TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2018) 42: 172-182

doi:10.3906/bot-1706-18

Genetic diversity and phylogenetic relationships of a potential cotton collection for European breeding research

Elena PELÁEZ-ANDÉRICA[1,*], Felipe REY[1], Manuel LÓPEZ[2], Juan GIL[3]
[1]Algodonera del Sur S.A., Los Palacios y Villafranca, Seville, Spain
[2]Agricultural Production Area, IFAPA Las Torres-Tomejil Center, Alcalá del Río, Seville, Spain
[3]Department of Genetics, University of Córdoba, Campus de Rabanales, Cordoba, Spain

Received: 08.06.2017 • **Accepted/Published Online:** 17.10.2017 • **Final Version:** 20.03.2018

Abstract: An understanding of the genetic diversity of any plant collection is essential for initiating future plant breeding strategies. Throughout the cotton regions in Europe (mainly Greece and Spain), there is an expanding need for new yielding varieties with shorter crop cycles and high quality fiber. Although cotton is decisive to many rural economies, it is surprising that there is a lack of public information available regarding genetic diversity of European cotton varieties. Thus, we began an examination of variability heritage by evaluating a subset of 48 commercial and experimental cotton varieties (36 *Gossypium hirsutum*, 10 *G. barbadense*, 1 *G. arboreum*, and 1 unknown species) using 67 microsatellite markers (SSRs). Sixty-two SSRs (92.54%) showed polymorphism and revealed 464 alleles. The total polymorphism information content value was 0.7 (0.57 in *G. hirsutum* and 0.54 in *G. barbadense*), which indicates a high informative content of our SSR set compared to that obtained in previous cotton studies and plant collections. Cluster analysis was performed to analyze diversity results and phylogenetic relationships between and within species. These results established a foundation for the future development of improved commercial cotton varieties and can thereby help the European cotton industry meet pressing needs.

Key words: *Gossypium* sp., European, polymorphism, SSR marker, cluster analysis

1. Introduction

Cotton (genus *Gossypium*, Malvaceae) is a major natural fiber crop used by the textile industry and is an important oil crop, as well as a byproduct provider (seeds for cattle feed, short fibers to obtain cellulose for chemistry, etc.). Although the EU represents only 1.2% of the world's cotton production with 283,400 metric tons of raw cotton (Cotton Inc., 2017), the crop plays a determinant role in economic and social aspects of countries like Greece and Spain. The two countries account for 81.4% and 17.3% of the total European output, respectively (Eurostat, 2016). These regions provide cotton plants with high boll yields and high fiber quality due to the specialized, professional, and intensive farming practices being implemented. Nevertheless, there is a scarcity of research on European cotton, especially related to genetic diversity and molecular marker analysis. An enhanced understanding of both of these topics would be highly beneficial for developing new breeding strategies. The success of plant breeding programs mainly depends on the available genetic variability of crops. This variance depends on the genetic relationship and diversity between and within plant groups (Yu et al., 2012a; Abdellatif and Soliman, 2013). Thus, having a set of varieties that globally represent useful alleles is critical in any breeding program. It requires the management, conservation, and evaluation of valuable cotton materials while maintaining a continued supply of new germplasm. This plant breeding set could be made of new or old improved varieties with yield potential to our crop environment and other related wild materials or different species with desirable traits (Kaur et al., 2014). Nowadays, more than 95% of world cotton production comes from American cotton species *Gossypium hirsutum* L. and *Gossypium barbadense* L. (Upland and Pima cotton, respectively) (Abdurakhmonov et al., 2012). Upland varieties, which represent 90% of the total, are appreciated for their high yield and wide adaptability. Conversely, Pima varieties are cultivated for their fiber length and quality. Both New World species are allotetraploidic (AD) with 2n = 4x = 52 chromosomes (Fryxell et al., 1992) and can be cross-pollinated for commercial purposes in order to interchange traits via backcrossing or segregating

* Correspondence: elepeland@hotmail.com

populations. Alternatively, they can be used to produce many intra- and interspecific F1 hybrid varieties (Boopathi et al., 2011). Since 2000, and mainly in India, approximately 20% of the total global acreage has been grown with cotton hybrids (Dong et al., 2006; Nacoulima et al., 2016). Another way to improve genetic variability of tetraploid cotton is the use of *Gossypium* diploid species (2n = 2x= 26) (Mergeai, 2006) or exotic tetraploid germplasms (McCarty and Percy, 2001). Notably, wild species related to most diploid genomes are currently used as important reservoirs for alleles of interest and can be used to enhance fiber fineness (e.g., Nacoulima et al., 2016).

The use of DNA-based molecular markers represents a powerful tool for any breeding objective, including characterization of germplasm collections, estimation of genetic diversity, acceleration of the selection process, varietal identification, seed certification, and plant breeder rights implementation (Yu et al., 2012a). Microsatellite markers (SSRs) are considered the marker system of choice for the majority of applications in the plant breeding field because of their high polymorphism level, PCR-based performance, multiallelic nature, high reproducibility, codominant inheritance, relative abundance, and uniform dispersion in the plant genome (Kantartzi et al., 2013). Public marker databases are available with 18,796 SSRs; their corresponding primer sequences are developed from genomic, EST, or BAC libraries (www.cottongen.org). In the past few years, SSR and ISSR markers have been widely used to characterize many cotton collections (Sapkal et al., 2011; Ullah et al., 2012; Abdellatif and Soliman, 2013; Tyagi et al., 2014; Zhao et al., 2015; Moiana et al., 2015; Poortavakoli et al., 2017). However, there is only one published study using SSRs for breeding in Europe that is based on the identification of quantitative trait loci for fiber quality in a Bulgarian cotton breeding collection (Ivanova and Bojinov, 2009).

Genetic variability remains unknown in European commercial cotton varieties for breeding purposes, as there are no diversity or phylogeny studies. Such information would hold immense value for countries like Spain where the need for adapted and improved cotton varieties is increasing. In the 2016/2017 season, the entirety of the Spanish cotton cultivation area (60,781 ha) was located in the valley of the Guadalquivir River (Andalusia, southern Spain). Over the last 100 years, cotton has brought great wealth to this region due to the generation of many daily jobs and the high prices of raw cotton. Even today, it remains one of the few crops that can support rural populations and preserve small farming economies (those who own less than 10 ha), which represents 82% of the total number of Spanish holdings (COAG, 2012). In addition to its rural society assets, cotton is fairly unique in its ability to tolerate the extreme summer temperatures (more than 40 °C) and moderate salt content soil found in the Andalusia region. Nevertheless, since changes to the European Common Agricultural Policy came into effect in 2006 (Council Regulation (EC) No. 1782/2003 of 29 September 2003), the cotton sector has been losing its profitability with these new environmental regulations and higher requirements of fiber quality. Hence, there is an increasing demand for suitable cotton varieties with shorter cycles, high fiber quality, high yields, improved agronomic adaptation, and resistance to *Verticillium*, which is the main cotton fungal disease found in European growing regions like Spain, and especially among early maturing cotton (Erdogan et al., 2013).

The purpose of this study was to evaluate genetic diversity and phylogenetic relationships of a potential cotton collection for European breeding efforts (mainly *G. hirsutum* and *G. barbadense* varieties). All generated information will contribute to facilitating the initiation of future breeding programs inside European cotton regions or in areas of similar climatic conditions.

2. Materials and methods

2.1. Plant materials and DNA extraction

A total of 48 cotton cultivars (inbred lines) were selected by Algodonera del Sur S.A. and the Andalusian Institute of Agricultural Research and Training (Spanish acronym: IFAPA) (Table 1). Thirty-nine commercial and experimental plant materials were selected for their potential adaptation to the Andalusian field and climate conditions or for being donors of desired phenotypic traits for breeding. The remaining 9 cultivars were obtained from the US National Plant Germplasm System (USDA-ARS) as well-known genetic variability sources, 6 of which have previously been used as genetic standards for the Cotton Marker Database (CMD) and other cotton genomic studies (Yu, 2004; Blenda et al., 2006; Lacape et al., 2007; Yu et al., 2012a).

For DNA extraction, 100 mg of fresh young leaves was taken from a pool of 5 plants per cultivar to minimize the genomic influence of possible off-types in commercial cultivar samples with less control of progeny. The total genomic DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method following Paterson et al. (1993). Modification to the traditional protocol involved adding 1 µL of RNAse once DNA was extracted, diluting in TE buffer, and incubating for less than 30 min at 37 °C to avoid DNA degradation caused by the high content of polyphenolic compounds in cotton tissues. DNA quality was then checked by electrophoresis on 0.8% (w/v) agarose gel and DNA quantity was checked using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) at 260/280 nm. Subsequently, DNA samples were diluted to 25 ng/mL and stored at 20 °C until used.

Table 1. The information on 48 cotton cultivars used in this study.

