Euphytica (2012) 187:203–213
DOI 10.1007/s10681-012-0643-y

Development of a core set of SSR markers for the characterization of *Gossypium* germplasm

John Z. Yu · David D. Fang · Russell J. Kohel · Mauricio Ulloa · Lori L. Hinze · Richard G. Percy · Jinfa Zhang · Peng Chee · Brian E. Scheffler · Don C. Jones

Received: 28 September 2011 / Accepted: 13 February 2012 / Published online: 27 March 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Molecular markers such as simple sequence repeats (SSR) are a useful tool for characterizing genetic diversity of *Gossypium* germplasm. Genetic profiles by DNA fingerprinting of cotton accessions can only be compared among different collections if a common set of molecular markers are used by different laboratories and/or research projects. Herein, we propose and report a core set of 105 SSR markers with wide genome coverage of at least four evenly distributed markers per chromosome for the 26 tetraploid cotton chromosomes. The core marker set represents the efforts of ten research groups involved in marker development, and have been systematically evaluated for DNA polymorphism on the 12 genotypes belonging to six *Gossypium* species [known collectively as the cotton marker database (CMD) panel]. A total of 35 marker bins in triplex sets were arranged from the 105 markers that were each labeled with one of the three fluorescent dyes (FAM, HEX, and NED). Results from this study indicated that the core marker set was robust in revealing DNA polymorphism either between and within species. Average value of polymorphism information content (PIC) among the CMD panel was 0.65, and that within the cultivated cotton

Electronic supplementary material The online version of this article (doi:10.1007/s10681-012-0643-y) contains supplementary material, which is available to authorized users.

J. Z. Yu · R. J. Kohel · L. L. Hinze · R. G. Percy
Crop Germplasm Research Unit,
USDA-ARS-SPARC, 2881 F&B Road,
College Station, TX 77845, USA

J. Z. Yu (✉)
Southern Plains Agricultural Research Center,
USDA-ARS, 2881 F&B Road, College Station,
TX 77845, USA
e-mail: john.yu@ars.usda.gov

D. D. Fang
Cotton Fiber Bioscience Research Unit,
USDA-ARS-SRRC, 1100 Robert E. Lee Blvd,
New Orleans, LA 70124, USA

M. Ulloa
Shafter Cotton Research Station,
USDA-ARS-WICSRU, 17053 N. Shafter Ave,
Shafter, CA 93263, USA

J. Zhang
Department of Plant and Environmental Sciences,
New Mexico State University, Skeen Hall N338,
Las Cruces, NM 88003, USA

P. Chee
Molecular Cotton Breeding Laboratory, University
of Georgia, 2356 Rainwater Road, Tifton,
GA 31793, USA

B. E. Scheffler
Genomics and Bioinformatics Research Unit,
USDA-ARS, 141 Experiment Station Road,
Stoneville, MS 38776, USA

D. C. Jones
Agricultural and Environmental Research,
Cotton Incorporated, 6399 Weston Parkway,
Cary, NC 27513, USA

species *Gossypium hirsutum* was 0.29. Based on the similarity matrix and phylogenetic analysis of the CMD panel, the core marker set appeared to be sufficient in characterizing the diversity within *G. hirsutum* and other *Gossypium* species. The portability of this core marker set would facilitate the systematic characterization and the simultaneous comparison among various research efforts involved in genetic diversity analysis and germplasm resource preservation.

Keywords Cotton (*Gossypium* spp.) · Chromosome location · Genetic diversity · Polymorphism information content (PIC) · Simple sequence repeats (SSR) markers

Abbreviations

BAC	Bacterial artificial chromosome
CMD	Cotton marker database
EST	Expressed sequence tag
PIC	Polymorphism information content
QTL	Quantitative trait loci
SSR	Simple sequence repeats

Introduction

From the perspective of germplasm utilization, genetic diversity is viewed in terms of genes that can be utilized for the improvement of the cultivated species. In this sense, more emphasis is placed on genes related to environmental fitness and physiological processes than on the morphological characteristics important in taxonomic classification. Thus, it is imperative to take advantage of the available gene pools for cotton improvement to meet demands for enhanced crop productivity while simultaneously increasing the fitness of the crop in response to changing environmental factors. The domesticated and wild species of the *Gossypium* genus represent an enormous genetic storehouse for potential exploitation by cotton breeders and geneticists. However, genetic diversity within *Gossypium* germplasm collections is underutilized as, in the U.S., less than 1% of the genetic diversity is exploited (Esbroeck and Bowman 1998; Wallace et al. 2009). This is in part due to the shortage of effective genomic tools that can discern and define the patterns of variability available to the breeders and geneticists.

