Theor Appl Genet (2009) 119:281–292
DOI 10.1007/s00122-009-1037-y

ORIGINAL PAPER

A new interspecific, *Gossypium hirsutum* × *G. barbadense*, RIL population: towards a unified consensus linkage map of tetraploid cotton

Jean-Marc Lacape · J. Jacobs · T. Arioli · R. Derijcker · N. Forestier-Chiron · D. Llewellyn · J. Jean · E. Thomas · C. Viot

Received: 9 January 2009 / Accepted: 5 April 2009 / Published online: 23 April 2009

Abstract We report the development of a new interspecific cotton recombinant inbred line (RIL) population of 140 lines deriving from an interspecific cross between *Gossypium hirsutum* (*Gh*) and *G. barbadense* (*Gb*), using the same two parents that have served for the construction of a BC_1 map and for the marker-assisted backcross selection program underway at CIRAD. Two marker systems, microsatellites and AFLPs, were used. An important feature of the RIL population was its marked segregation distortion with a genome-wide bias to *Gh* alleles (parental genome ratio is 71/29). The RIL map displays an excellent colinearity with the BC_1 map, although it is severely contracted in terms of map size. Existence of 255 loci in common (between 6 and 14 per chromosome) allowed the integration of the two data sets. A consensus BC_1–RIL map based upon 215 individuals (75 BC1 + 140 RIL) was built. It consisted of 1,745 loci, spanned 3,637 cM, intermediate between the sizes of the two component maps, and constituted a solid framework to cross align cotton maps using common markers. The new RIL population will be further exploited for fiber property QTL mapping and eQTL mapping.

Communicated by C. Gebhardt.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1037-y) contains supplementary material, which is available to authorized users.

J.-M. Lacape (✉) · N. Forestier-Chiron · J. Jean · E. Thomas · C. Viot
CIRAD, UMR DAP, Avenue Agropolis, 34398 Montpellier Cedex 5, France
e-mail: marc.lacape@cirad.fr

J. Jacobs · R. Derijcker
Bayer BioScience N.V., Technologiepark 38, Ghent, Belgium

T. Arioli
Bayer CropScience, BioScience Research, Lubbock, TX, USA

D. Llewellyn
CSIRO, Plant Industry, Canberra, ACT 2601, Australia

Introduction

Both of the cultivated AD allotetraploid ($2n = 4x = 52$) *Gossypium* species, *G. hirsutum* and *G. barbadense*, were independently domesticated for their epidermal seed trichomes (Wendel et al. 1989) used as textile fibers. Highly productive and agronomically versatile, *G. hirsutum* produces fibers of good quality (with present day cultivars) and comprises more than 90% of world production of cotton. Conversely, *G. barbadense* produces extra long staple (ELS) fibers, which are much longer, stronger and finer, but is much less productive and of narrower agronomic adaptability. Given their complementary agronomic and fiber quality characteristics, numerous attempts have been made to combine their favorable properties by conventional breeding and selection from crosses between these two sexually compatible species. The limited success of interspecific *G. hirsutum* × *G. barbadense* cultivars has been caused in part by reduced fertility, differences in maturity and hybrid breakdown often observed in the later generations.

Cotton molecular genetic research since the early 90s has demonstrated a workable level of molecular polymorphism between *G. hirsutum* and *G. barbadense* genomes in contrast to the relatively low levels within each of these species individually. This has encouraged the development of interspecific *G. hirsutum* × *G. barbadense* framework