Name	Species	Genome	Status[a]	Origin[b]	Donor[c]	Year[d]
ALEPO	*G. barbadense*	$(AD)_2$	C	Caucasus	Hazera	2016
ARMADA	*G. barbadense*	$(AD)_2$	C	Caucasus	Spirou	2013
CIMA	*G. barbadense*	$(AD)_2$	E	Caucasus	Hazera	2006
E1	*G. barbadense*	$(AD)_2$	C	USA, CA	Israel S	2014
GW-4265	*G. barbadense*	$(AD)_2$	E	Caucasus	Spirou	2010
GW-4269	*G. barbadense*	$(AD)_2$	E	USA	Spirou	2014
LAGIRALDA	*G. barbadense*	$(AD)_2$	C	Caucasus	Spirou	2016
LANOVIA	*G. barbadense*	$(AD)_2$	C	Caucasus	Spirou	2016
SAN DIEGO	*G. barbadense*	$(AD)_2$	Cc	USA, CA	-	2010
AI-270	*G. hirsutum*	$(AD)_1$	E	Israel	Hazera	1999
AI-292	*G. hirsutum*	$(AD)_1$	E	Israel	Hazera	1999
ALEXANDROS	*G. hirsutum*	$(AD)_1$	C	Greece	Mons	2002
AMAZONA	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2012
AVANGARD 264	*G. hirsutum*	$(AD)_1$	C	Bulgaria	Chirpan	2001
BABYLON	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2002
CAMPO	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2001
CARISMA	*G. hirsutum*	$(AD)_1$	C	USA	ProGen	2013
CELIA	*G. hirsutum*	$(AD)_1$	C	Australia	CSIRO	2001
COKER 312	*G. hirsutum*	$(AD)_1$	C	USA, TX	USDA	1974
DARMI	*G. hirsutum*	$(AD)_1$	C	Bulgaria	Chirpan	2010
DP ACALA 90	*G. hirsutum*	$(AD)_1$	C	USA	Mons	1992
DP-332	*G. hirsutum*	$(AD)_1$	C	USA, MS	Mons	2010
DP-377	*G. hirsutum*	$(AD)_1$	C	USA, MS	Mons	2010
ELPIDA	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2015
ELSA	*G. hirsutum*	$(AD)_1$	C	Australia	CSIRO	2009
FANTOM	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	1999
JULIA	*G. hirsutum*	$(AD)_1$	C	Australia	CSIRO	2005
JUNCAL	*G. hirsutum*	$(AD)_1$	C	USA	Mons	2004
KOLORIT	*G. hirsutum*	$(AD)_1$	C	Bulgaria	Chirpan	2010
LAMBADA	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2007
LIDER	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2001
MASSALA	*G. hirsutum*	$(AD)_1$	Cc	Israel	-	2014
MISTRAL	*G. hirsutum*	$(AD)_1$	C	Greece	Spirou	2012
NATALIYA	*G. hirsutum*	$(AD)_1$	C	Bulgaria	Chirpan	2008
RUMI	*G. hirsutum*	$(AD)_1$	C	Bulgaria	Chirpan	2011
SP-57	*G. hirsutum*	$(AD)_1$	E	Greece	Spirou	2013
SS-UPL-01	*G. hirsutum*	$(AD)_1$	E	Spain	SSI	2002
ST-467	*G. hirsutum*	$(AD)_1$	E	USA, MS	Stonev	2001
VIKY	*G. hirsutum*	$(AD)_1$	C	Spain	Eusem	2005
180F	*G. hirsutum*	$(AD)_1$	BM	Uzbekistan	USDA	1940
ACALA 3080	*G. hirsutum*	$(AD)_1$	BM	USA, TX	USDA	1989
TEX 1425	*G. hirsutum*	$(AD)_1$	WM	Bulgaria	USDA	1989

Table 1. (Continued).

WC-19NSSL	*G. hirsutum*	$(AD)_1$	BM	USA, AZ	USDA	1991
TM-1 (CMD1)	*G. hirsutum*	$(AD)_1$	BM	USA, TX	USDA	1956
ACALA MAXXA (CMD3)	*G. hirsutum*	$(AD)_1$	Cu	USA, CA	USDA	1990
PIMA S-6 (CMD8)	*G. barbadense*	$(AD)_2$	Cu	USA, AZ	USDA	1983
A2-0008 (CMD9)	*G. arboreum*	A_2	BM	USA, TX	USDA	1989
KIBALA	unknown	-	WM	Angola	-	2014

[a] C — commercial, E — experimental, Cc — commercial collecting of interesting off types, BM — breeding material, WM — wild material, Cu — cultivar.
[b] Breeding origin (varieties with traits adapted or domesticated to that climatic region conditions).
[c] Hazera — Hazera Seeds Ltd., Shikmim, Israel; Spirou — House of Agriculture Spirou AEBE, Athens, Greece; Israel S — Israel Seeds Ltd., Shefaim, Israel; CSIRO — Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia; Mons — Monsanto Company, St. Louis, MO, USA; SSI — Semillas Sostenibles Ibérica S.L.U., Seville, Spain; Chirpan — Institute of Agriculture Crops, Chirpan, Bulgaria; Stonev — Stoneville Pedigreed Seed Company, Stoneville, MS, USA; Eusem — Eurosemillas, Cordoba, Spain; USDA — Agricultural Research Service of the United States Department of Agriculture (USDA-ARS); ProGen — ProGen Seeds S.A., Antalya, Turkey.
[d] Year of admission to European plant variety catalogues (C) / year of experimentation testing (E) / year of own collection or donation (Cc, Cu, BM, WM).

2.2. Genotyping by SSR markers

A total of 67 pairs of primers were selected to represent 6 diverse SSR sources (BNL, CIR, JESPR, MGHES, NAU, and TMB) developed by different research groups of the cotton community (Blenda et al., 2006) (Table S1). Markers were chosen to be well distributed among the 26 chromosomes of allotetraploidic cotton based on previous genetic linkage mapping studies (chiefly from Guo et al., 2008; Yu et al., 2012b) (Table S1). Each chromosome was represented by at least 1 or 2 SSR loci in different chromosomal arms, and those with the longest chromosomes were represented by 4 or 5 SSR loci (Yu et al., 2012b). Other criteria to equilibrate the marker distribution included single-copy versus multiple-copy markers and markers that amplified in both A- and D-subgenomes. In addition, most of the selected markers were reported as polymorphic in previous diversity studies (Bertini et al., 2006; Abdurakmonov et al., 2007; Lacape et al., 2007; Shen et al., 2011; Yu et al., 2012a; Mishra and Fougat, 2013; Cai et al., 2014; Elçi et al., 2014). The sequence of individual primer pairs can be found in the CMD (www.cottongen.org). They were custom synthesized by Integrated DNA Technologies Inc. in Coralville, IA, USA.

Polymerase chain reactions (PCRs) were carried out in a volume of 10 µL containing 25 ng of DNA, 2 µL of 5X GoTaq reaction buffer, 2.5 mM $MgCl_2$, 0.3 mM dNTP, 0.3 mM each forward and reverse SSR primer, and 0.25 U of GoTaq G2 (Promega, Madison, WI, USA). The PCR conditions were as follows: 3 min at 94 °C, followed by 40 cycles of 30 s at 94 °C, 30 s at 55 °C, and 30 s at 72 °C and then 10 min at 72 °C for final extension. $MgCl_2$ concentration and annealing temperature were adjusted in some SSR reactions (Table S1). Forward primers were then labeled at the 5' end with 6-FAM, HEX, or NED fluorescent dyes, multiplexing the reactions by 21 triplex PCR bin sets. Amplified DNA products were separated using capillary electrophoresis on an automated ABI Prism 3130XL genetic analyzer (Applied Biosystems/Life Technology, Foster City, CA, USA) using GeneScan-400 Rox as an internal DNA standard size.

2.3. Genetic diversity, phylogeny, and ploidy analysis

These SSR amplification products were analyzed by Peak Studio v.2.2 software (Bioinformatics Solutions Inc., Waterloo, ON, Canada). Using a GeneCall threshold of 0.45, present peaks or supposed alleles were scored as 1 (present) or 0 (absent) in each cultivar to generate a binary matrix. Based on Jaccard algorithms, a genetic distance matrix was computed by running the NTSYSpc 2.1 software using the unweighted pair group method with arithmetic means (UPGMA) to construct the genetic dendrogram. The PIC values of each SSR marker were used as indicators of existing diversity and were calculated as $PIC = 1 - \Sigma p_j^2$, where p_j is the frequency of the jth alleles in the examined cultivars (Anderson et al., 1993).

We also implemented flow cytometer analysis for an unknown accession that had an unknown ploidy level and botanical origin. Following Moreno et al. (2006), samples of tetraploid asparagus landrace Morado de Huetor were employed as the internal standard and then compared with cultivar (cv.) Amazona (*G. hirsutum*, tetraploid) and our unknown accession. A Partec PA flow cytometer was employed (Sysmex Partec GmbH, Münster, Germany) with its corresponding nuclei extraction and staining buffers (CyStain UV Precise P, Partec).

3. Results

3.1. Polymorphism information content (PIC) values

Among the 67 SSRs chosen, 62 primer pairs (92.54%) were amplified successfully and showed polymorphisms between all 48 cultivars, confirming the ability of most of our markers to reveal genetic diversity within this plant collection (Table 2). Based on the results, 464 allele loci were detected with an average of 2.38 markers per chromosome and 7.48 alleles per SSR loci (ranging from 2 to 17). These 62 SSRs were represented in all 26 chromosomes, between 1 SSR (c.12 and c.22) and 6 SSRs (c.19 and c.24) (Table S1).

The total average PIC value of markers for all 48 accessions including both *G. hirsutum* and G. *barbadense* was 0.70, ranging from 0.3 (NAU2649) to 0.89 (JESPR0152) (Table 2). Thirty-six out of the 62 SSR marker sets were potentially informative (0.7 < PIC < 0.9), 19 SSRs were highly informative (0.5 < PIC < 0.7), and 7 SSR markers were moderately informative (0.25 < PIC < 0.50). Within *G. hirsutum*, the mean PIC value was 0.57 and ranged from 0 (NAU2649 and BNL3308) to 0.88 (BNL3255). In the *G. barbadense* group, the average PIC value was 0.54, ranging from 0 (BNL3563, BNL3580, BNL2544, BNL2495, NAU1322, and BNL3590) to 0.83 (BNL3255). Most markers were more informative among *G. hirsutum* varieties, with the exception of 7 markers (NAU2649, NAU3770, JESPR0127, BNL3594, CIR0203, MGHES0031, TMB0366) that were more informative in *G. barbadense* varieties (PIC < 0.5 in *G. hirsutum*). In general, more than half of the employed SSR markers showed high PIC values in both species. For example, markers BNL3255 on c.08, BNL3545 on c.02/c.14, and BNL1521 on c.24 had very high PIC values not only among *G. hirsutum* (0.88, 0.85, and 0.83, respectively) but also within the *G. barbadense* varieties (0.83, 0.75, and 0.72, respectively). These highly informative SSR markers amplified between 11 and 15 alleles in our plant collection.