The *Gossypium* genus has about 50 species, including the four domesticated species (*G. arboreum* L., *G. barbadense* L., *G. herbaceum* L., and *G. hirsutum* L.), that embody considerable genetic diversity across a wide geographic range covering most tropical and subtropical regions of the world (Percival et al. 1999; Wendel and Cronn 2003; Campbell et al. 2010). Based on pairing relationships of 13 chromosomes, the diploid *Gossypium* species are grouped into A through G and K genomes (Endrizzi et al. 1985; Stewart 1994). The allotetraploid species are the products of a presumed single polyploidization event between the A-genome and D-genome diploids that occurred around 1–2 million years ago (Wendel and Cronn 2003). The two diploid species *G. arboreum* and *G. herbaceum* ($n = x = 13$) of the A-genome are known as Old World Asiatic cottons, while two allotetraploid species *G. hirsutum* and *G. barbadense* ($n = 2x = 26$) of the AD-genome are known as New World cottons (Lee 1984). The number of 26 disomic pairing gametic chromosomes in allotetraploid cotton species is the largest among all major crop plant species (Yu and Kohel 2001; Stelly et al. 2005). Cultivated cotton is the world's leading fiber crop and the second most important oilseed crop.

As DNA markers, genome maps, and other molecular tools are being developed for cotton, it becomes possible to look at the genotypes (not just phenotypes) of accessions of the *Gossypium* germplasm collections to identify redundancies and duplications, and to provide guidance for evaluation and selection. Many novel alleles can thus be discovered that are otherwise buried in the background noise of the genome, but may contribute to the well-being of the cotton plant (Tanksley and McCouch 1997). To make significant and timely advances in genetic improvement of cotton, portable DNA markers based on polymerase chain reaction (PCR) are needed for the tetraploid genome of cultivated cottons. Over the last decade, thousands of PCR-based DNA markers including simple sequence repeat (SSR) markers were identified and many were mapped in the cotton genome (Blenda et al. 2006; Guo et al. 2007; Lacape et al. 2009; Yu et al. 2011; Yu et al. 2012). These cotton SSR markers were developed either from random enriched small genomic clones, large insert bacterial artificial chromosome (BAC) clones (or physical contigs) or expressed sequence tags (ESTs) of *Gossypium* species. Ad hoc studies to characterize *Gossypium* germplasm were conducted

that used various sets of mapped and/or unmapped SSR markers (Liu et al. 2006; Lacape et al. 2007; Abdurakhmonov et al. 2008, 2009). Subsets of different collections that represent the range of diversity of *G. hirsutum* and outliers of other *Gossypium* species were subjected to molecular characterization.

Over the course of such studies, an understanding of the germplasm structure emerged and valuable information was generated toward developing a core set of markers. However, comparative analyzes and systematic characterizations of the genus were made difficult and potential information was lost due to inability to make direct comparisons between studies and the inability to combine data sets. Fragmented data and information on germplasm characterization should be coordinated and integrated as a whole. Coordination with the cotton research community would require a consensus on the common set of portable DNA markers (Kohel and Yu 2002). A strategy was proposed that at least four markers would be necessary to cover both arms of individual chromosomes (Fig. 1). The objective of this study was to identify and validate an appropriate set of core SSR markers to serve as initial molecular descriptors to characterize *Gossypium* germplasm collections and as desirable genomic tools for other studies including marker-assisted selection in cotton.

Materials and methods

Plant materials and DNA extraction

Table 1 shows the 12 cultivated and exotic cotton genotypes representing six *Gossypium* species that were assembled and maintained as individual plants in a USDA-ARS greenhouse in College Station, Texas, USA. The 12 diverse genotypes provide the cotton research community with a manageable reference panel of core *Gossypium* germplasm, which are the standards for the cotton marker database (CMD) and other cotton genomic studies (Yu 2004; Blenda et al. 2006). The cotton genotype panel consists of the tetraploid genetic standards for *G. hirsutum* (upland inbred TM-1) and *G. barbadense* (double haploid Pima 3–79), which were also used as parents for a recombinant inbred (RI) population. In addition, the panel includes two wild tetraploid introgression species (*G. tomentosum* Nuttall ex Seemann and *G. mustelinum* Miers ex Watt). The panel also includes five commercially grown *G. hirsutum* cultivars, i.e., Acala Maxxa (PVP 9000168), Delta and Pine Land (DPL) 458BR (PVP 9800206), Paymaster (PM) 1218BR (PVP 200000213), Fibermax 832 (PVP 9800258) and Stoneville (STV) 4892BR (PVP 200000253), representing major cottonseed companies in the U.S.; one

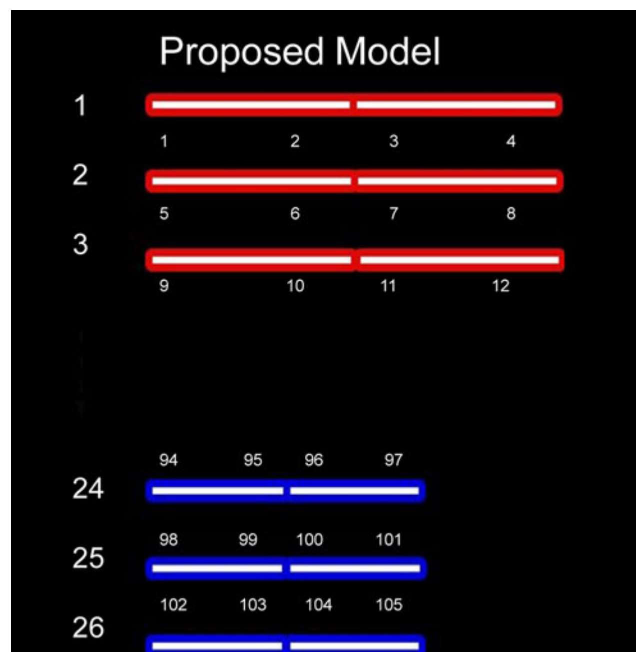

Fig. 1 Schematic illustration of core DNA markers distributed on the 26 chromosomes of tetraploid cotton genomes (Kohel and Yu 2002)

