13 forward and reverse primers. The size of the inserts was estimated by agarose-gel electrophoresis, and the DNA was subsequently transferred onto Hybond-N^+ (Amersham, UK) membranes. The filters were then probed with γ-(^{32}P)-end-labeled $(CA)_{15}$ oligonucleotides. Only the clones containing positive inserts larger than 300 bp were retained for further analysis.

Sequence analysis and primer design

Sequencing of the positive clones was carried out at the Centre National de Séquençage, Evry, France. Vector and adaptor sequences were eliminated using an inhouse-built algorithm. In order to eliminate redundant clones, each cleaned sequence was aligned against all others by using the SEQUENCHER 4.0 program (Gene Codes, Ann Arbor, Mich., USA). The flanking primers for the unique clones were designed using the PRIMER3 software (Rozen and Skaletsky 2000) (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) with the optimum parameters set at a primer length between 15 and 27 bases, an annealing temperature of between 40°C and 65°C, and a minimum GC content of 35%. All primers were synthesized commercially (Eurogentec, Belgium).

In order to verify for duplications between our newly developed SSRs and the publicly available ones (BNL and JESPR), multiple alignments were carried out using the SEQUENCHER 4.0 program.

Amplification of SSR loci and polyacrylamide gel electrophoresis

PCR amplifications were performed in 20 µl reaction mixture containing 25 ng DNA, 0.2 mM dNTP, 2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.2 µM radiolabeled primer, 0.2 µM unlabeled primer, and 0.6 U *Taq* DNA polymerase. All PCR amplifications were performed using a PTC100 thermal cycler (MJ Research, Watertown, Mass., USA) under the following conditions: an initial denaturation step at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing temperature for 1 min, and extension at 72°C for 1 min. A final extension step of 8 min at 72°C was included. Radiolabeled PCR products were obtained by 5′-end labeling one of the primers with γ-(^{33}P) ATP using T4 polynucleotide kinase (Fermentas, Germany). After adding 20 µl loading buffer (98% formamide, 10 mM EDTA, bromophenol blue, and xylene cyanol) to the PCR reactions, the mixture was denatured at 94°C for 3 min, and 5 µl from each sample was loaded onto a 5% (w/v) polyacrylamide gel containing 7.5 M urea and electrophoresed in 0.5× TBE buffer at 55 W for 1 h, 20 min to 1 h, 50 min. After the run, gels were transferred onto Whatman 3MM paper, dried at 80°C for 20 min in a gel dryer, and exposed to X-ray films (Kodak, X-Omat) for 2 to 3 days. Fragment sizes were determined by comparison with a γ-(^{33}P)-labeled, 30–330-bp AFLP ladder (Invitrogen, France) used as a size marker on all gels.

In the first step, each primer couple was tested alone for its ability to detect polymorphism between the two parents of the mapping population. The genotyping of the BC_1 progenies with the polymorphic SSRs was performed by multiplexing two, three, or four primer couples having a similar T_m and amplifying fragments of compatible molecular weight. In this case, the PCR conditions of multiplexed microsatellites were not modified, except for a reduction in the volume of H_2O added to the reaction mixture.

Mapping of SSRs

Mapping of the SSRs was achieved by genotyping the 75 BC_1 individuals as well as the two parents and the F_1 progeny. The χ^2 test for goodness of fit was used to assess the Mendelian 1:1 inheritance of *G. hirsutum* or *Gh*, and *G. barbadense* or *Gb* allele combinations (*Gh*/*Gh* and *Gh*/*Gb*).

The combined RFLP/SSR/AFLP linkage with 888 loci (Lacape et al. 2003) was used as a backbone map for positioning additional loci. MapMaker version 3.0 (Lander et al. 1987) and the commands "group" (LOD 5, recombination fraction 30 cM), "try," "ripple," and "map" were used. In a few cases, alternate loci orders (all with the "error detection" command on) were tested using the "order" command. Recombination frequencies were converted into map distances (centiMorgans) by the Kosambi mapping function (Kosambi 1944). Homoeologous duplications, as well as comparisons between different maps for the position of common loci, were used to confirm the previously reported (Lacape et al. 2003) cases of loose linkages (c5sup, D08bot, c6sup, c6bot, c9bot, and D02sup).

Primer and locus designation

The SSR markers were designated "m*Gh*CIR" (for *m*icrosatellite *Gossypium hirsutum CIR*AD), using "CIR" as a short prefix. When different polymorphic loci were revealed with the same SSR primer pair, an a/b/c lowercase letter was assigned, starting from the higher-molecular-weight bands and moving toward the lower-weight ones. The sequences of the clones and primers were deposited in GenBank as "m*Gh*CIR" under accession numbers AJ566951 to AJ567342.