mapping populations in several global cotton research centers, including in France at CIRAD from the cross Guazuncho 2 × VH8 (BC_1) (Nguyen et al. 2004), in the USA from crosses race palmeri × K101 (F_2) and TM1 × 3–79 (RIL) (Rong et al. 2004; Frelichowski et al. 2006), and in China from crosses TM1 × Hai7124 (BC_1), CRI36 × Hai7124 (F_2) and Handan 208 × Pima90 (BC_1) (Guo et al. 2007; He et al. 2007; Yu et al. 2007). The main features of the resulting genetic maps are the following: (1) map size typically ranges from 3,500 to 5,700 cM. The relatively large variation in map size is caused by differences in population structure and population size, as well as differences in analysis tools, marker density and the quality of the mapping data (2) PCR-based markers and particularly markers derived from genomic or EST-derived SSR are abundant in most of the maps (3) the various maps have a significant and increasing number of markers in common allowing comparative displays. Several websites offer CMap as a tool for performing map comparisons, including the Cotton Marker Database (CMD) (http://www.cottonmarker.org/), CottonDB (http://cottondb.org/) and TropGeneDB http://tropgenedb.cirad.fr). Most of these framework maps have also served for QTL mapping, mainly focusing on fiber quality QTL (Paterson et al. 2003; Mei et al. 2004; Chee et al. 2005; Lacape et al. 2005; Park et al. 2005; Frelichowski et al. 2006; Wang et al. 2006a; He et al. 2008). Although the various maps show reasonable global consistency, an integrative approach for comparing QTL data from the various parental combinations and/or environments has met with limited success in revealing consistent QTL locations. This is partly due to contradictions in locus orders between the maps at a higher resolution (Lacape et al. 2005; Rong et al. 2007; Wang et al. 2007).

As compared to F_2 and backcross (BC) populations, recombinant inbred lines (RILs) present several advantages for gene and QTL mapping: the high level of homoeozygosity enables identical replication in progeny and effectively renders RILs as perpetual populations. This allows for replicated studies by many investigators and in a multiple environments (Burr et al. 1988). The main advantage of the immortal status of a RIL population is that phenotypes can be estimated much more accurately. Drawbacks of RILs are the long development timelines and, in some cases, reduced fertility in a subset of the lines. Until recently, the use of RILs has been limited in cotton. To date, four RIL populations have served for genetic mapping in cotton, two derived from interspecific *G. hirsutum* × *G. barbadense* crosses (Frelichowski et al. 2006; He et al. 2008) and two from intraspecific (*G. hirsutum*) crosses (Shen et al. 2007; Wang et al. 2007).

Here, we report the development of a new interspecific, *G. hirsutum* × *G. barbadense*, RIL population, which forms the basis of an ongoing international research project aimed at a genetic and genomic dissection of cotton fiber quality (Génoplante project nr ANR-06-GPLA-018). The parents of this new RIL population are highly contrasting for fiber characteristics and were previously used for the development of a BC_1 genetic framework map (Lacape et al. 2003) and an advanced backcross marker-assisted selection program based on the BC_1 materials is underway at CIRAD (Lacape et al. 2005). As a first milestone, this paper reports on the development of a combined SSR-AFLP genetic map for the RIL population, and on the merging of this RIL genetic map with the earlier BC_1 map to create an integrated consensus genetic map permitting numerous bridging points to other published cotton maps.

Materials and methods

Generation of the RIL population

The RIL population comprises 140 lines in an F_6 to F_9 (6 F_6, 19 F_7, 89 F_8 and 26 F_9) stage of selfing through single seed descent (SSD). The two parents, Guazuncho 2 (*G. hirsutum*), and VH8-4602 (*G. barbadense*), were chosen from the germplasm bank of CIRAD for their excellent agronomic performance (Guazuncho 2) and superior fiber quality parameters (VH8) (Lacape et al. 2005). The same *G. hirsutum* and *G. barbadense* parents also served to develop the backcross BC_1 mapping population (Lacape et al. 2003) and further BC_1, BC_2, and BC_2S_1 generations that were used for QTL mapping of fiber quality parameters (Lacape et al. 2005) and leaf hairiness (Lacape and Nguyen 2005).

DNA extraction and marker analysis

DNA was extracted from 140 RILs grown in the CIRAD greenhouse in Montpellier and from 133 RILs grown in the Bayer greenhouse in Belgium. At CIRAD, DNA was extracted from a pooled sample (25 mg) of the first expanded leaf of two to three young seedlings, ground using a MM mixer mill (Qiagen). DNA was isolated using the MATAB (mixed alkylmethylammonium bromide) method as previously reported (Risterucci et al. 2000; Lacape et al. 2003), without further purification steps. At Bayer, DNA was extracted from young leaf tissue from one plant per line (100 mg), following the protocol of Dellaporta et al. (1983) with minor modifications.