3.2. Phylogenetic relationships clusters

The UPGMA analysis based on the genetic distance matrix (Table S2) clustered the 48 genotypes (36 *G. hirsutum*, 10 *G. barbadense*, 1 *G. arboreum*, and 1 unknown) into three main groups (A, B, and H) that mainly comprise our 3 different species analyzed (Figure): Group A represented the diploid *G. arboreum* species (A2-0008). It was distinctly separated from the remaining varieties, where the highest genetic distances were found with *G. barbadense* cv. Cima (0.92) and *G. hirsutum* cv. Massala (0.9). Group B consisted of all 10 *G. barbadense* varieties and the unknown accession from Angola, which is separated from the 10 *G. barbadense* with a mean genetic distance of 0.56. In parallel to cluster analysis, flow cytometry analysis of this unknown genotype showed around 30% less fluorescence intensity for the G0/G1 peak than for the allotetraploid Amazona (*G. hirsutum*), which indicated different and lower DNA content. The varieties of *G. barbadense* were divided into two subgroups: B-I, which corresponded to Caucasian-origin breeding varieties, and B-II, which was mainly composed of breeding varieties with USA origin. In the B-I group, it is interesting to note the lowest genetic distance of 0.12 between cv. Armada and cv. Lanovia. Cv. Armada and cv. Lanovia have very similar bush structure and leaf color and shape, although cv. Lanovia shows higher precocity and lower height. Only 5 SSRs showed polymorphism between them (BNL0569, BNL1672, BNL2570, JESPR0152, TMB0471).

Finally, group H was composed of 36 *G. hirsutum* varieties and cv. 180-F, which clustered alone in the lower part of the group, completely separated from the rest of the *G. hirsutum* varieties. The remaining *G. hirsutum* accessions could not clearly be separated in other subgroups, although it has been possible to differentiate all of them with at least 5 polymorphic markers from the initial 62 SSR markers. Some *G. hirsutum* accessions from the same breeding origin clustered together, like Australian varieties (cv. Celia to cv. Julia) and Bulgarian varieties (cv. Avangard 264 to cv. TEX-1425). Our average genetic distance for the *G. hirsutum* group was 0.43, ranging from a highest value of 0.59 between cv. Fantom and cv. AI-270 or cv. TM-1 to the lowest value of 0.22 between cv. DP Acala 90 and cv. AI-292.

4. Discussion

The molecular methods used in this genetic analysis provided us with a better understanding of the genetic diversity of this cotton collection, chosen for its diverse origins and phenotypic potential for future European breeding purposes.

A high level of amplification success (92.54%) was obtained with our initial selection of 67 markers, probably due to the selection of previous polymorphic markers. Comparing our PIC value results with previous studies, Yu et al. (2012a) obtained an average PIC value of 0.713 for 25 SSRs in 12 genotypes belonging to six *Gossypium* species (known collectively as the CMD panel). In this study, a similar PIC value (0.69) was obtained for these same SSRs in our 48 accessions (4 *Gossypium* species). Similar results were obtained when compared with another 25 SSRs used previously by Lacape et al. (2007) in 4 tetraploid species. They reported an average PIC value of 0.712, and in our study it had a value of 0.70. Therefore, in terms of different species diversity, our results were similar to previous studies.

However, our *G. hirsutum* mean PIC value (0.57) was very high compared with other diversity studies with more varieties and markers used, or with lower PIC values. For example, Abdurakhmonov et al. (2008) obtained an average PIC value of 0.122 (287 accessions and 95 SSRs),

Table 2. Polymorphism information content (PIC) values, number of alleles found in all 48 accessions, and allele size range in *G. hirsutum* and *G. barbadense* accessions obtained from 62 SSR (microsatellite) markers.

	48 varieties		*Gossypium hirsutum*			*Gossypium barbadense*		
Marker[abc]	PIC	Total alleles	PIC	No. alleles	Range	PIC	No. alleles	Range
BNL0387[b]	0.64	7	0.5	3	222–309	0.49	3	237–241
BNL0530[ab]	0.72	6	0.58	5	155–197	0.59	4	153–197
BNL0569[b]	0.73	8	0.54	4	145–161	0.64	4	149–177
BNL1122[a]	0.74	6	0.74	4	182–192	0.5	2	182/186
BNL1227[b]	0.81	14	0.76	12	136–202	0.52	3	176–188
BNL1423[c]	0.64	5	0.54	3	134–146	0.41	2	134/224
BNL1521[abc]	0.86	15	0.83	13	104–207	0.72	4	107–167
BNL1531[b]	0.8	11	0.76	6	134–155	0.79	6	134–170
BNL1551[bc]	0.82	10	0.74	6	188–264	0.67	3	188–260
BNL1672[c]	0.83	10	0.74	7	103–166	0.64	4	99–127
BNL1897[bc]	0.6	5	0.5	2	142/158	0.55	3	142–152
BNL2495[bc]	0.65	4	0.52	3	202–212	0	1	202
BNL2544[bc]	0.43	4	0.12	3	93–233	0	1	221
BNL2570[bc]	0.84	11	0.79	7	239–296	0.68	5	247–296
BNL2572[bc]	0.38	3	0.1	3	252–266	0.18	2	252/266
BNL2921[c]	0.81	7	0.73	4	158–173	0.5	2	162/175
BNL3090[bc]	0.8	8	0.73	5	227–285	0.54	3	242–305
BNL3255[a]	0.89	12	0.88	10	223–391	0.83	6	223–391
BNL3257[c]	0.84	14	0.75	8	147–225	0.75	6	147–207
BNL3308	0.44	4	0	1	212	0.18	2	212/214
BNL3408[c]	0.82	10	0.72	5	134–187	0.76	5	134–187
BNL3474[ab]	0.67	5	0.51	3	188–205	0.69	4	188–207
BNL3545[b]	0.86	11	0.85	8	131–233	0.75	5	121–233
BNL3563[c]	0.45	4	0.07	2	223/243	0	1	259
BNL3580[bc]	0.42	3	0.1	3	215–229	0	1	223
BNL3590	0.75	8	0.68	6	180–205	0	1	191
BNL3594[bc]	0.72	5	0.36	2	198/213	0.63	3	196–215
BNL3992[b]	0.8	7	0.73	7	135–170	0.55	3	135–164
CIR0105[bc]	0.85	11	0.78	9	94–256	0.7	4	108–242
CIR0165[bc]	0.68	5	0.63	4	224–243	0.55	3	224–245
CIR0169[bc]	0.38	2	0.06	2	152/154	0.2	2	152/154
CIR0203[bc]	0.55	8	0.39	4	268–278	0.66	3	262–274
CIR0246[c]	0.83	17	0.74	9	136–258	0.74	7	142–231
CIR0413[c]	0.84	17	0.81	13	95–225	0.57	4	100/130
JESPR0056	0.79	9	0.72	5	136–206	0.75	5	136–200
JESPR0127[a]	0.69	6	0.32	3	214–233	0.65	3	200/236
JESPR0152[c]	0.89	12	0.81	7	224–283	0.74	5	195–245
JESPR0153[abc]	0.79	11	0.66	8	113–232	0.59	4	113–164
JESPR0208[c]	0.69	6	0.54	3	105–128	0.75	5	94–132

Table 2. (Continued).

JESPR0220[b]	0.69	7	0.63	3	145–178	0.67	4	145–200
JESPR0274[b]	0.85	13	0.77	6	118–245	0.69	4	122–194
MGHES0031	0.52	4	0.43	4	216–228	0.5	2	216/222
MGHES0055	0.71	4	0.67	3	216–236	0.75	4	216–241
NAU0905[a]	0.76	9	0.59	3	171–183	0.65	3	168–195
NAU0934[a]	0.77	7	0.7	5	211–231	0.61	3	211–223
NAU0943[a]	0.57	5	0.21	2	195/207	0.17	2	164/176
NAU0998[a]	0.79	6	0.67	4	173–185	0.63	4	173–195
NAU1042[a]	0.72	5	0.67	5	236–260	0.61	3	236–260
NAU1043[a]	0.79	7	0.63	4	240–254	0.6	3	167–257
NAU1070[a]	0.68	5	0.64	5	168–194	0.5	2	178/194
NAU1162[a]	0.66	4	0.62	3	198–210	0.5	2	198/201
NAU1167[a]	0.79	7	0.79	7	201–224	0.64	3	201–217
NAU1200[a]	0.76	6	0.7	4	220–249	0.48	2	204/220
NAU1221[a]	0.79	8	0.74	4	115–261	0.79	7	115–261
NAU1322[a]	0.68	4	0.65	4	189–213	0	1	189
NAU2508[a]	0.55	4	0.54	3	161–174	0.5	2	161/174
NAU2649	0.3	2	0	1	202	0.5	2	202/207
NAU3341	0.66	5	0.51	3	260–270	0.55	3	266–272
NAU3770	0.61	6	0.1	3	181–191	0.71	5	179–193
NAU6634	0.78	6	0.66	4	158–190	0.65	4	167–199
TMB0366	0.64	5	0.48	3	220–235	0.58	3	194–231
TMB0471	0.87	14	0.77	9	212–276	0.73	5	183–221
Average	0.70	7.48	0.57			0.54		

[a] Markers also used in the study of Cai et al. (2014).
[b] Markers also used in the study of Yu et al. (2012a).
[c] Markers also used in the study of Lacape et al. (2007).