Results

Screening $(CA)_n$ repeats and designing primers

A total of 846 clones from the enriched library were sequenced. Sequence analysis showed that 202 sequences (24%) were redundant within the library. From the remaining 644 unique sequences, we were finally able to design primers for 392 (61%) of them (detailed in the Electronic Supplementary Material). All except two sequences [CIR154 with $(AACA)_4$ and CIR271 with $(GA)_{17}$] had a $(CA)_n$ or a $(GT)_n$ motif. Apart from the $(CA/GT)_n$ motif, 18% (71) of the clones contained other repeats (of 2, 3, or 4 bp). In most cases, the compound repeats were adjacent or very close.

The average number of (CA/GT) repeated units present was nine with 33 (CIR413) as a maximum (Fig. 1). The expected sizes of the products ranged from 55 bp to 350 bp. Annealing temperatures ranged from 41.1°C to 54.5°C, and were finally rounded into four groups (46, 49, 51, and 55°C) for ease of multiplexing.

In some cases, the presence of the microsatellite repeat region towards the end of the sequence resulted in 21 primer pairs designed as overlapping the repeat region (CIR398 to CIR418). These 21 SSRs did not deviate from other SSRs for the number of PCR products, although a lower hybridization specificity could have been expected.

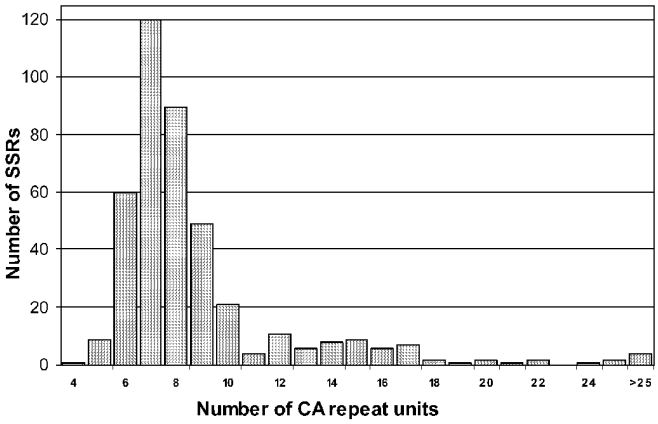

Fig. 1 Distribution of (CA) repeat length among 390 microsatellite sequences of the m*Gh*CIR [for *m*icrosatellite *Gossypium hirsutum CIR*AD (CIR)] collection

Microsatellite sequence comparison

The raw sequences of 390 unique (CA/GT) microsatellites (all except CIR154 and CIR271) reported in this work were compared with the sequences from two series of cotton microsatellites, BNL and JESPR, having CA or GT repeats. When comparing our sequences with those 76 BNL and 66 JESPR microsatellites, 17 (22%) and 9 (14%), respectively, proved to be identical. Two microsatellites were common between the three libraries (CIR184/JESPR101/BNL3590 and CIR409/JESPR242/JESPR222/BNL2634).

Evaluation of polymorphism

After discarding the 28 SSR primer pairs that did not amplify correctly, we observed that 193 of 364 remaining SSRs yielded a single PCR product on the *Gh* parental variety, 126 yielded two bands, and 43 produced three or more bands. The average number of amplified products on the *Gh* parent was 1.60, and the proportion of SSRs showing more than one PCR product was 47%.

Considering only the 364 SSRs yielding PCR products, *Gh*/*Gb* polymorphism was observed for 204 SSRs, or 56.0%. Globally, we observed no significant correlation between microsatellite length and polymorphism, as monomorphic and polymorphic SSRs had on average 8.0 and 9.6 repeats, respectively. Interestingly, we observed that the two biggest differences in allele sizes between our parents were obtained for CIR413, which shows the highest repeat number (33 CA repeats in Guazuncho 2), and for CIR120, with a compound motif of $(CT)_8$, $(CA)_8$, and $(CT)_{10}$, where the VH8-4602 product was 28 bp longer than that of Guazuncho 2.

The 204 SSRs showing at least one allele segregating in the GV population revealed 261 scorable segregating bands/loci.