Table 1 A reference panel of *Gossypium* germplasm standardized for cotton marker database (CMD) (Yu 2004)

CMD	Scientific name	Common name	Description
CMD01	*G. hirsutum*	TM-1	AD1 genetic standard (RI parent)
CMD02	*G. barbadense*	3–79	AD2 genetic standard (RI parent)
CMD03	*G. hirsutum*	Acala Maxxa	California upland cotton (AD1)
CMD04	*G. hirsutum*	DPL 458BR	Upland cotton (AD1) with significant acreage
CMD05	*G. hirsutum*	Paymaster 1218BR	Upland cotton (AD1) with significant acreage
CMD06	*G. hirsutum*	Fibermax 832	Upland cotton (AD1) with significant acreage
CMD07	*G. hirsutum*	Stoneville 4892BR	Upland cotton (AD1) with significant acreage
CMD08	*G. barbadense*	Pima S-6	Pima cotton (AD2) germplasm breeding source
CMD09	*G. arboreum*	*G. arboreum* (A2-8)	Diploid A-genome representative
CMD10	*G. raimondii*	*G. raimondii* (D5-3)	Diploid D-genome representative
CMD11	*G. tomentosum*	*G. tomentosum* (AD3)	Introgression breeding source
CMD12	*G. mustelinum*	*G. mustelinum* (AD4)	Introgression breeding source

G. barbadense cultivar (Pima S-6); and two ancestral diploid genome representatives (*G. arboreum*, A_2-8 and *G. raimondii* Ulbrich, D_5-3). This panel represents a balanced diversity of *Gossypium* germplasm that are important in cotton cultivar improvement and that are well accepted by the cotton research community (Yu 2004; Blenda et al. 2006).

Genomic DNA was extracted from fresh young leaf tissue of individual cotton plants grown in the greenhouse following the modified CTAB DNA extraction procedure as described by Kohel et al. (2001) and Yu et al. (2012).

SSR primers and PCR assays

The SSR primer pairs used in this study were developed by different research groups of the cotton community (Blenda et al. 2006). A total of 105 pairs of primers were identified that represented ten diverse sources (BNL, CIR, DPL, GH, JESPR, MGHES, MUSB, NAU, STV, and TMB) of cotton SSR discovery efforts (Table 2). SSRs exhibit minimal cross interaction among the primer pairs, least overlapping of PCR products or fragment sizes, good binding capacity to amplify genomic DNA, sufficient power to detect DNA polymorphism of *Gossypium* species, and even distribution among the 26 chromosomes of the tetraploid cotton. Each chromosome was represented by four SSR loci except for the genetically longest chromosome (c05) that had five SSR loci on the basis of the genetic map developed from the *G. hirsutum* TM-1/*G. barbadense* 3–79 RIL population(Yu et al. 2012). The sequence of individual primer pairs and the source clone for each SSR can be found in the CMD (http://www.cottonmarker.org/).

Table 2 Source and number of cotton SSR primer pairs identified for the core marker set (http://www.cottonmarker.org/)

Code prefix	SSR number	Original source
BNL	39	Brookhaven National Laboratory
TMB	17	USDA-ARS, College Station, TX
DPL	14	Delta and Pine Land Company
CIR	13	CIRAD, Montpellier, France
JESPR	9	Texas A&M University and USDA-ARS (MS)
GH	6	Texas A&M University and USDA-ARS (TX)
MGHES	3	USDA-ARS, Starkville, MS
NAU	2	Nanjing Agricultural University, China
MUSB	1	USDA-ARS, Shafter, CA
STV	1	USDA-ARS, Stoneville, MS
Total	105	

Each of the 105 individual forward primers was labeled at the 5′ end with one of the three fluorescent dyes (FAM, HEX, or NED) to allow the creation of 35 triplex PCR bin sets. Three pairs of primers with different dyes were multiplexed in each PCR reaction according to Fang et al. (2010). All PCR experiments were conducted using the same protocol and temperature profile described by Fang et al. (2010). Amplified DNA products were separated with GeneScan-500 Rox® as an internal DNA size standard in

capillary electrophoresis of an automated Genetic Analyzer 3730xl (Applied Biosystems/Life Technology, Foster City, CA, USA).