Tyagi et al. (2014) obtained 0.17 (378 accessions with 120 SSRs), Moiana et al. (2015) obtained 0.361 (20 accessions with 27 SSRs), and Qin et al. (2015) obtained an average PIC of 0.3 (241 accessions and 333 SSRs). To explain this result, we selected 20 SSRs that were used previously by Cai et al. (2014), who employed similar electrophoresis resolution systems. They analyzed 99 *G. hirsutum* and 2 *G. barbadense* genotypes and they obtained an average PIC value of 0.48 for these same 20 SSRs, while in our study a higher average PIC value of 0.63 was obtained. It is known that most *G. hirsutum* current cultivars share common parental inbred ancestors in their pedigrees (Meredith, 1998; Boopathi et al., 2011) and this usually causes low variability between most commercial materials worldwide. Nevertheless, our higher mean PIC value must be related to our different genotypes coming from different species, origins, and breeding companies, suggesting that our 36 *G. hirsutum* varieties represent a collection with high variability. Unexpectedly, a similar PIC value was found between 36 *G. hirsutum* and 10 *G. barbadense* accessions (0.57 and 0.54, respectively). These results suggest that we could find higher diversity with fewer *G. barbadense* accessions, probably because *G. barbadense*'s genetic background has suffered less breeding pressure from the cotton seed community.

The high variability obtained in our set of *G. hirsutum* varieties did not show a clear distribution according to their breeding origins. This could be due to the fact that most breeding programs have used similar germplasm, like the Spanish experimental variety SS-UPL-01 that comes from crosses involving cv. Lider, which has a Greek origin (Semillas Sostenibles Iberica S.L.U., personal communication). Another example is the case of Acala parentals, which are cultigens derived from early maturing Latifolium or Mexican Highlands and has been highly used in many intervarietal hybridizations by most

breeding programs worldwide (Iqbal et al., 2001). In our phylogenetic study it was observed that Acala germplasms (Acala Maxxa, Acala 3080, DP Acala 90, AI-270, and AI-292) did not cluster together, as probably they have more in common with other varietal lineages that could have been developed from them than with different Acala germplasms.

On the other hand, introgressions from other species materials have been extensively employed to increase *G. hirsutum* variability, as was shown in our phylogenetic results. First, we found cv. 108-F at the bottom of the H block, separated from the rest (Figure). This variety was developed by Rumshevich in 1940 and was widely cultivated during the 1960s in Uzbekistan (74% of the cotton growing area), Tajikistan, and Turkmenistan. Cv. 108-F was bred from *G. hirsutum* spp. *mexicanum* var. *nervosum* (Galicinsky et al., 1962) and this may be the reason for its genetic distance from the other *G. hirsutum*. Second, we also observed some cases of *G. barbadense* genes in our *G. hirsutum* collection. An example is the breeding material WC-19NSSL, which clustered far from the other *G. hirsutum* varieties of group H (Figure). This genotype is commonly named Okra leaf type because it possesses a semidominant mutation of cotton (L_2^0) that came from *G. barbadense* and alters leaf shape by increasing the length of lobes and decreasing lamina expansion (Dolan and Poethig, 1998). Another example is cv. Massala, which had the lowest genetic distance mean of 0.75 from *G. barbadense* varieties. Cv. Massala probably has *G. barbadense* genes that could have restored cv. Massala fertility, as it was derived from fertile off-type plants that came from male sterile parental blocks in an interspecific hybrid field (Semillas Sostenibles Iberica S.L.U., personal communication). Similar *G. barbadense*

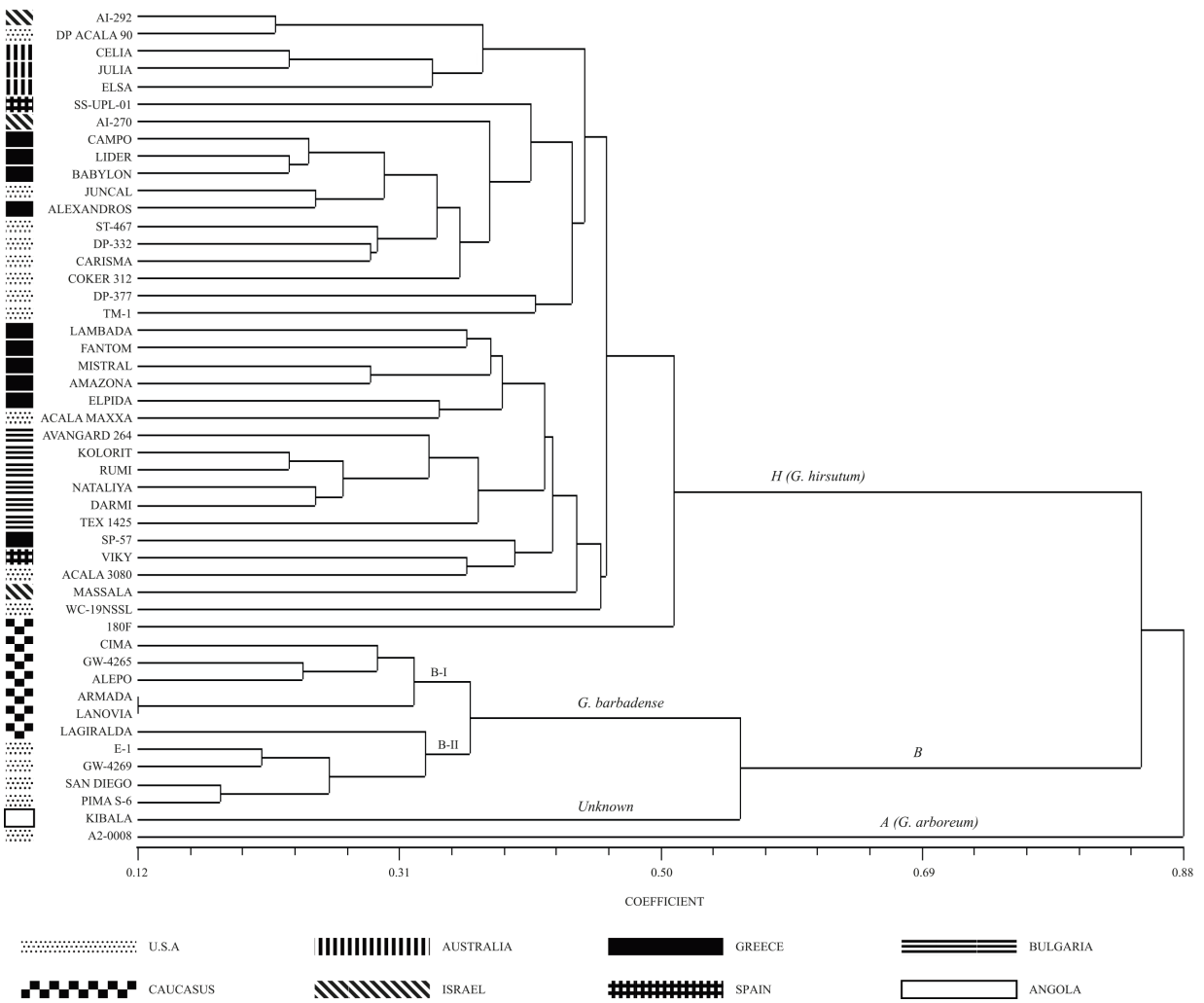

Figure. Dendrogram of 48 *Gossypium* genotypes using the UPGMA algorithm, constructed with 62 SSR markers.

introgressions have been found in *G. hirsutum* commercial materials, like the first 5 Bulgarian varieties (cv. Avangard 264 to cv. Darmi), whose pedigrees are well known and obtained by the crossing of *G. hirsutum* and inbred lines of *G. hirsutum* × *G. barbadense* (Stoilova, 2011; Stoilova et al., 2014). In addition, it seems that some of the Greek varieties have some *G. barbadense* origin, like cv. Fantom with its columnar brush structure and high fiber length (Stoilova, 2013), or cv. Amazona and cv. Elpida, which have very high fiber quality, typically from the *G. barbadense* varieties (Semillas Sostenibles Iberica S.L.U., personal communication). Other diversity studies obtained a lower genetic distance value compared to our value of 0.43. For example, Bertini et al. (2006) obtained 0.40, Tyagi et al. (2014) obtained 0.195, and Moiana et al. (2015) obtained 0.378. Consequently, the presence of varieties with introgression of *G. barbadense* germplasm and coming from different countries could explain the high average genetic distance and diversity found in our collection of Upland cotton.

The unknown accession that showed lower DNA content was defined as diploid *G. herbaceum* (A_1). Besides its significant phenotypical differences, this unknown accession produced spinnable fibers, which is an exclusive trait of A-genome diploids (*G. arboreum* and *G. herbaceum*) (Applequist et al., 2001). It has also been shown to be genetically similar to *G. barbadense*. These facts point at diploid *G. herbaceum* (A_1), which is the closest living ancestor of the A-genome of allotetraploidic cotton (Paterson et al., 2012). In addition, *G. herbaceum* has around two-thirds of the tetraploid DNA content (Hendrix and Stewart, 2005), which is similar to what we observed in the flow cytometer analysis. Nacoulima et al. (2016) and Kulkarni et al. (2009) reinforced the importance of diploid species to cotton breeders, especially *G. herbaceum* and *G. arboreum*. More research is required to explore their genetic potential, such as being resistant to drought, pests (white flies, thrips, and aphids), and leaf curl virus; having higher cotton fiber qualities (Mergeai, 2006); and solving lodging problems in columnar *G. barbadense* varieties like cv. Armada and cv. Alepo when they support the high weight of their green bolls.

In conclusion, the polymorphic markers and phylogenetic relationships between different varieties identified by this study may guide future breeding efforts in order to produce more competitive varieties. Taking into account our results, the high genetic variability found in our cotton germplasm collection could be employed in future breeding programs.

Acknowledgment

This research was involved in the "PhD Graduates at the Business Place" program, cofinanced by Algodonera del Sur S.A.; CeiA3 (Agrifood Campus of International Excellence); the Ministry of Education, Culture, and Sport; the Ministry of Economy, Industry, and Competitiveness; and Banco Santander.

References

Abdellatif KF, Soliman YA (2013). Genetic relationships of cotton (*Gossypium barbadense* L.) genotypes as studied by morphological and molecular markers. Afr J Biotechnol 12: 4736-4746.