Genome coverage and distribution

The segregation of 21 of the CIR loci was distorted (16 towards an excess of heterozygotes and five towards an excess of homozygotes). Nine of these were mapped along the five arm regions of c6, c12, c15, c17, and c20, previously described as covered by distorted loci (Lacape et al. 2003). Of the 261 scored VH8-specific alleles, 28 were unmapped, and the 233 mapped loci originate from 193 different SSRs—162 SSRs mapped one locus, 26 mapped two loci, three mapped three loci, one mapped four loci, and one SSR mapped six loci. Apart from two pairs of CIR loci mapping as NL6 and NL7 with no clear linkage to any known linkage group, the new SSR markers hereby described contributed to 229 additional loci mapped along the 26 chromosomes/linkage groups (Fig. 2). Table 1 details the locus composition of the updated genetic map of Guazuncho 2 × VH8-4602, including the 233 reported CIR loci. We previously re-ported six cases of putative linkages of small groups as top or bottom parts of c5, c6, c9, c20 and D02 (Lacape et al. 2003). Additional evidence for these associations results either from additional homoeologous duplications (CIR224 on c5-D08, CIR267 on c6-c25, CIR069 on A03-D02) or from the densification of the map (Fig. 2). Four (NL1 to NL4) of the five former unlinked short groups of Lacape et al. (2003) are now anchored to known chromosomes: NL1 to A01, NL2 to c26, NL3 to c10, and NL4 to c16 (Fig. 2). The number of CIR loci per chromosome varies from three on chromosomes 9 and 16 to 19 on group D08 (Table 1). Chromosome 7, previously low in microsatellite markers with only three BNL loci (Lacape et al. 2003) now has eight additional CIR loci (Fig. 2). One CIR locus, CIR376c, maps within the bottom half of c2, which was previously only covered with AFLPs. The six new CIR loci on c12 are all mapped within a 51-cM region at the top of the chromosome. Although the CIR loci were globally evenly distributed over the chromosomes, 45 mapped within the previously reported (Lacape et al. 2003) denser regions of 18 of the 26 chromosomes (Fig. 2).

Though some of the CIR markers proved to be identical to existing BNLs after sequence comparison, primers were designed independently. Of the 17 CIR/BNL redundant microsatellites, all mappable loci (ten) colocalized: CIR241 and BNL3580 on c1, CIR184 and BNL3590 on c2, CIR364 and BNL852a on c5, CIR210 and BNL3932 on c14, CIR180 and BNL834 on c17, CIR391 and BNL3510b on c26, CIR209 and BNL3792 on A02, CIR416 and BNL3442 on A03, and lastly, CIR165 and BNL3452, as well as CIR222a and BNL3535 on D08 (Fig. 2).

As compared to our previous map (Lacape et al. 2003) that contains 888 loci and spans 4,397 cM, the updated GV linkage map, including the 233 new CIR loci as well as 42 candidate genes (unpublished results from our laboratory), now shows 1,160 loci and reaches 5,519 cM. (Table 1). The groups belonging to the A and D subgenomes account for 3,018 cM (638 loci) and 2,414 cM (511 loci), respectively. The 13 homeologous pairs span 5,432 cM with 1,149 loci (4.7 cM as an average distance between two loci), and the four unassigned groups, namely NL5 to NL8, span 86.2 cM with 11 loci. The CIR loci mapped on either ends of 14 of the 26 chromosomes account for an extension of 390 cM of the map; in particular, the bottoms of c3, c14 (five CIR and BNL3932 as new loci), and c5 are all extended by over 30 cM (Fig. 2). The remaining extension of 732 cM resulted from both the coalescing of the 37 linkage groups (Lacape et al. 2003) into 31 groups (accounting for an extension of 465 cM) and from a stretching in distances between common loci after the mapping of additional loci (this accounts for an extension of 267 cM).

Half of the duplicated CIR markers bridge the homoeologous A/D chromosomes, thus increasing the corre-

Table 1 Locus composition and recombination distances of chromosomes and linkage groups (LG) in the genetic map of *Gossypium hirsutum* cv. Guazuncho 2 × *G. barbadense* cv. VH8-4602. Chromosomes or LGs are ordered by pairs of A/D homoeologies (for example, homoeologous chromosomes 1 and 15 of A and D subgenomic affinity, respectively), and cumulative values of the A and D subgenomes. The number of new m*Gh*CIR [for *m*icrosatellite *G. hirsutum CIR*AD (CIR)] microsatellite loci is given. *NL* Unassigned groups

	Total loci	CIR loci	Size (cM)	Average distance (cM)
Chromosome/LG no. (A subgroup)				
c1	35	7	205.1	5.9
c2	42	4	195.9	4.7
c3	46	10	152.7	3.3
c4	37	9	189.5	5.1
c5	54	15	360.4	6.7
c6	46	10	296.2	6.4
c7	49	8	206.8	4.2
c9	43	3	287.1	6.7
c10	43	8	191.7	4.5
c12	55	6	185.6	3.4
A01	63	11	232.5	3.7
A02	60	9	243.5	4.1
A03	65	10	271.3	4.2
A subgroup	638	110	3,018.3	4.7
Chromosome/LG no. (D subgroup)				
c15	43	11	185.6	4.3
c14	44	12	197.0	4.5
c17	30	6	90.8	3.0
c22	29	7	138.9	4.8
D08	56	19	236.5	4.2
c25	41	7	183.0	4.5
c16	21	3	165.8	7.9
c23	38	6	172.6	4.5
c20	41	11	268.3	6.5
c26	35	11	195.4	5.6
c18	43	5	157.5	3.7
D03	42	9	132.4	3.2
D02	48	12	290.2	6.0
D subgroup	511	119	2,414.0	4.7
Chromosome/LG no.				
NL5	5		53.0	10.6
NL6	2	2	8.2	4.1
NL7	2	2	8.3	4.2
NL8	2		16.7	8.4
Total NL	11	4	86.2	7.8
A subgroup	638	110	3,018.3	4.7
D subgroup	511	119	2,414.0	4.7
Total	1,160	233	5,518.5	4.8