Data analysis

Individual PCR fragments of each cotton accession were recorded in actual base pairs (bp) with the ABI GeneMapper 4.0 software. The polymorphism information content (PIC) values of each SSR marker was calculated according to Botstein et al. (1980) using the PowerMarker 3.0 program (Liu and Muse 2005). SSR products were treated as presence or absence when calculating a similarity coefficient. The Jaccard coefficient was used to compute a similarity matrix using a server running the program DendroUPGMA (http://genomes.urv.es/UPGMA/) (Garcia-Vallvé et al. 1999). This matrix was used to construct a dendrogram using the program PhyloWidget (http://www.phylowidget.org/) (Jordan and Piel 2008).

Results and discussion

Genome distribution of a core SSR marker set

A recently developed tetraploid genetic linkage map with robust framework of SSR-linked loci (Yu et al. 2012) was used to identify the SSR markers used in this study. Four SSR markers were identified from each one of the 26 cotton chromosomes, except for chromosome 5 (c05). Based on the genetic map of Yu et al. (2012), c05 was the longest in genetic distance and thus was represented with five SSR markers. Table 3 shows the chromosome location of each SSR marker as placed on cotton linkage groups/chromosomes by Yu et al. (2012). An approximate 30 cM interval between SSRs was selected as the criteria to ensure even distribution of the markers on each chromosome. Even though telomeric regions of plant chromosomes are primarily heterochromatic, they are important not only in maintaining the integrity of chromosome structure but also in harboring the genes of interest (Gill et al. 1993; Fang et al. 2010). Six markers were selected from these telomeric regions, including TMB1648 on c04, JESPR119 on c06, JESPR291 on c08, TMB2295 on c18, BNL1047 and DPL0520 on c25 (Table 3). Other criteria to balance the marker distribution in the genome included single-copy versus multiple-copy markers and recombination-rich versus recombination-poor regions for potential tapping into gene islands that may be buried in the cotton genome (Xu et al. 2008).

Polymorphism information content (PIC) values of core SSR markers

Each of the SSR markers from the core set was examined for its PCR specificity and DNA polymorphism on the CMD panel of 12 core *Gossypium* genotypes (Table 1). Analysis of the CMD panel and other *Gossypium* germplasm accessions demonstrated the ability of the core markers to reveal genetic diversity within the collections (Abdurakhmonov et al. 2008). Table 3 shows the PIC values of each core SSR marker among all the 12 *Gossypium* genotypes and within the six *G. hirsutum* cultivars and accessions. As expected, the core SSR marker set detected higher levels of genetic diversity among the *Gossypium* species than within the *G. hirsutum* genotypes. PIC values of the core SSR marker set on the complete CMD panel ranged from 0.23 to 0.88, with an average value of 0.65. The majority (90) of SSRs from the 105 core SSR marker set were highly informative (PIC >0.50) and 13 SSR markers were moderately informative ($0.25 < \text{PIC} < 0.50$), while only two were slightly informative (PIC <0.25). PIC values of the core SSR marker set on six *G. hirsutum* genotypes ranged from 0 to 0.84, with the average value of 0.29. Thirty-seven markers specifically detected genetic diversity among the six *Gossypium* species but not within the six *G. hirsutum* cultivars (PIC $= 0$). Of the 68 markers that detected *G. hirsutum* diversity, 27 were highly informative (PIC >0.50), 30 moderately informative ($0.25 < \text{PIC} < 0.50$) and 11 slightly informative (PIC <0.25). These results can be used to guide researchers when they plan individual studies with this core SSR marker set. For example, SSR markers such as BNL3545 on c02 and BNL1531 on c16 had very high PIC values not only among the six *Gossypium* species (0.88 and 0.76, respectively) but also within the *G. hirsutum* species (0.84 and 0.77, respectively). Such highly informative SSR markers usually resulted from multiple copies of PCR products, and would be of particular interest in identification of unique DNA profiles and potentially novel genes within the cultivated cottons (*G. hirsutum*) (Botstein et al. 1980; Tanksley and McCouch 1997).