Abdurakhmonov IY, Buriev ZT, Saha S, Pepper AE, Musaev JA, Almatov A, Shermatov SE, Kushanov FN, Mavlonov GT, Reddy UK et al. (2007). Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*. Euphytica 156: 141-156.

Abdurakhmonov IY, Buriev ZT, Shermatov SE, Abdullaev A, Urmonov K, Kushanov F, Egamberdiev SS, Shapulatov U, Abdukarimov A, Saha S et al. (2012). Genetic diversity in *Gossypium* genus. In: Galiskan M, editor. Genetic Diversity in Plants. Rijeka, Croatia: InTech, pp. 331-338.

Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, Salakhutdinov IB, Buriev ZT, Saha S, Scheffler BE et al. (2008). Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 92: 478-487.

Anderson JA, Churchill GA, Autrique JE (1993). Optimizing parental selection for genetic linkage maps. Genome 36: 181-186.

Applequist WL, Cronn R, Wendel JF (2001). Comparative development of fiber in wild and cultivated cotton. Evol Dev 1: 3-17.

Bertini C, Schuster I, Sediyama T, Barros E, Moreira M (2006). Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet Mol Biol 29: 321-329.

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape JM, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P et al. (2006). CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics 7: 132-141.

Boopathi NM, Thiyagu K, Urbi B, Santhoshkumar M, Gopikrishnan A, Aravind S, Swapnashri G, Ravikesavan R (2011). Marker-assisted breeding as next-generation strategy for genetic improvement of productivity and quality: can it be realized in cotton? International Journal of Plant Genomics 2011 :1-17.

Cai C, Ye W, Zhang T, Guo W (2014). Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum* L.). J Integr Plant Biol 56: 51-62.

COAG (2012). Cotton Put at Risk. Dossier Cotton COAG (Agricultural Trade Union of Andalusia) [online]. Website: http://www.coagandalucia.com/extras/noticias/DOSSIER_Algodon_El_algodon_se_la_juega.pdf [accessed 2 June 2017].

Cotton Inc. (2017) onward (continuously updated). Monthly Economic Letter, April 2017. Website: http://www.cottoninc.com/corporate/Market-Data/MonthlyEconomicLetter/ [accessed 20 April 2017].

Dolan L, Poethig RS (1998). The okra leaf shape mutation in cotton is active in all cell layers of the leaf. Am J Bot 85: 322-327.

Dong J, Wu F, Jin Z, Huang Y (2006). Heterosis for yield and some physiological traits in hybrid cotton. Euphytica 151: 71-77.

Elçi E, Akışcan Y, Akgöl B (2014). Genetic diversity of Turkish commercial cotton varieties revealed by molecular markers and fiber quality traits. Turk J Bot 38: 1274-1286.

Erdogan O, Ozbek N, Unay A, Gore ME (2013). The determination of relationship between Verticillium wilt (*Verticillium dahliae* Kleb.) and early maturity in cotton (*Gossypium hirsutum* L.). Turk J Field Crops 18: 8-12.

Eurostat (Statistics Explained, European Union) (2016) onward (continuously updated). Crop Statistics (From 2000 Onwards). Website: http://ec.europa.eu/eurostat/web/products-datasets/-/apro_acs_a [accessed 2 June 2017].

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

Galicinsky P, Demidov S, Obuhov M, Samoilov A (1962). Cotton varieties in Uzbekistan. In: Results of State Varieties Testing during 1950–1959. Tashkent, Uzbekistan: Government Press, pp. 17-19.

Guo W, Cai C, Wang C, Zhao L, Wang L, Zhang T (2008). A preliminary analysis of genome structure and composition in *Gossypium hirsutum*. BMC Genomics 9: 314.

Hendrix B, Stewart JM (2005). Estimation of the nuclear DNA content of *Gossypium* species. Ann Bot-London 95: 789-797.

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

Ivanova B, Bojinov B (2009). Identification of QTLs for fiber quality in Bulgarian cotton breeding collection. Genet Breed 38: 55-60.

Kantartzi SK (2013). Microsatellites: methods and protocols. In: Kantartzi SK, editor. Methods in Molecular Biology. New York, NY, USA: Humana Press.

Kaur S, Cogan NOI, Forster JW, Paull JG (2014). Assessment of genetic diversity in faba bean based on single nucleotide polymorphism. Diversity 6: 88-101.

Kulkarni VN, Khadi BM, Maralappanavar MS, Deshapande LA, Narayanan SS (2009). The worldwide gene pools of *Gossypium arboreum* L. and *G. herbaceum* L., and their improvement. In: Paterson AH, editor. Genetics and Genomics of Cotton. New York, NY, USA: Springer-Verlag, pp. 69-97.

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 Breeding 19: 45-58.

McCarty JC, Percy RG (2001). Genes from exotic germplasm and their use in cultivar improvement in *Gossypium hirsutum* L. and *G. barbadense* L. In: Jenkins JN, Saha S, editors. Genetic Improvement of Cotton. Emerging Technologies. Enfield, UK: Science Publishers Inc., pp. 65-80.

Meredith WR Jr (1998). Continued progress for breeding for yield in the USA. In: Proceedings of the World Cotton Research Conference II, Athens (Greece), pp. 97-101.

Mergeai G (2006). Cotton improvement through interspecific hybridization. Cah Agric 15: 135-143 (in French with an abstract in English).

Mishra KK, Fougat RS (2013). Genetic relationship among different species of cotton as revealed by SSR markers for fiber quality traits. International Journal of Pure & Applied Bioscience 1: 81-93.

Moiana LD, Vidigal Filho PS, Gonçalves-Vidigal MC, Carvalho LP (2015). Genetic diversity and population structure of upland cotton Brazilian cultivars (*Gossypium hirsutum* L. race *latifolium* H.) using SSR markers. Aust J Crop Sci 9: 143-152.

Moreno R, Espejo JA, Cabrera A, Millan T, Gil J (2006). Ploidic and molecular analysis of 'Morado de Huetor' asparagus (*Asparagus officinale* L.) population; a Spanish tetraploid landrace. Genet Resour Crop Ev 53: 729-736.

Nacoulima NL, Diouf HF, Konan ON, Mergeai G (2016). Production of new cotton interspecific hybrids with enhanced fiber fineness. J Agr Sci 8: 46-56.

Paterson AH, Brubaker CL, Wendel JF (1993). A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP and PCR analysis. Plant Mol Biol Rep 11: 112-127.

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J et al. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 492: 423-427.

Poortavakoli S, Sheidai M, Alishah O, Noormohammadi Z (2017). Genetic diversity analysis in drought stress tolerant cottons. Nucleus 60: 57-62.

Qin H, Chen M, Yi X, Bie S, Zhang C, Zhang Y, Lan J, Meng Y, Yuan Y, Jiao C (2015). Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. PLoS One 10: e0118073.

Sapkal DP, Sutar SR, Thakre PB, Patil BR, Paterson AH, Waghmare VN (2011). Genetic diversity analysis of maintainer and restorer accessions in upland cotton (*Gossypium hirsutum* L.). J Plant Biochem Biot 20: 20-28.

Shen X, Cao Z, Singh R, Lubbers EL, Xu P, Smith CW, Paterson AH, Chee PW (2011). Efficacy of qFL-chr1, a quantitative trait locus for fiber length in cotton (*Gossypium* spp.). Crop Sci 51: 1-6.

Stoilova A (2011). Breeding of Bulgarian cotton varieties with improved fiber quality. Vavilov Journal of Genetics and Breeding 15: 550-553.

Stoilova A (2013). Inheritance of Fiber Length in Diallel Cotton Crosses [Online]. Website https://www.icac.org/meetings/ [accessed 16 October 2017].

Stoilova A, Valkova N, Hadzhiivanova B, Koleva M, Nedyalkova S (2014). The cotton breeding in Bulgaria. Turkish Journal of Agricultural and Natural Sciences 1: 992-999.

Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V (2014). Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). Theor Appl Genet 127: 283-295.

Ullah I, Iram A, Iqbal MZ, Nawaz M, Hasni SM, Jamil S (2012). Genetic diversity analysis of Bt cotton genotypes in Pakistan using simple sequence repeat markers. Genet Mol Res 11: 597-605.

Yu JW, Zhang K, Li SY, Yu SX, Zhai HH, Wu M, Li X, Fan S, Song M, Yang D et al. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. Theor Appl Genet 126: 275-287.

Yu JZ (2004). A standard panel of *Gossypium* genotypes established for systematic characterization of cotton microsatellite markers. Plant Breeding News 2004: 148.

Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, Zhang J, Chee P, Scheffler BE, Jones DC (2012a). Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. Euphytica 187: 203-213.

Yu JZ, Kohel RJ, Fang DD, Cho J, Van Deynze A, Ulloa M, Hoffman SM, Pepper AE, Stelly DM, Jenkins JN et al. (2012b). A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. G3-Genes Genom Genet 2: 43-58.

Zhao Y, Wang H, Chen W, Li Y, Gong H, Sang X, Huo F, Zeng F (2015). Genetic diversity and population structure of elite cotton (*Gossypium hirsutum* L.) germplasm revealed by SSR markers. Plant Syst Evol 301: 327-336.

Table S1. Information about sequences, original source, publication origin (www.cottongen.org/), linkage position, and chromosome location of the final 62 SSR markers, with their optimization of multiplexed PCR conditions for this diversity study.