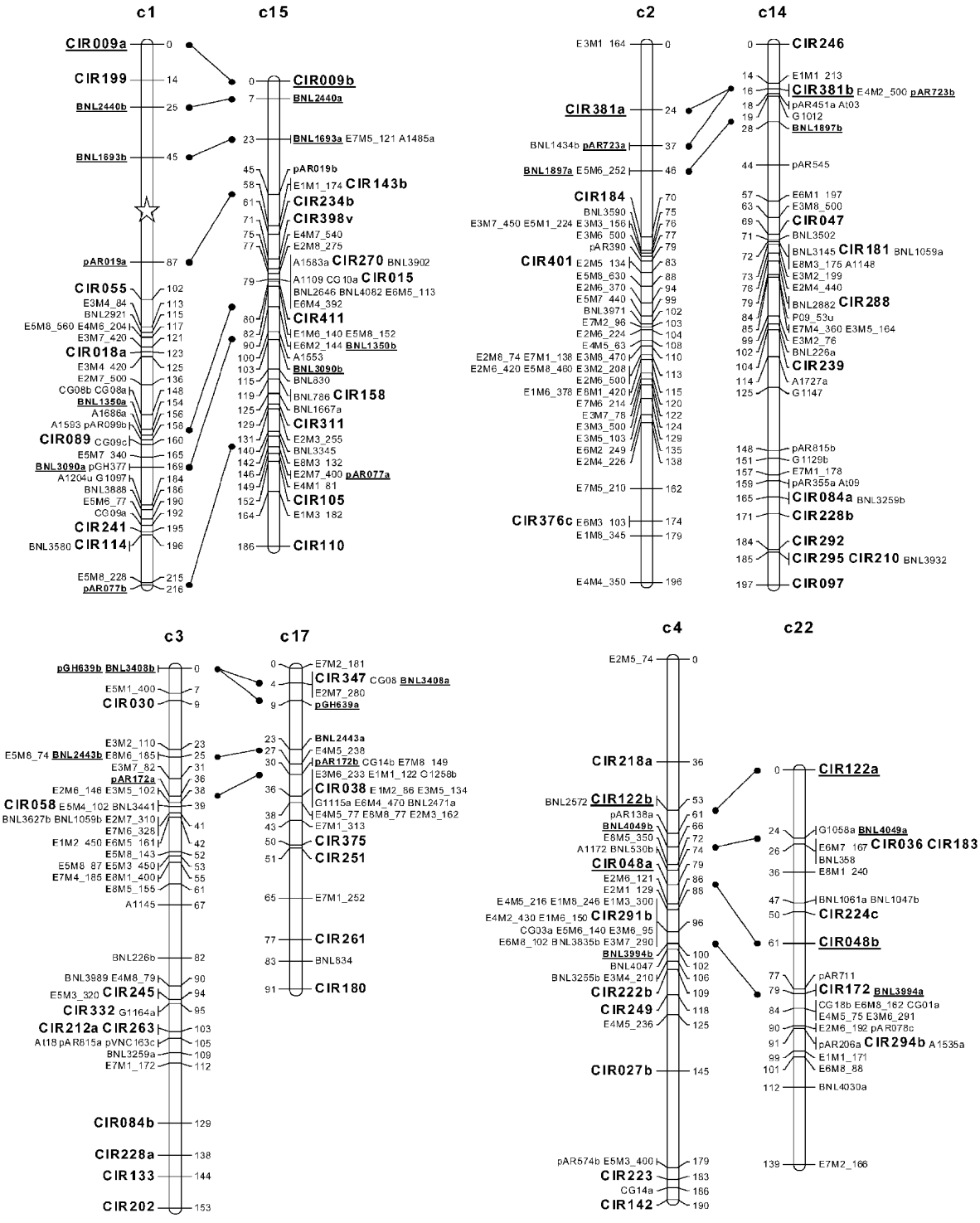

Fig. 2 The updated linkage map of *Gossypium hirsutum* cv. Guazuncho 2 × *G barbadense* cv. VH8-4602 showing the positions of newly mapped CIR loci (*enlarged* and *bolded characters*) along 26 chromosomes/linkage groups. Chromosomes and linkage groups are associated in 13 homoeologous pairs, following the nomenclature of Lacape et al. (2003). The map also contains cotton RFLP markers (prefix *A*, *G*, *M*, *P*, *pAR*, *pGH*, *pVNC*, and *pXP*), candidate genes (prefix *CG*), *Arabidopsis thaliana* RFLP markers (prefix *At*), BNL microsatellites, AFLPs derived from *Eco*RI/*Mse*I (*ExMy* prefix followed by fragment size in base pairs), and two morphological markers—petal color (*PETCOL*) and linter color (*LTCOL*) (see Lacape et al. 2003 for details). When duplicated (*underscored letter attached to the locus name*) between homoeologous chromosomes, locus names are *boldface* and *underlined*, and their positions are connected by a *solid bar*. *Star symbols* on c1, c5, c6, c7, c10, c20, c26, D02, and D03, indicate linkages at LOD values superior to LOD2 and inferior to LOD5