Table 3 Genomic location, PIC value and triplex set of cotton core SSR markers

Marker #	Marker name[a]	Chromosome #	Map position (cM)	CMD PIC	Upland PIC	Dye label	Triplex#
1	BNL0409	c13	13.43	0.66	0.24	FAM	1
2	BNL0387	c24	30.31	0.82	0.55	HEX	1
3	CIR169[a]	c07	57.28	0.48	0.00	NED	1
4	BNL0569	c18	60.36	0.74	0.55	FAM	2
5	BNL1551[a]	c21	110.54	0.63	0.24	HEX	2
6	BNL0530	c04	61.46	0.72	0.00	NED	2
7	BNL1231	c11	140.86	0.66	0.24	FAM	3
8	BNL1495[a]	c13	56.78	0.57	0.38	HEX	3
9	CIR081	c12	94.94	0.59	0.35	NED	3
10	BNL1417	c25	46.74	0.71	0.35	FAM	4
11	BNL1047	c25	126.88	0.57	0.00	HEX	4
12	BNL0256	c10	92.40	0.74	0.48	NED	4
13	BNL1521[a]	c24	61.28	0.69	0.24	FAM	5
14	BNL2570[a]	c20	36.52	0.68	0.35	HEX	5
15	JESPR153[a]	c13	75.00	0.82	0.56	NED	5
16	BNL2443	c17	92.59	0.60	0.00	FAM	6
17	BNL2572[a]	c04	82.71	0.63	0.24	HEX	6
18	TMB1295	c19	139.97	0.81	0.00	NED	6
19	BNL2495[a]	c26	62.72	0.73	0.35	FAM	7
20	BNL1897[a]	c02	86.46	0.63	0.00	HEX	7
21	BNL3090[a]	c15	51.86	0.82	0.55	NED	7
22	BNL2496	c17	7.13	0.76	0.00	FAM	8
23	BNL1673	c12	76.17	0.62	0.00	HEX	8
24	DPL0168	c16	86.25	0.63	0.54	NED	8
25	BNL2544[a]	c18	15.02	0.39	0.00	FAM	9
26	BNL1531	c16	37.90	0.76	0.77	HEX	9
27	BNL3644	c14	122.54	0.70	0.54	NED	9
28	BNL2960[a]	c10	40.73	0.74	0.38	FAM	10
29	BNL3441[a]	c03	68.58	0.56	0.00	HEX	10
30	DPL0039	c26	24.23	0.73	0.54	NED	10
31	BNL3545	c02	117.49	0.88	0.84	FAM	11
32	BNL4071	c05	43.34	0.78	0.54	HEX	11
33	DPL0570	c11	5.93	0.70	0.48	NED	11
34	BNL3580[a]	c01	23.15	0.69	0.24	FAM	12
35	BNL2681	c21	133.71	0.61	0.00	HEX	12
36	BNL3594[a]	c06	12.27	0.75	0.61	NED	12
37	BNL3650	c06	43.66	0.48	0.00	FAM	13
38	BNL3474	c08	57.66	0.75	0.48	HEX	13
39	CIR347	c03	95.50	0.74	0.54	NED	13
40	BNL3992	c05	90.08	0.82	0.56	FAM	14
41	BNL3985	c23	14.76	0.71	0.67	HEX	14
42	DPL0717	c21	13.45	0.71	0.54	NED	14
43	BNL4017	c03	39.66	0.70	0.54	FAM	15
44	BNL4028	c09	48.57	0.68	0.35	HEX	15

Table 3 continued

Marker #	Marker name[a]	Chromosome #	Map position (cM)	CMD PIC	Upland PIC	Dye label	Triplex#
45	JESPR274	c23	104.97	0.86	0.67	NED	15
46	GH537	c25	80.06	0.64	0.00	FAM	16
47	CIR009	c01	125.90	0.60	0.00	HEX	16
48	JESPR220	c22	66.60	0.66	0.48	NED	16
49	BNL4061	c13	99.52	0.74	0.35	FAM	17
50	BNL4030	c19	198.94	0.73	0.38	HEX	17
51	JESPR065	c05	140.20	0.82	0.45	NED	17
52	CIR165[a]	c19	22.99	0.69	0.54	FAM	18
53	CIR105[a]	c15	16.31	0.78	0.35	HEX	18
54	BNL1227	c26	32.15	0.85	0.57	NED	18
55	CIR286	c23	140.12	0.61	0.43	FAM	19
56	CIR203[a]	c06	110.27	0.55	0.24	HEX	19
57	DPL0520	c25	0.00	0.59	0.00	NED	19
58	JESPR119	c06	130.99	0.39	0.00	FAM	20
59	CIR253	c05	163.98	0.80	0.48	HEX	20
60	DPL0541	c09	13.94	0.57	0.48	NED	20
61	CIR307	c15	75.21	0.66	0.29	FAM	21
62	CIR320	c07	22.24	0.53	0.00	HEX	21
63	NAU2277	c02	3.69	0.73	0.38	NED	21
64	DPL0196	c22	24.93	0.79	0.56	FAM	22
65	DPL0405	c14	45.66	0.76	0.35	HEX	22
66	TMB1640	c08	88.74	0.77	0.00	NED	22
67	DPL0600	c20	116.39	0.28	0.55	FAM	23
68	JESPR008	c11	124.01	0.56	0.00	HEX	23
69	DPL0094	c01	48.67	0.65	0.61	NED	23
70	DPL0618	c09	120.02	0.69	0.24	FAM	24
71	MGHES-029	c24	106.68	0.44	0.00	HEX	24
72	STV023	c07	70.79	0.73	0.00	NED	24
73	JESPR293	c14	82.61	0.56	0.48	FAM	25
74	DPL0071	c19	91.51	0.70	0.64	HEX	25
75	TMB1356	c10	56.05	0.77	0.00	NED	25
76	JESPR300	c12	120.31	0.72	0.00	FAM	26
77	DPL0249	c18	32.13	0.73	0.54	HEX	26
78	TMB2295	c18	104.84	0.82	0.48	NED	26
79	MGHES-076	c16	9.25	0.63	0.48	FAM	27
80	TMB0083	c26	73.01	0.44	0.00	HEX	27
81	TMB1648	c04	101.58	0.77	0.55	NED	27
82	MUSB1015	c11	90.44	0.50	0.00	FAM	28
83	JESPR291	c08	0.00	0.69	0.24	HEX	28
84	MGHES-022	c24	5.82	0.58	0.00	NED	28
85	NAU2140	c05	112.36	0.24	0.24	FAM	29
86	TMB0043	c21	59.72	0.75	0.00	HEX	29
87	TMB0514	c02	56.78	0.63	0.00	NED	29
88	TMB0382	c23	64.45	0.63	0.38	FAM	30