Procedures for gel-based SSR detection (CIRAD) were as detailed in Lacape et al. (2007). In total, 451 SSRs were amplified and electrophoresed in 197 different PCR reactions, the majority of the SSRs being tested as 3× PCR multiplexes. The SSRs originate from various libraries

(predominantly BNL, CIR, NAU, MU, and JESPR) that are all described and their sequences posted on the website of the Cotton Marker Database, CMD (http://www.cottonmarker.org). Eighty percent of the SSRs were chosen from the BC_1 map of the Guazuncho × VH8 cross in order to obtain an even coverage of each of the 26 cotton chromosomes. Following this first round of genotyping, additional SSRs were identified from an examination of other international linkage maps to increase the saturation of a few regions of lower coverage.

In total, 597 reliable SSR loci, deriving from 411 different SSRs, were scored for their segregation over 140 RILs. Fragment sizes were estimated from their mobility relative to a 30–330 bp size standard, and for those markers in common with the BC_1, the radiogram of the RILs was aligned with the corresponding BC_1 radiogram, and an identical locus name was adopted for the 2 maps. The majority of SSR loci were codominantly scored, though for a small portion (<5%) of them, only a single parental allele could be traced and was thus dominantly scored.

AFLP analysis for 25 primer pairs (Bayer) was performed essentially as described in the protocol of Vos et al. (1995), except that genomic DNA was first amplified according to the whole genome amplification (WGA) using the Illustra GenomePhi V2 DNA Amplification Kit (GE Healthcare Bioscience, Piscataway, NJ). Pilot experiments showed that original genomic DNA and WGA-amplified DNA gave nearly identical AFLP band patterns. Bands that were not reproducible between genomic DNA and WGA-amplified DNA were excluded from further analysis. In total, 952 AFLP markers were scored using AFLP QuantarPro software (Keygene products N·V.), of which 763 were used in genetic map construction. Based on segregation patterns, around 15% of AFLP loci could be considered allelic and were coded as codominant markers. Part of the primer pairs, 8 out of 25, used in this study were common with the ones used for the BC_1 s, but due to differences in revelation techniques and fragment size estimation (visual in the case of BC_1 s and using QuantarPro in the case of RILs), comparisons of RIL and BC profiles and of the positional correspondence among AFLP loci were not made possible.

Linkage analysis and construction of the RIL map

In a preliminary screen of the complete set of 1,360 loci (597 SSRs and 763 AFLPs), the option "*similarity of loci*" of the software JoinMap® 4.0 (Stam and Van Ooijen 1995; Van Ooijen 2006) was used to identify a series of 238 pairs of similar AFLP loci that were discarded from the analysis. The final SSR + AFLP data set comprises 1,122 loci, 597 SSR and 525 non-redundant AFLPs that were effectively grouped and ordered with JoinMap. Population type was defined as an "*RI8*" (Recombinant Inbred in generation F_8) in JoinMap.

Locus localization derived from other maps was used to examine RIL linkage groups as proposed by JoinMap, i.e., to putatively assign them to chromosomes. The denomination of the 26 chromosomes followed the most recent accepted nomenclature of the 13 A (numbered c1–c13) and 13 D (c14–c26) chromosomes of tetraploid AD cotton, by which the six formerly un-assigned linkage groups are now numbered: A01 as chromosome 13, c13; A02 as c8, A03 as c11, D02 as c21, D03 as c24 and D08 as c19 (Wang et al. 2006b). Despite the use of multiplex PCR and loading of 3 SSRs, the comparison of the RILs and BC_1 profiles was straightforward as the parents were the same. The comparison with the other international maps relied on the existence of several bridge loci per chromosome.

To maintain consistency of loci designation within RIL groups, we had to modify default calculation options in the JoinMap software. When necessary, a conservative high-LOD value (up to 15) was used to break apart large and mixed inconsistent groups into smaller, more tightly linked and consistent ones. Smaller groups known to belong to a given chromosome were then treated as separate data sets and resolved at LOD values between 1 and 4. Manual improvements of the draft map were realized iteratively as new loci were added to the map. When possible, these new loci were used as bridges with other maps and thus alternative locus orders were sometimes tested accordingly. We also used the JoinMap statistical outputs ("*genotype probabilities*" and "*mean square contributions*") as well as graphical genotypes to come back to raw data and check SSR radiograms and AFLP scans for possible coding errors generating singletons. Last, in an advanced version of the map, graphical genotypes were also used to re-check radiograms and gel scans to recode some dominant markers into codominant markers based upon signal intensity (H codings were generated from C or D) and complementary segregation (a pair of AFLP bands, being co-segregating and co-localized on the map, would be associated as a single codominant locus).