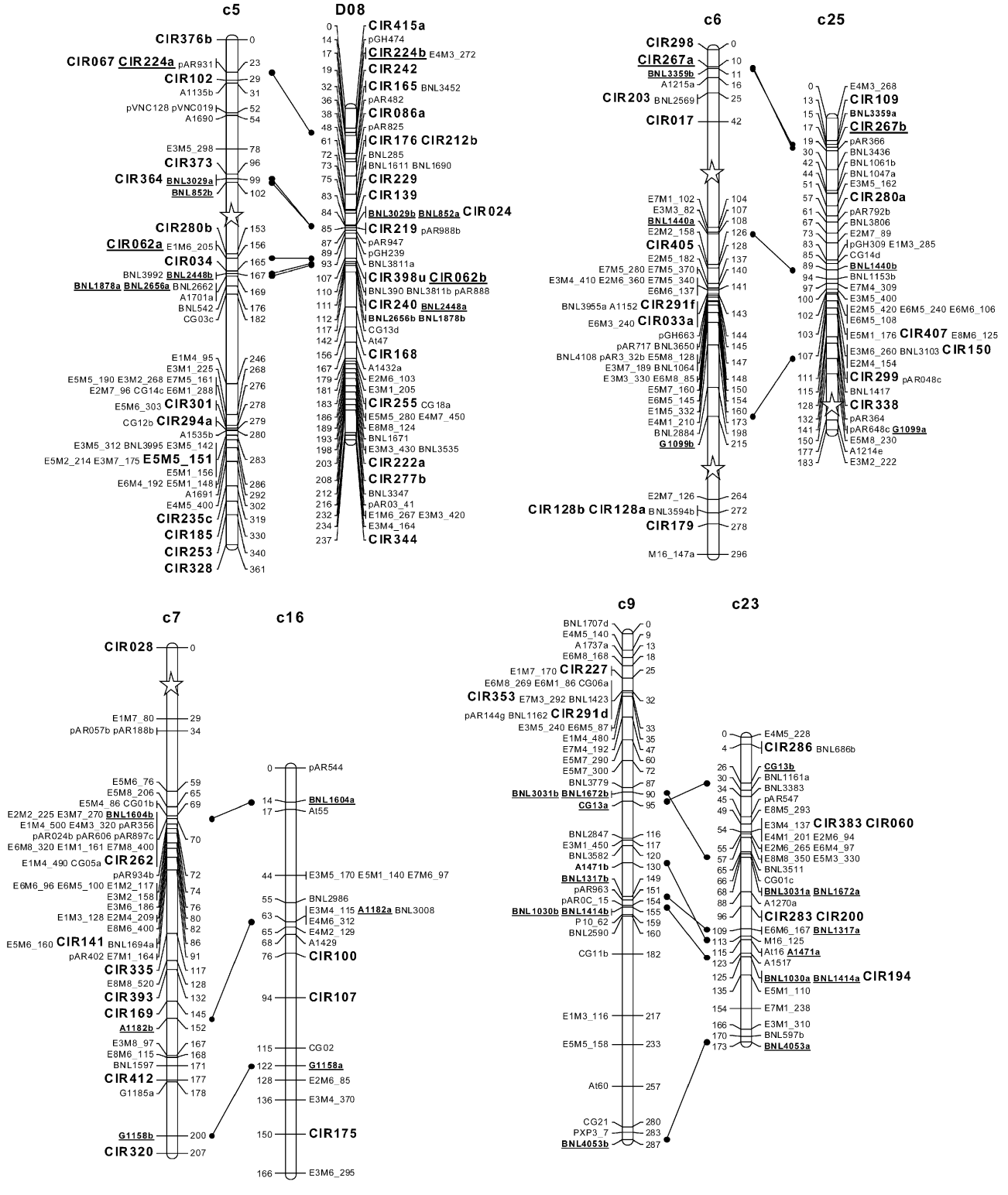

Fig. 2 (continued)

sponding numbers of reported duplications (Reinish et al. 1994; Lacape et al. 2003). Two markers (CIR171 and CIR305) bridge the pair c10/c20, which had no homoeologous duplication in Lacape et al. (2003). Four new markers substantiate the two known translocations involving c2 and c3 (with CIR084 and CIR228) or c4 and c5 (with CIR222 and CIR294) (Endrizzi et al. 1985; Reinish et al. 1994; Lacape et al. 2003).