Table 3 continued

Marker #	Marker name[a]	Chromosome #	Map position (cM)	CMD PIC	Upland PIC	Dye label	Triplex#
89	TMB0836	c03	3.23	0.77	0.54	HEX	30
90	CIR202	c14	2.54	0.46	0.24	NED	30
91	TMB1421	c01	71.07	0.46	0.00	FAM	31
92	TMB2018	c17	46.52	0.63	0.00	HEX	31
93	GH048	c20	86.60	0.44	0.00	NED	31
94	TMB1427	c08	137.91	0.73	0.36	FAM	32
95	TMB2068	c16	100.19	0.54	0.00	HEX	32
96	BNL3371	c17	37.91	0.61	0.48	NED	32
97	TMB2955	c07	119.82	0.63	0.28	FAM	33
98	TMB0799	c12	7.46	0.70	0.48	HEX	33
99	TMB1910	c15	114.68	0.57	0.00	NED	33
100	GH107	c04	39.67	0.47	0.00	FAM	34
101	GH283	c10	6.55	0.71	0.35	HEX	34
102	CIR218	c22	74.06	0.78	0.00	NED	34
103	GH495	c09	78.04	0.48	0.00	FAM	35
104	GH052	c22	14.09	0.23	0.00	HEX	35
105	DPL0135	c20	66.49	0.67	0.54	NED	35
Average				0.65	0.29		

[a] Markers also used in the study of Lacape et al. (2007)

We compared our results with those obtained by Lacape et al. (2007). They used 184 SSR markers to analyze 47 *Gossypium* genotypes that included four tetraploid species. Of the 184 markers used in their study, 18 were also included in our core set (Table 3). In their study, the average PIC value of these 18 markers was 0.67 with the range from 0.34 (marker BNL2544) to 0.86 (JESPR153a). In our present research, the average PIC value of the same 18 markers was 0.66 with the range from 0.39 (marker BNL2544) to 0.82 (JESPR153). The results of these two studies were similar although Lacape et al. (2007) did not calculate the PIC values of each marker among *G. hirsutum* accessions.

Phylogenetic relationship of the CMD panel as revealed by the core SSR marker set

The assembly of the 12 genotype CMD panel has provided the cotton research community with a *Gossypium* germplasm reference (Yu 2004; Blenda et al. 2006). However, knowledge about the phylogenetic relationship of the 12 cotton genotypes is needed to better interpret research results with this genetic reference set. Using the 105 core SSR marker set, a matrix was generated to reveal the genetic similarity between any two members of the CMD panel (Supplemental Table 1). A dendrogram was constructed to display the genetic similarities within the CMD panel (Fig. 2). While all six *G. hirsutum* genotypes (AD_1) were clustered in one major clade, DPL 458BR and STV 4892BR were most similar (0.755) while Acala Maxxa and PM 1218BR were the most different (0.575) among the *G. hirsutum* cultivars. TM-1, the *G. hirsutum* genetic standard, is an inbred derived from DPL 14, a highly productive and widely adapted cultivar (Kohel et al. 1970). Figure 2 shows that TM-1 was located within proximity to the branch of DPL 458BR and STV 4892BR that had a TM-1 similarity coefficient of 0.683 and 0.724, respectively. Two *G. barbadense* cottons (AD_2), the genetic standard 3–79 and cultivar Pima S-6, were very closely related to each other with similarity coefficient of 0.855. Two diploid species representatives, *G. arboreum* A_2-8 and *G. raimondii* D_5-3, were the probable progenitors of the A and D subgenomes which formed the tetraploid cottons (Wendel and Cronn 2003). However, *G. raimondii* D_5-3 (similarity coefficients with the tetraploid cottons all under 0.1) exhibited a more distant relationship than *G. arboreum* A_2-8 (similarity