	Marker	Repeat type	Forward (primer 5'–3')	Reverse (primer 5'–3')	Original source	SSR type	Species	SSR publication*	Cmap	Locus (chromosome, cM)	Nº chromos. amplification	Triplex	Primer label	Tª annealing	Cf $MgCl_2$
1	BNL0387	(AG)15	GAAGGGGAATTTATAGCGGG	AGAGACTCCCGCACTTGAAA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Yu et al., 2012	A8(55.645)/D8(30.31)	8/24	10	FAM	58	2.5
2	BNL0530	(GA)10	CGTAGGATGGAAACGAAAGC	GCCACACTTTTCCCTCTCAA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A4(67.344)	4	18	HEX	55	2.5
3	BNL0569	(AG)20	TTGAGAAGTACTACCATTAATTATCCA	GACTGATGCCAGTTGACCCT	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Yu et al., 2012	A13(88.63)/ D13(60.359)	13/18	15	HEX	55	2.5
4	BNL1122	(AG)16	TCGATAACGGCTATAGTAATCTCTC	CAACAAATAAGCAGCCAAGAAA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D7(89.00)	16	2	HEX	55	2.5
5	BNL1227	(AG)15	CATCAAGATCTATCTCTCTCTATACCG	TTTACCCTCCGATCTCAACG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D12(48.302)	26	5	FAM	55	2.5
6	BNL1423	(AG)12	AAAAACCAATTGCCTCCAAA	CTCCTAACGATCAATGGGGA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Yu et al., 2012	A9(113.255)	9	1	NED	55	2.5
7	BNL1521	(AG)26	TGAAGAAAGAAAAAGAGAAAGGG	CTCACCACGTGGCACTTATG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D8(73.48)	24	5	NED	55	2.5
8	BNL1531	(AG)14	CTGCAACAAGAGCCTGTGTC	ATGGAGATTGGCTGAGATGG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Yu et al., 2012	A7(43.29)/D7(37.90)	7/16	12	HEX	55	3
9	BNL1551	(AG)22	CGCAAGCCACCTGTAAAAC	TCGAATTTTCTCTCTCTCTCTCTCT	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D11(63.38)	21	6	HEX	55	2.5
10	BNL1672	(AG)14	TGGATTTGTCCCTCTGTGTG	AACCAACTTTTCCAACACCG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A9(60.015)/D9(70.92)	9/23	21	FAM	58	3
11	BNL1897	TA+GA	TTAAAAGAAGATTGAGATGACATTATG	ATTTGTACTTAGATGTGCCAAATG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A2(41.94)	2	6	FAM	55	3
12	BNL2495	(AG)14. (TC)14	ACCGCCATTACTGGACAAAG	AATGGAATTTGAACCCATGC	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D12(73.35)	26	9	HEX	55	2
13	BNL2544	(AG)11	GCCGAAACTAAAACGTCCAA	TCCTTACTCACTAAGCAGCCG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D13(94.59)	18	2	NED	55	3
14	BNL2570	(GA)13	TTCTACAAAAAAAGAAAAAATGGG	AAATACGGATGGGACCAACC	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	D10(45.18)	20	21	FAM	55	2.5
15	BNL2572	(GA)23	GTCCTATTACTAAAATTGTTAATTTAGCC	CGATGTTAAATCAATCAGGTCA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A4(93.7)	4	9	FAM	55	3
16	BNL2921	(AG)10	CGAGAGATTTTAAAGGGAAACA	GGGAGTGGTCTGATGGAAAA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A1(71.545)	1	3	FAM	55	2
17	BNL3090	(AG)31	GAAATCATTGGAAGAACATATACTACA	TTGCTCCGTATTTTCCAGCT	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A1(63.465)/D1(54.70)	15	7	FAM	55	2.5
18	BNL3255	(GC)6AT (AC)14	GACAGTCAAACAGAACAGATATGC	TTACACGACTTGTTCCCACG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A8(51.161)	8	3	NED	55	2.5
19	BNL3257	(AC)13+ (AT)10	CAATCTGGGATCAAAAAAACC	GGTGAAACATAGCGTGTTGC	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A8(69.707)	8	11	FAM	55	2.5
20	BNL3308	(GA)10	TTCCCCTGCTTGTCAAAAAC	GCAAGTCTCGTTGGGAAAAA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A7(88.989)	7	11	NED	55	2.5
21	BNL3408	(GT)2AT (GT)12	ATCCAAACCATTGCACCACT	GTGTACGTTGAGAAGTCATCTGC	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Yu et al., 2012	A3(94.677)/D3(100.34)	3/17	7	HEX	55	2.5
22	BNL3474	(CA)16	AAGGTAATGCAGTGCGGTTC	ATAATGGCATTGATTATAGAGTGTG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A8(79.31)/D8(92.40)/ D8(97.039)	8/24	10	NED	55	2.5
23	BNL3545	(CA)10	AGTCAGTTTTTTGTTAGCAATATGC	AACCATTAATTCCCTATTTAACCG	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Yu et al., 2012	A2(117.49)/D2(118.87)	2/14	16	NED	58	2.5
24	BNL3563	(CA)13 (TA)4	AAGCATAAACTTGACACAAGCC	AATGGGCAAGAAAAGGGAAC	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A10(72.15)	10	4	FAM	55	3
25	BNL3580	(AC)15	CTTGTTTACATTCCCTTCTTTATACC	CAAAGGCGAACTCTTCCAAA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A1(49.03)	1	15	NED	55	2.5

Table S1. (Continued).

26	BNL3590	(CA)20	TCTTCCCTCTCTTTCTCTTTCG	ACACGGAAGACCAACCAAGT	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A2(50.676)/ D3(17.772)/D3(4.804)	17	4	NED	55	2.5
27	BNL3594	(TC)37	AGGGATTTTGATTGTTGTGC	TGAATTCAAAACAAATGTTAGCC	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A6(6.76)/D6(0)	6/25	16	HEX	58	2.5
28	BNL3992	(TC)26. (GA)26	CAGAAGAGGAGGAGGTGGAG	TGCCAATGATGGAAAACTCA	Brookhaven National Laboratory (USA)	Genomic	*G. hirsutum*	Liu et al., 2000	Guo et al., 2008a	A5(95.804)	5	20	HEX	58	2.5
29	CIR0105	(AC)9	GTCTCTTGTCTTTCTTTCTTAC	AACCAAACTGAACCCA	CIRAD, Montpellier (France)	Genomic	*G. hirsutum*	Nguyen et al., 2004	Yu et al., 2012	D1(16.31)	15	4	NED	55	2.5
30	CIR0165	(AC)13	ATAAGTGGAGACAGGCA	GACCAGCACAGGAAAC	CIRAD, Montpellier (France)	Genomic	*G. hirsutum*	Nguyen et al., 2004	Yu et al., 2012	A5(22.32)/D5(22.99)	5/19	14	NED	55	2.5
31	CIR0169	(AC)8	GAAGCACAATAAGGCAA	CAAACAAGCGATGAAAC	CIRAD, Montpellier (France)	Genomic	*G. hirsutum*	Nguyen et al., 2004	Yu et al., 2012	A7(57.28)	7	17	HEX	55	2.5
32	CIR0203	(TG)11 (N)1(TG)5	AGTTCAAGGGCACAAA	ATCTCCAAGTCCCACC	CIRAD, Montpellier (France)	Genomic	*G. hirsutum*	Nguyen et al., 2004	Guo et al., 2008a	A6(123.61)	6	14	FAM	52	2.5
33	CIR0246	(TG)6	TTAGGGTTTAGTTGAATGG	ATGAACACACGCACG	CIRAD, Montpellier (France)	Genomic	*G. hirsutum*	Nguyen et al., 2004	Guo et al., 2008a	D2(112.572)	14	18	NED	55	3
34	CIR0413	(CA)33	TTAAAGCTCACACACACA	CAACAGTAACGAAGAACAAT	CIRAD, Montpellier (France)	Genomic	*G. hirsutum*	Nguyen et al., 2004	Guo et al., 2008a	D5(127.86)/D6(56.11)/ D8(120.22)	19/25/24	10	HEX	58	2.5
35	JESPR0056	(GAA)23	CCAGTTAGCACCAATTTAGG	CCACAATAACACACTGGAATC	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	A1(62.1)/D10(60.42)/ D13(56.88)	1/20/18	12	NED	55	2.5
36	JESPR0127	(GA)9AA (GA)5	GATTTGGGTAACATTGGCTC	CTGCAGTGTTGTGTTGGGTAGA	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	D8(97.939)	24	17	FAM	55	2.5
37	JESPR0152	(GAA)50	GATGCACCAGATCCTTTTATTAG	GGTACATCGGAATCACAGTG	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	D1(104.91)	15	21	NED	55	2.5
38	JESPR0153	(CTA)18	GATTACCTTCATAGGCCACTG	GAAAACATGAGCATCCTGTG	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	A13(63.046)/ D13(65.823)	13/18	16	FAM	58	2.5
39	JESPR0208	(CT)15	CGCAACCAAACATATACTTCACAC	CCCTTTCCATCCATAGAACG	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	A9(96.388)/ D9(119.133)	9/23	13	HEX	58	2.5
40	JESPR0220	(GA)20	CGAGGAAGAAATGAGGTTGG	CTAAGAACCAACATGTGAGACC	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	D4(102.491)	22	17	NED	58	2.5
41	JESPR0274	(CA)13	GCCCACTCTTTCTTCAACAC	TGATGTCATGTGCCTTGC	Texas A&M University and USDA-ARS (MS, USA)	Genomic	*G. hirsutum*	Reddy et al., 2001	Guo et al., 2008a	A9(14.968)/ A9(42.802)/D9(52.97)	9/9/23	1	FAM	55	2
42	MGHES0031	(CAT)9	AAGTTAGCGGCTTCTTGTGG	GGGTCAGAACTGGACAAGGA	USDA-ARS, Starkville (MS, USA)	EST-74%fiber	*G. hirsutum*	Qureshi et al., 2004	Yu et al., 2007	D12(87.5)	26	6	NED	55	2
43	MGHES0055	(CAT)5	CGAACCCTAGCTTTCAATCG	CGGCTTCAATTGTACGGTCT	USDA-ARS, Starkville (MS, USA)	EST-74%fiber	*G. hirsutum*	Qureshi et al., 2004	-	-		15	FAM	55	2.5
44	NAU0905	(AAT)17	TGGCTGAACTTTGCAATTTA	AAGCAAGGGAGGTAATCCTT	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A6(60.245)/D6(51.415)	6/25	19	HEX	55	2.5
45	NAU0934	(TC)11	TGCTTTCGTATCCTTTTTCC	ATTAGAGAAGCCAGGGAGGT	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A5(186.59)	5	20	FAM	55	2.5
46	NAU0943	(AAT)20	ATCTGTTCAATTTCTCGTCA	CAGTTGTTGGTTGATCTGGA	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A12(84.933)	12	20	NED	58	2.5
47	NAU0998	(TCATG)4	CTCTCTCTCACACACGCACT	CTGGCCAAGTGACACAATTA	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	D2(48.11)	14	14	HEX	55	2.5
48	NAU1042	(TCAGGC)4	CATGCAAATCCATGCTAGAG	GGTTTCTTTGGTGGTGAAAC	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A5(46.74)/D5(70.123)/ D5(56.1)	5/19/19	3	HEX	55	2.5
49	NAU1043	(TTC)14	GTATCCGCCCACAAATAAAG	GCATCGTGAGAGAAAGTGAA	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Shen et al., 2007	A7	7	19	FAM	55	2.5
50	NAU1070	(AGG)10	CCCTCCATAACCAAAAGTTG	ACCAACAATGGTGACCTCTT	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A3(50.113)/D2(12.779)	3/14	4	HEX	55	2.5

Table S1. (Continued).