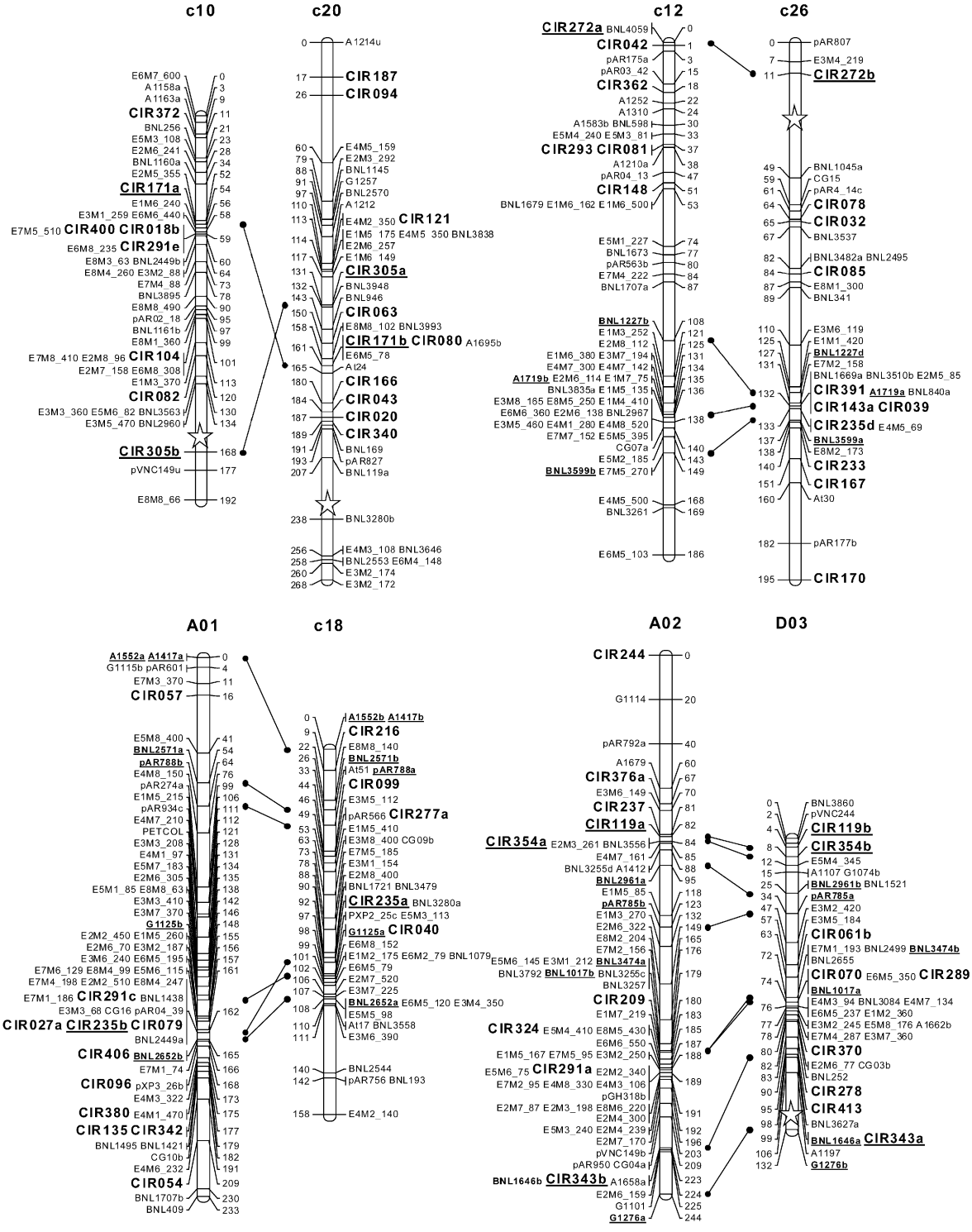

Fig. 2 (continued)

Discussion

Generally in plants (unlike in mammals), the (GA/CT) microsatellite motifs are more frequent than the (CA/GT) motifs (Lagercrantz et al. 1993). In contrast to this general trend, in taro (Mace and Godwin 2002), black poplar (van der Schoot et al. 2000), and some Poaceae genomes (Jones et al. 2001), (CA/GT) motifs were more abundant. Though possibly biased by the fact that the enrichment procedure used a mixture of different repeat motifs, Reddy et al. (2001) found that the most frequently occurring microsatellite motif in *G. hirsutum* genomic