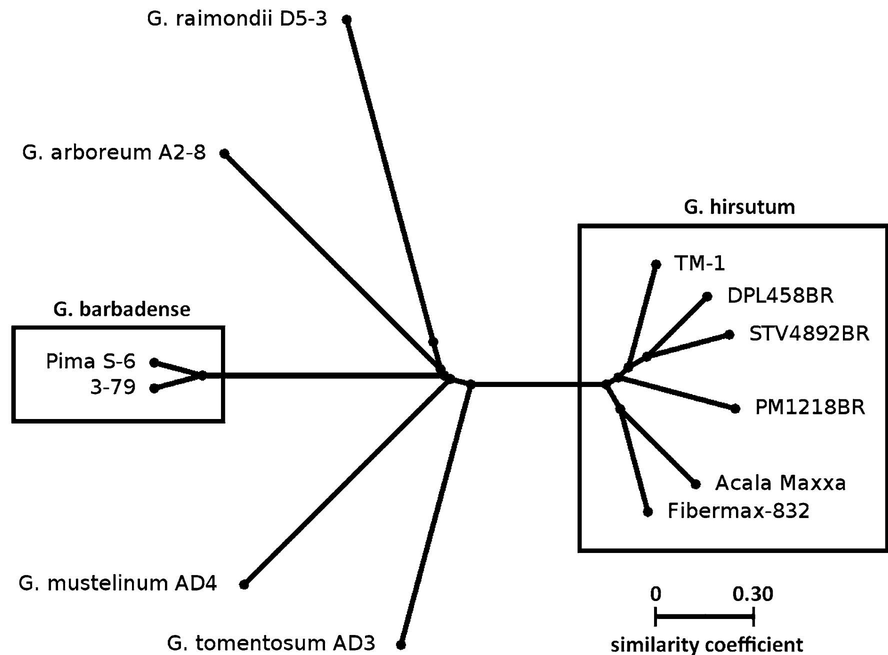

Fig. 2 Unrooted dendrogram for 12 *Gossypium* genotypes of the CMD panel constructed with the 105 core SSR markers using the UPGMA algorithm (Jordan and Piel 2008)

coefficients with the tetraploid cottons all above 0.1) to all other CMD panel members (Supplemental Table 1). Two wild tetraploid species, *G. tomentosum* (AD_3) and *G. mustelinum* (AD_4), also were clearly distinguishable in forming their own branches of the dendrogram. Results from this analysis reflect the original considerations of the CMD panel assembly and confirm the excellent value of the 105 core SSR marker set for use in cotton research.

Future implications of core DNA markers

The *Gossypium* genus comprises species with differing ploidy levels and presents a high degree of variability, from highly improved allotetraploid species to wild diploid forms. Research utilizing *Gossypium* germplasm is essential, as this is a complex genus. Cotton improvement programs, either public or private, use *Gossypium* germplasm in their programs with the ultimate goal of developing breeding lines or cultivars to increase lint yield and fiber quality among other traits. The variability in *Gossypium* germplasm has only recently begun to be tapped as a source of beneficial characteristics because there is still a shortage of effective genomic tools compared with other major crops. This report represents a major effort to improve the accuracy and efficiency of molecular characterization of *Gossypium* germplasm. This core SSR marker set, representing the efforts of many research groups, is portable, and thus is readily available to any researchers who wish to use it in their own research facilities.

Although additional portable markers are needed for detailed characterization, all evaluations of cotton germplasm across different gene pools and germplasm sources can be shared and pooled into a single database for systematic analysis with the 105 core SSR marker set presented in this report. Cotton germplasm accessions with unique DNA profiles can be investigated in detail with additional DNA markers for mining new genes of interest. Future augmentation of this core SSR set can be made with newly developed single nucleotide polymorphism (SNP) markers upon appropriate characterization and mapping in the cotton genomes (Van Deynze et al. 2009). The development of an appropriate set of portable core DNA markers that are highly informative, evenly distributed in the genome, able to generate unambiguous PCR products, and able to be effectively multiplexed offers a previously unavailable opportunity to integrate molecular characterization of the U.S. cotton germplasm collection and other *Gossypium* collections around the

world into a single cohesive international effort to assess *Gossypium* diversity.

Acknowledgments We thank members of the cotton research community for valuable comments. We also thank Ryan McCormick for technical assistance.

Disclaimer Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U. S. Department of Agriculture that is an equal opportunity provider and employer.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov IB, Buriev ZT, Saha S, Scheffler BE, Jenkins JN, Abdukarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 92(6): 478–487. doi:10.1016/j.ygeno.2008.07.013

Abdurakhmonov IY, Saha S, Jenkins J, Buriev Z, Shermatov S, Scheffler B, Pepper A, Yu J, Kohel R, Abdukarimov A (2009) Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. Genetica 136(3):401–417. doi:10.1007/s10709-008-9337-8

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape JM, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu SL, Zhang TZ, Fang DQ, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006) CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics 7:132–141. doi:13210.1186/1471-2164-7-132

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3): 314–331

Campbell BT, Saha S, Percy R, Frelichowski J, Jenkins JN, Park W, Mayee CD, Gotmare V, Dessauw D, Giband M, Du X, Jia Y, Constable G, Dillon S, Abdurakhmonov IY, Abdukarimov A, Rizaeva SM, Abdullaev A, Barroso PAV, Pádua JG, Hoffmann LV, Podolnaya L (2010) Status of the global cotton germplasm resources. Crop Sci 50(4): 1161–1179. doi:10.2135/cropsci2009.09.0551

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. In: Caspari EW, John GS (eds) Advances in genetics, vol 23. Academic Press, New York, pp 271–375

Esbroeck G, Bowman DT (1998) Cotton germplasm diversity and its importance to cultivar development. J Cotton Sci 2(3):121–129

Fang D, Xiao J, Canci P, Cantrell R (2010) A new SNP haplotype associated with blue disease resistance gene in cotton (*Gossypium hirsutum* L.). Theor Appl Genet 120(5): 943–953. doi:10.1007/s00122-009-1223-y