51	NAU1162	(CAG)6	CTGAGTGAGCCATGAAGTTG	TTCGGCTTCTGCTTTTACTT	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A11(99,95)	11	1	HEX	55	3
52	NAU1167	(GATAGG)4	CTGACTTGGACCGAGAACTT	AAGAGCCCTGGACAATGATA	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	A3(105,33)	3	18	FAM	55	2.5
53	NAU1200	(CAG)11	CAACAGCAACAACCACAA	CTGCCTCGAGGACAAATAGT	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Shen et al., 2007	A5(16,6)	5	19	NED	55	2.5
54	NAU1221	(TCAAGC)5	CATGCAAATCCATGCTAGAG	AGGTTTCTTTGGTGGTGAAA	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	D5(69,644)	19	8	FAM	55	2.5
55	NAU1322	(CATCTC)3	CTCCAATCGAATGATTTTT	GGTAGGGTTTTGGAGGTTTT	Nanjing Agricultural University (China)	EST	*G. arboreum*	Han et al., 2004	Guo et al., 2008a	D8(89,78)	24	8	HEX	55	2.5
56	NAU2508	(TGGGGT)5	TGGAGGAGGGTGTAACATCT	GGCATTCAAGGAGATGAGTT	Nanjing Agricultural University (China)	EST	*G. hirsutum*	Han et al., 2006	Guo et al., 2008a	A10(126,83)	10	11	HEX	55	2.5
57	NAU2649	(TTG)5	GGTGAAACTTTCTCGTTGCT	ATGGCTTCACTTTCTCTCCA	Nanjing Agricultural University (China)	EST	*G. raimondii*	Guo et al., 2007	Guo et al., 2008a	D3(28,51)	17	13	FAM	55	2.5
58	NAU3341	(ATAC)5	TTTGATACCACCCATCACAG	CAGCCATGGATATGTTTCAG	Nanjing Agricultural University (China)	EST	*G. raimondii*	Guo et al., 2007	Guo et al., 2008a	A11(150.43)/ D11(159.149)	11/21	2	FAM	55	2.5
59	NAU3770	(TA)6CAT (GT)9	GCAGAAACCTCGAATCTTGT	AACCCAGCAAATTCATTCAT	Nanjing Agricultural University (China)	EST	*G. raimondii*	Guo et al., 2007	Guo et al., 2008a	A11(17,846)	11	13	NED	55	2.5
60	NAU6634	TAT(15)	TGCCTTATATACCCCATTTTTC	CAGACTTCAAAATAGGCTTATGG	Nanjing Agricultural University (China)	BAC	*G. hirsutum*	Guo et al., 2008b	Guo et al., 2008a	D7(74.854)/D3 (33.495)	17/16	12	FAM	55	2.5
61	TMB0366	(GA)17+ (CA)8	GAGCCCACCATTATCACTCC	GGTGGTCATGTGAGAGAGGA	USDA-ARS, College Station (TX, USA)	BAC	*G. hirsutum*	Yu et al., 2002; He et al., 2007	He et al., 2007	D5 (69,0)	19	9	NED	58	2.5
62	TMB0471	(GA)30	AAGAATTAGCGGAAGTGGTCA	TTTGACAAAACATGGATGGA	USDA-ARS, College Station (TX, USA)	BAC	*G. hirsutum*	Yu et al., 2002; He et al., 2007	Guo et al., 2008a	A2(53.529)/D3(0)	2/17	21	HEX	55	2.5

*References:

Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007). A microsatellite-based, gene-rich linkage map reveals genome structure, function, and evolution in *Gossypium*. Genetics 176: 527-541.

Guo W, Cai C, Wang C, Zhao L, Wang L, Zhang T (2008a). A preliminary analysis of genome structure and composition in *Gossypium hirsutum*. BMC Genomics 9: 314.

Guo Y, Saha S, Yu JZ, Jenkins JN, Kohel RJ, Scheffler BE, Stelly DM (2008b). BAC-derived SSR markers chromosome locations in cotton. Euphytica 161: 361.

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

Han ZG, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006). Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor Appl Genet 112: 430-439.

He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, Zhang YX, Li W (2007). QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. Euphytica 153: 181.

Liu S, Saha S, Stelly D, Burr B, Cantrell RG (2000). Chromosomal assignment of microsatellite loci in cotton. J Hered 91: 326-332.

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.

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004). EST-SSR: A new class of genetic markers in cotton. Journal of Cotton Science 8: 112-123.

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

Shen X, Guo WZ, Lu QX, Zhu XF, Yuan YL, et al., (2007). Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155: 371-380.

Yu JW, Yu SX, Lu CR, Wang W, Fan SL, Song MZ, Lin ZX, Zhang XL, Zhang JF, Wu W (2007). High-density linkage map of cultivated allotetraploid cotton based on SSR, TRAP, SRAP and AFLP Markers. J Integr Plant Biol 49: 716-724.

Yu JZ, Kohel RJ, Dong RJ (2002). Development of integrative SSR markers from TM-1 BACs. In: Proceedings of the Beltwide Cotton Improvement Conference by National Cotton Council of America, Atlanta, GA, USA. CD-ROM.

Yu JZ, Kohel RJ, Fang DD, Cho J, Van Deynze A, Ulloa M, Hoffman SM, Pepper AE, Stelly DM, Jenkins JN et al. (2012b). A high-density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. G3 (Bethesda) 2: 43-58.

Table S2. Genetic distance matrix computed with Jaccard coefficient among the 48 cotton varieties.