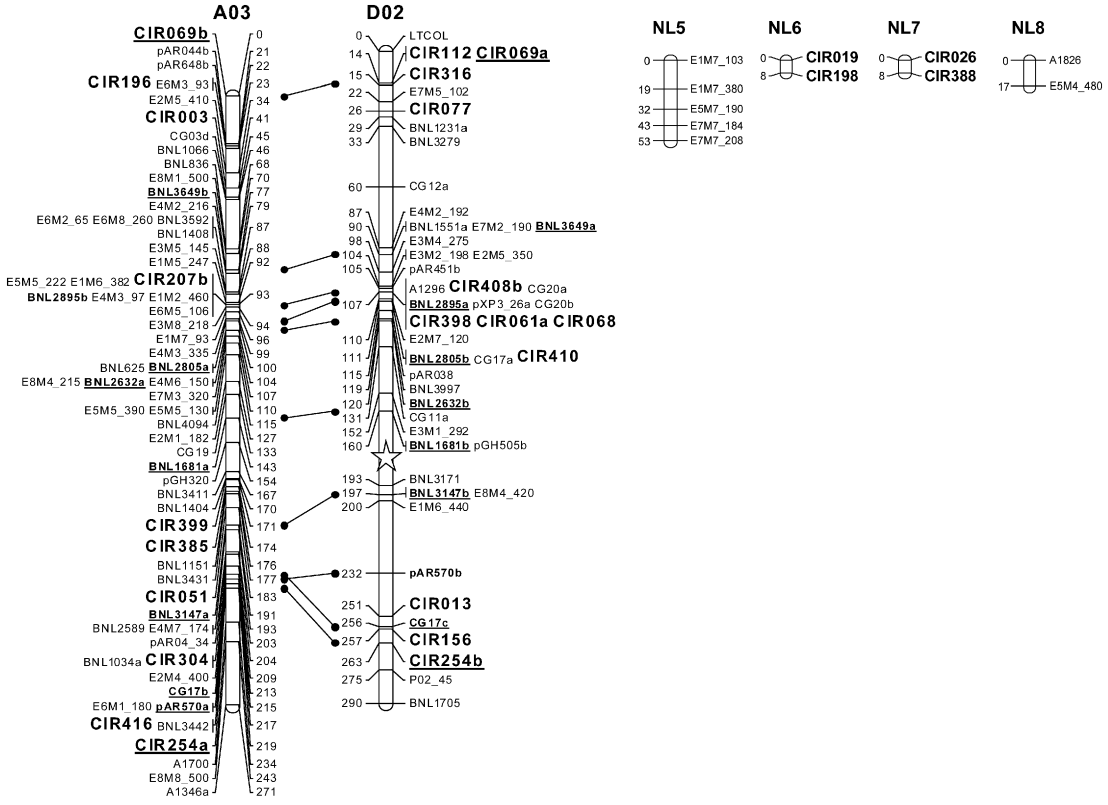

Fig. 2 (continued)

DNA was (AAG) (around 32% of all ten different types of di- to hexanucleotide repeats), followed by (GA) (18%) and (CA) (14%). In their study on cotton EST-SSRs, Saha et al. (2003) found that in *G. hirsutum*, the (GA/CT) motif was the most abundant (24%), followed by the (AAG/TTC) (18%) and the (AGT/TCA) (15%) trinucleotide motifs. In contrast, in *G. arboreum* ESTs, the trinucleotide repeat (AAG/TTC) was the most abundant (25%), followed by (GA/CT) with 21%, and (AT/TA) with 17%. We decided to develop microsatellite markers containing (CA/GT) repeats, as this motif was an under-represented class [140 (GA/CT) and 76 (CA/GT)] in the only existing library of cotton SSRs (BNL) at the start of this research. We isolated a set of 392 microsatellite-containing clones, with a high rate of successful amplification (93%). The number of (CA) repeats of the new SSRs is nine on average, which is inferior to that reported in Reddy et al. (2001), where the majority of the repeated (CA/GT) units ranged between 10 and 20. We aligned the sequences of (CA/GT) microsatellites developed from three independent libraries that all used a similar protocol (enzyme-shearing of *G. hirsutum* DNA, followed by hybridization with biotin-labeled oligoprobes and captured with streptavidin-coated magnetic beads). The libraries consist of 390 (CIR library), 66 (JESPR library), and 76 (BNL library) sequences containing (CA/GT) microsatellites. They have nine (CIR and JESPR), 17 (CIR and BNL) and ten (JESPR and BNL) sequences in common.