Garcia-Vallvé S, Palau J, Romeu A (1999) Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. Mol Biol Evol 16(9):1125–1134

Gill KS, Gill BS, Endo TR (1993) A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma 102(6):374–381. doi:10.1007/bf00360401

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(1):527–541. doi:10.1534/genetics.107.070375

Jordan GE, Piel WH (2008) PhyloWidget: web-based visualizations for the tree of life. Bioinformatics 24(14): 1641–1642. doi:10.1093/bioinformatics/btn235

Kohel RJ, Richmond TR, Lewis CF (1970) Texas marker-1. Description of a genetic standard for *Gossypium hirsutum*. Crop Sci 10:670–671

Kohel RJ, Yu JZ (2002) Molecular characterization of *Gossypium* germplasm for cotton improvement. In: Engels JMM, Rao VR, Brown AHD, Jackson MT (eds) Managing plant genetic resources. CAB International Publishing, Wallingford, pp 67–75

Kohel RJ, Yu JZ, Park Y-H, Lazo G (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121(2):163–172. doi:10.1023/a:1012263413418

Lacape JM, Dessauw D, Rajab M, Noyer JL, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. Mol Breed 19(1):45–58. doi:10.1007/s11032-006-9042-1

Lacape JM, Jacobs J, Arioli T, Derijcker R, Forestier-Chiron N, Llewellyn D, Jean J, Thomas E, Viot C (2009) A new interspecific, *Gossypium hirsutum* × *G. barbadense*, RIL population: towards a unified consensus linkage map of tetraploid cotton. Theor Appl Genet 119(2):281–292. doi:10.1007/s00122-009-1037-y

Lee JA (1984) Cotton as a world crop. In: Kohel RJ, Lewis CF (eds) Cotton, agronomy monograph 24. American Society of Agronomy, Madison, pp 1–25

Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21(9): 2128–2129

Liu D, Guo X, Lin Z, Nie Y, Zhang X (2006) Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. Genet Res Crop Evol 53(6):1145–1152. doi:10.1007/s10722-005-1304-y

Percival AE, Wendel JF, Stewart JM (1999) Taxonomy and germplasm resources. In: Smith CW, Cothren JT (eds) Cotton: origin, history, technology, and production. Wiley, New York, pp 33–63

Stelly DM, Saha S, Raska DA, Jenkins JN, McCarty JC, Gutierrez OA (2005) Registration of 17 upland (*Gossypium hirsutum*) cotton germplasm lines disomic for different *G.*

barbadense chromosome or arm substitutions. Crop Sci 45(6):2663–2665
Stewart JM (1994) Potential for crop improvement with exotic germplasm and genetic engineering. Paper presented at the Challenging the Future: Proceedings of the World Cotton Research Conference-1. Brisbane, Australia, 14–17 February
Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277(5329):1063–1066
Van Deynze A, Stoffel K, Lee M, Wilkins TA, Kozik A, Cantrell RG, Yu JZ, Kohel RJ, Stelly DM (2009) Sampling nucleotide diversity in cotton. BMC Plant Biol 9:125. doi:12510.1186/1471-2229-9-125
Wallace T, Bowman D, Campbell B, Chee P, Gutierrez O, Kohel R, McCarty J, Myers G, Percy R, Robinson F, Smith W, Stelly D, Stewart J, Thaxton P, Ulloa M, Weaver D (2009) Status of the USA cotton germplasm collection and crop vulnerability. Genet Res Crop Evol 56(4):507–532. doi:10.1007/s10722-008-9382-2
Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. In: Advances in agronomy, vol 78. Academic Press, San Diego, pp 139–186
Xu ZY, Kohel RJ, Song GL, Cho JM, Alabady M, Yu J, Koo P, Chu J, Yu SX, Wilkins TA, Zhu YX, Yu JZ (2008) Gene-rich islands for fiber development in the cotton genome. Genomics 92(3):173–183. doi:10.1016/j.ygeno.2008.05.010
Yu JZ (2004) A standard panel of *Gossypium* genotypes established for systematic characterization of cotton microsatellite markers. An electronic newsletter of applied plant breeding sponsored Food and Agriculture Organization of the United Nations. Plant Breed News Ed 148(107):148
Yu JZ, Kohel RJ (2001) Cotton genome research in the United States. In: Jenkins JN, Saha S (eds) Genetic improvement of cotton: emerging technologies. The Science Publishers Inc., Polymouth, pp 103–121
Yu JZ, Kohel RJ, Fang DD, Cho J, Van Deynze A, Ulloa M, Hoffman SM, Pepper AE, Stelly DM, Jenkins JN, Saha S, Kumpatla SP, Shah MR, Hugie WV, Percy RG (2012) A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. Genes Genomes Genet 2:43–58
Yu Y, Yuan DJ, Liang SG, Li XM, Wang XQ, Lin ZX, Zhang XL (2011) Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC1 population between *Gossypium hirsutum* and *G. barbadense*. BMC Genomics 12:15. doi:1510.1186/1471-2164-12-15