	ALEPO	CIMA	ARMADA	LANOVIA	GW-4265	LAGIRALDA	E1	SAN DIEGO	GW-4269	SS-UPL-01	AI-292	AI-270	CAMPO	LIDER	BABYLON	LAMBADA	MISTRAL	AMAZONA	CELIA	ELSA	JUNCAL	ST-467	ALEXANDROS	AVANGARD 264	KOLORIT	NATALIYA	DARMI	RUMI	DP ACALA 90	ELPIDA	SP-57	MASSALA	DP-332	DP-377	COKER 312	CARISMA	VIKY	JULIA	FANTOM	TEX 1425	WC-19NSSL	KIBALA	A2-0008 (CMD9)	PIMA S-6 (CMD8)	TM-1 (CMD1)	ACALA MAXXA (CMD3)	I80F	ACALA 3080
ALEPO	0	0.26	0.36	0.33	0.24	0.38	0.37	0.35	0.37	0.85	0.86	0.87	0.88	0.88	0.88	0.85	0.87	0.85	0.88	0.86	0.88	0.87	0.85	0.84	0.85	0.86	0.85	0.86	0.88	0.89	0.85	0.75	0.88	0.88	0.88	0.88	0.87	0.87	0.85	0.85	0.85	0.56	0.9	0.35	0.85	0.87	0.83	0.84
CIMA		0	0.34	0.4	0.33	0.37	0.35	0.31	0.31	0.84	0.86	0.87	0.88	0.89	0.88	0.84	0.86	0.83	0.88	0.87	0.88	0.85	0.85	0.83	0.83	0.84	0.83	0.84	0.88	0.86	0.84	0.76	0.88	0.85	0.9	0.87	0.85	0.87	0.83	0.85	0.85	0.57	0.92	0.31	0.83	0.84	0.84	0.84
ARMADA			0	0.12	0.26	0.35	0.4	0.37	0.34	0.84	0.86	0.86	0.87	0.88	0.88	0.84	0.85	0.83	0.86	0.85	0.87	0.86	0.85	0.83	0.83	0.84	0.84	0.84	0.87	0.87	0.84	0.76	0.87	0.86	0.88	0.86	0.86	0.86	0.84	0.84	0.82	0.56	0.91	0.4	0.84	0.86	0.82	0.84
LANOVIA				0	0.24	0.38	0.4	0.41	0.38	0.85	0.87	0.87	0.88	0.88	0.88	0.84	0.85	0.83	0.87	0.86	0.88	0.87	0.86	0.83	0.84	0.85	0.84	0.85	0.88	0.88	0.84	0.76	0.87	0.86	0.88	0.87	0.87	0.86	0.85	0.84	0.83	0.59	0.91	0.43	0.85	0.86	0.82	0.84
GW-4265					0	0.32	0.36	0.35	0.34	0.85	0.87	0.88	0.88	0.89	0.89	0.85	0.87	0.85	0.87	0.86	0.89	0.88	0.85	0.84	0.85	0.85	0.85	0.86	0.89	0.88	0.85	0.76	0.88	0.87	0.89	0.88	0.87	0.87	0.85	0.85	0.84	0.57	0.9	0.37	0.85	0.87	0.83	0.84
LAGIRALDA						0	0.36	0.32	0.33	0.84	0.88	0.86	0.88	0.89	0.88	0.86	0.88	0.85	0.89	0.86	0.89	0.87	0.85	0.82	0.85	0.85	0.84	0.86	0.89	0.88	0.85	0.77	0.89	0.86	0.89	0.87	0.85	0.87	0.84	0.85	0.86	0.57	0.91	0.31	0.83	0.86	0.84	0.84
E1							0	0.26	0.21	0.85	0.87	0.88	0.89	0.89	0.89	0.86	0.86	0.85	0.89	0.87	0.89	0.86	0.87	0.84	0.85	0.84	0.83	0.85	0.89	0.88	0.85	0.76	0.87	0.85	0.89	0.88	0.85	0.87	0.85	0.86	0.84	0.58	0.91	0.31	0.84	0.85	0.84	0.85
SAN DIEGO								0	0.24	0.83	0.86	0.86	0.87	0.88	0.87	0.85	0.86	0.83	0.87	0.85	0.87	0.84	0.86	0.82	0.83	0.82	0.82	0.84	0.88	0.87	0.83	0.73	0.87	0.85	0.87	0.85	0.83	0.86	0.83	0.85	0.84	0.53	0.9	0.18	0.84	0.86	0.83	0.84
GW-4269									0	0.84	0.88	0.88	0.89	0.89	0.88	0.86	0.87	0.85	0.88	0.86	0.88	0.86	0.86	0.84	0.85	0.85	0.84	0.85	0.89	0.87	0.85	0.74	0.89	0.86	0.89	0.87	0.85	0.87	0.84	0.85	0.86	0.56	0.91	0.23	0.84	0.85	0.84	0.85
SS-UPL-01										0	0.54	0.4	0.34	0.39	0.37	0.41	0.46	0.48	0.53	0.43	0.45	0.41	0.42	0.43	0.42	0.43	0.41	0.37	0.51	0.44	0.44	0.51	0.43	0.47	0.41	0.45	0.4	0.49	0.49	0.46	0.48	0.83	0.87	0.82	0.52	0.46	0.51	0.46
AI-292											0	0.48	0.45	0.47	0.46	0.47	0.47	0.48	0.37	0.43	0.48	0.47	0.47	0.47	0.45	0.44	0.43	0.48	0.22	0.55	0.55	0.56	0.47	0.49	0.44	0.49	0.5	0.4	0.5	0.49	0.49	0.83	0.89	0.87	0.52	0.49	0.58	0.46
AI-270												0	0.32	0.34	0.34	0.52	0.45	0.44	0.52	0.46	0.44	0.38	0.38	0.51	0.46	0.42	0.48	0.46	0.51	0.48	0.51	0.57	0.37	0.44	0.41	0.41	0.47	0.53	0.59	0.54	0.46	0.86	0.87	0.86	0.47	0.53	0.57	0.48
CAMPO													0	0.23	0.26	0.43	0.42	0.5	0.42	0.38	0.31	0.36	0.29	0.42	0.41	0.43	0.45	0.39	0.39	0.41	0.45	0.5	0.36	0.44	0.31	0.37	0.42	0.38	0.46	0.48	0.47	0.86	0.87	0.87	0.43	0.47	0.49	0.44
LIDER														0	0.23	0.47	0.41	0.5	0.39	0.33	0.32	0.39	0.32	0.48	0.44	0.43	0.4	0.41	0.43	0.43	0.41	0.49	0.28	0.4	0.33	0.3	0.45	0.41	0.52	0.45	0.46	0.87	0.86	0.89	0.43	0.47	0.52	0.44
BABYLON															0	0.43	0.45	0.47	0.43	0.36	0.29	0.35	0.27	0.5	0.41	0.44	0.41	0.38	0.44	0.48	0.43	0.52	0.27	0.36	0.37	0.29	0.45	0.39	0.5	0.46	0.47	0.86	0.88	0.87	0.41	0.5	0.5	0.47
LAMBADA																0	0.41	0.34	0.44	0.41	0.46	0.45	0.4	0.36	0.34	0.39	0.4	0.35	0.43	0.42	0.4	0.42	0.48	0.45	0.44	0.48	0.48	0.49	0.36	0.37	0.51	0.82	0.88	0.84	0.52	0.39	0.48	0.4
MISTRAL																	0	0.29	0.4	0.41	0.44	0.46	0.44	0.41	0.39	0.38	0.36	0.41	0.49	0.4	0.41	0.43	0.42	0.41	0.43	0.51	0.4	0.48	0.36	0.46	0.4	0.82	0.85	0.85	0.55	0.36	0.49	0.37
AMAZONA																		0	0.48	0.44	0.48	0.46	0.39	0.41	0.41	0.4	0.42	0.42	0.46	0.35	0.46	0.44	0.49	0.38	0.5	0.47	0.47	0.48	0.4	0.45	0.48	0.81	0.88	0.84	0.55	0.41	0.46	0.42
CELIA																			0	0.32	0.38	0.53	0.42	0.46	0.45	0.46	0.46	0.48	0.32	0.48	0.48	0.5	0.52	0.49	0.38	0.47	0.43	0.23	0.46	0.48	0.56	0.85	0.88	0.87	0.51	0.43	0.56	0.44
ELSA																				0	0.38	0.44	0.38	0.47	0.39	0.42	0.34	0.35	0.41	0.42	0.41	0.42	0.4	0.43	0.38	0.32	0.39	0.35	0.47	0.45	0.46	0.83	0.88	0.85	0.51	0.42	0.49	0.39
JUNCAL																					0	0.36	0.25	0.46	0.43	0.45	0.43	0.4	0.42	0.47	0.38	0.5	0.35	0.43	0.31	0.36	0.4	0.34	0.46	0.48	0.49	0.87	0.88	0.87	0.5	0.43	0.48	0.43
ST-467																						0	0.35	0.5	0.44	0.45	0.44	0.42	0.5	0.44	0.45	0.55	0.3	0.44	0.38	0.29	0.47	0.48	0.46	0.51	0.42	0.85	0.88	0.85	0.46	0.47	0.49	0.46
ALEXANDROS																							0	0.41	0.37	0.39	0.41	0.35	0.4	0.44	0.43	0.48	0.36	0.36	0.38	0.32	0.48	0.36	0.46	0.44	0.46	0.85	0.87	0.85	0.38	0.45	0.5	0.45
AVANGARD 264																								0	0.32	0.35	0.32	0.34	0.41	0.42	0.49	0.44	0.55	0.51	0.47	0.48	0.48	0.44	0.42	0.37	0.49	0.82	0.89	0.81	0.51	0.41	0.52	0.42
KOLORIT																									0	0.26	0.31	0.23	0.44	0.46	0.45	0.46	0.46	0.46	0.43	0.42	0.47	0.49	0.42	0.4	0.44	0.81	0.88	0.82	0.46	0.41	0.48	0.41
NATALIYA																										0	0.25	0.27	0.46	0.51	0.46	0.48	0.43	0.43	0.42	0.45	0.43	0.48	0.5	0.38	0.42	0.81	0.88	0.82	0.45	0.47	0.49	0.42
DARMI																											0	0.24	0.42	0.44	0.38	0.4	0.45	0.45	0.4	0.41	0.39	0.45	0.47	0.35	0.39	0.8	0.89	0.82	0.49	0.38	0.5	0.4
RUMI																												0	0.42	0.41	0.34	0.43	0.44	0.48	0.41	0.4	0.44	0.44	0.42	0.34	0.44	0.81	0.88	0.83	0.5	0.37	0.46	0.42
DP ACALA 90																													0	0.47	0.48	0.54	0.52	0.49	0.39	0.44	0.51	0.3	0.46	0.49	0.54	0.85	0.89	0.88	0.5	0.44	0.56	0.5
ELPIDA																														0	0.42	0.41	0.51	0.48	0.47	0.42	0.46	0.45	0.39	0.45	0.46	0.84	0.86	0.87	0.57	0.34	0.53	0.48

Table S2. (Continued).

SP-57																															0	0.49	0.45	0.48	0.38	0.43	0.4	0.46	0.37	0.4	0.42	0.83	0.87	0.84	0.57	0.4	0.47	0.39
MASSALA																																0	0.53	0.53	0.5	0.5	0.49	0.53	0.47	0.39	0.53	0.73	0.9	0.73	0.57	0.41	0.51	0.44
DP-332																																	0	0.37	0.38	0.29	0.47	0.47	0.54	0.47	0.41	0.85	0.87	0.88	0.46	0.54	0.5	0.47
DP-377																																		0	0.47	0.43	0.49	0.48	0.48	0.47	0.44	0.84	0.88	0.84	0.41	0.48	0.56	0.46
COKER 312																																			0	0.38	0.43	0.42	0.45	0.48	0.48	0.85	0.87	0.89	0.51	0.45	0.56	0.39
CARISMA																																				0	0.52	0.41	0.5	0.46	0.45	0.86	0.88	0.86	0.43	0.5	0.52	0.47
VIKY																																					0	0.44	0.44	0.47	0.5	0.82	0.87	0.83	0.55	0.41	0.52	0.36
JULIA																																						0	0.48	0.47	0.5	0.85	0.89	0.86	0.49	0.45	0.56	0.52
FANTOM																																							0	0.47	0.5	0.82	0.88	0.82	0.59	0.37	0.5	0.37
TEX 1425																																								0	0.47	0.81	0.9	0.85	0.53	0.44	0.49	0.44
WC-19NSSL																																									0	0.81	0.87	0.84	0.5	0.46	0.51	0.42
KIBALA																																										0	0.89	0.51	0.86	0.84	0.83	0.82
A2-0008 (CMD9)																																											0	0.9	0.86	0.88	0.87	0.86
PIMA S-6 (CMD8)																																												0	0.84	0.84	0.83	0.83
TM-1 (CMD1)																																													0	0.55	0.58	0.56
ACALA MAXXA (CMD3)																																														0	0.5	0.36
180F																																															0	0.46
ACALA 3080																																																0