Our rate of polymorphism (56%) detected between *G. hirsutum* and *G. barbadense* is slightly superior to that reported by Reddy et al. (2001), 45.4% between *G. hirsutum* cv. TM1 and *G. barbadense* cv. Pima 3-79 using 152 SSRs, and that of Reinish et al. (1994), who detected 46.2% polymorphism between *G. hirsutum* race *palmeri* and *G. barbadense* acc. K101 using 550 RFLP probes. Using 64 different AFLP primer couples, an average of 11 VH8-4602-specific markers out of 91 bands per primer couple was found, allowing for an estimation of a 22% polymorphism (Lacape et al. 2003). The 204 polymorphic microsatellites revealed 261 segregating alleles, of which 233 were integrated to the previously published map (Lacape et al. 2003). Increasing the number of loci by 30% (1,160 now in total) resulted in a 25% increase in overall map size (5,500 cM in total). Although the majority of this increase was due to a linear extension of either distal sides of chromosomes and to the coalescing of small groups into larger ones, a 6% stretching in distances was a result of a higher density of markers on chromosomes. While coding/typing errors, whose impact on distance inflation increases with marker density (Lincoln and Lander 1992), cannot be excluded, this stretching could result from the fact that we have added loci to an unsaturated map which comprised 37 linkage groups for 26 chromosomes.

The map now comprises 1,160 loci that map along 26 major linkage groups and five small ones (that account for 1% of the loci). The overall map size is of 5,519 cM, with

an average marker distance of 4.7 cM, calculated using the complete set of loci, and of 7.5 cM after excluding 400 colocalized loci. With respect to microsatellite markers, the map includes a total of 466 loci, equally partitioned between BNL and CIR. The number of SSR loci varies from six on c16 to 34 on D08, with an average genome-wide density of one SSR locus per 12 cM. Forty-four cases of gaps larger than 30 cM (and 19 larger than 50 cM) between two SSRs are found along the 26 chromosomes. Due to their fairly even distribution throughout the genome, the microsatellite markers reported in this work could serve as a useful tool for an effective tracking of QTLs and for marker-assisted breeding. In addition, the alignment of linkage maps across different populations produced in different laboratories using SSRs as framework markers will provide a powerful means for the production of a high-density, consensus linkage map of cotton, as it has been achieved in other crops (Cregan et al. 1999; Danin-Poleg et al. 2000; Cone et al. 2002; Ruiz and Asins 2003).

References

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

Billotte N, Lagoda PJ, Risterucci A-M, Baurens F-C (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits 54:277–288

Brubaker CL, Wendel JF (2001) RFLP diversity in cotton. In: Jenkins JN, Saha S (eds) Genetic improvement of cotton, emerging technologies. Science, Enfield, N.H., USA, pp 81–101

Brubaker CL, Cantrell RG, Giband M, Lyon BR, Wilkins TA (2000) Letter to Journal of Cotton Science community: formation of the International Cotton Genome Initiative, ICGI. J Cotton Sci 4:149

Cone KC, McMullen MD, Bi IV, Davis GL, Yim YS, Gardiner JM, Polacco ML, Sanchez-Villeda H, Fang Z, Schroeder SG, Havermann SA, Bowers JE, Paterson AH, Soderlund CA, Engler FW, Wing RA, Coe EH Jr (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. Plant Physiol 4:1598–1605

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

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Danin-Poleg Y, Reis N, Baudrasso-Arnas S, Pitrat M, Staub JE, Oliver M, Arus P, deVicente CM, Katzir N (2000) Simple sequence repeats in *Cucumis* mapping and map merging. Genome 43:963–974

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758–760

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

Holton TH, Christopher JT, McClure L, Harker N, Henry RH (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Mol Breed 9:63–71

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterization of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the *Citrus* genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16:656–662

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 [Suppl]:28

Lacape J-M, Nguyen TB, Thibivilliers S, Courtois B, Bojinov BM, 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

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Lander E, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, Newburg L (1987) MapMaker: an interactive computer package for constructing primary linkage maps of experimental and natural populations. Genomics 1:174–181

Lincoln SE, Lander ES (1992) Systematic detection of errors in genetic linkage data. Genomics 14:604–610

Mace ES, Godwin ID (2002) Development and characterization of polymorphic microsatellite markers in taro (*Colocasia esculenta*). Genome 45:823–832

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, MW, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerk G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2,240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

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

Reinish A, Dong J-M, Brubaker CL, Stelly DM, 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

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana, Totowa, N.J., pp 365–386

Ruiz C, Asins MJ (2003) Comparison between *Ponsirus* and *Citrus* linkage maps. Theor Appl Genet 106:826–836

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

Schoot J van der, Popiskova M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theor Appl Genet 101:317–322

Shappley Z, Jenkins JN, Meredith W, McCartty JCJ (1998) An RFLP linkage map of upland cotton, *Gossypium hirsutum* L. Theor Appl Genet 97:756–761

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

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Ulloa M, Meredith W, Shappley Z, Kahler A (2002) RFLP genetic linkage maps from four $F_{2.3}$ populations and a JoinMap of *Gossypium hirsutum* L. Theor Appl Genet 104:200–208

Yu J, Kohel RJ (2001) Cotton genome research in the United States. In: Jenkins JN, Saha S (eds) Genetic improvement of cotton. Emerging technologies. Science, Enfield, N.H., USA, pp 103–122

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