Map distances were calculated using Kosambi's mapping function, a goodness-of-fit jump threshold for removal of loci of three, pairwise recombination estimates smaller than 0.50 and LOD scores larger than 0.5.

Map integration

After construction of the RIL linkage map, its integration with the BC_1 map was accomplished with the program JoinMap v. 4.0 (Stam 1993; Van Ooijen 2006). Both BC_1 (1,271 loci, as 5 of the 1,276 loci of the original map were AFLP grouped in an un-assigned linkage group NL5) and RIL (800 loci) datasets were loaded into JoinMap and each

Table 1 Descriptive parameters of the Guazuncho 2 × VH8–RIL map (constructed using JoinMap 4.0) and the Guazuncho 2 × VH8-consensus map obtained by integration of the BC_1 and RIL data using JoinMap 4.0

Chromosome	RIL map			Consensus map		Common BC_1–RIL
	cM	No loci	CO/100 cM	cM	No loci	No loci
c1	97.9	23	2.5	128.3	50	9
c2	81.2	37	2.6	107.9	76	10
c3	76.6	33	2.7	116.5	71	11
c4	69.7	27	2.9	92.1	51	8
c5	108.9	43	3.7	138.3	81	14
c6	48.9	20	3.6	102.9	54	8
c7	74.5	21	2.7	113.0	65	7
c8	66.8	33	3.1	181.0	68	6
c9	73.3	42	4.6	131.2	80	10
c10	67.7	23	2.4	157.9	63	7
c11	66.3	30	4.2	245.2	91	9
c12	75.4	41	4.6	134.9	89	13
c13	72.9	35	3.0	136.8	86	8
Sum A	980.2	408	3.3	1786.0	925	120
c14	98.9	39	3.0	125.0	73	14
c15	71.7	25	2.9	150.7	64	8
c16	58.7	23	2.6	142.0	45	7
c17	45.8	15	3.4	82.2	42	6
c18	99.7	38	3.0	135.6	72	12
c19	75.9	25	3.9	199.8	76	12
c20	100.2	34	2.7	171.6	68	14
c21	167.2	47	2.4	190.0	76	14
c22	72.8	34	2.6	106.0	57	9
c23	60.9	30	3.3	128.5	65	9
c24	83.1	37	3.0	124.0	69	12
c25	53.4	23	2.9	155.5	60	10
c26	75.5	22	2.7	139.8	53	8
Sum D	1063.6	392	2.9	1850.6	820	135
Overall	2043.9	800	3.1	3636.6	1745	255

of the 26 chromosomes assembled separately (26 individual group nodes in each population). The option "*fixed orders*" was occasionally used in a few cases when contradictions between the BC_1 and RIL orders occurred, to impose the best locus order (verified across populations). Map integration was realized for each chromosome using the JoinMap function "*combine groups for map integration*". JoinMap used the data obtained from each population to estimate recombination frequencies that were then used to determine the arrangement of markers by minimizing recombination events (Stam 1993). The recombination frequency between a given pair of markers was calculated after applying the appropriate weighting (LOD-based). Map distances were calculated using function and values as mentioned earlier.

Graphical genotypes, used to check for singletons, were displayed using Excel formatting tools and GGT software (Van Berloo 2008). Comparative mapping and verification of consistency of marker order and distances between maps were done using MapChart (Voorrips 2002).

All genetic maps were finally imported into the web database of CIRAD, TropGeneDB, (http://tropgenedb.cirad.fr/en/cotton.html) for comparative visualization using CMap tool.

Results

Construction of the RIL genetic map

The final genetic map of the RIL population was computed from over 150,000 genotype data points, derived from 1,122 markers in 140 RILs. It comprised 800 loci and spanned 2,044 cM (Table 1, Fig. 1). All chromosomes emerged as single linkage groups that were well saturated

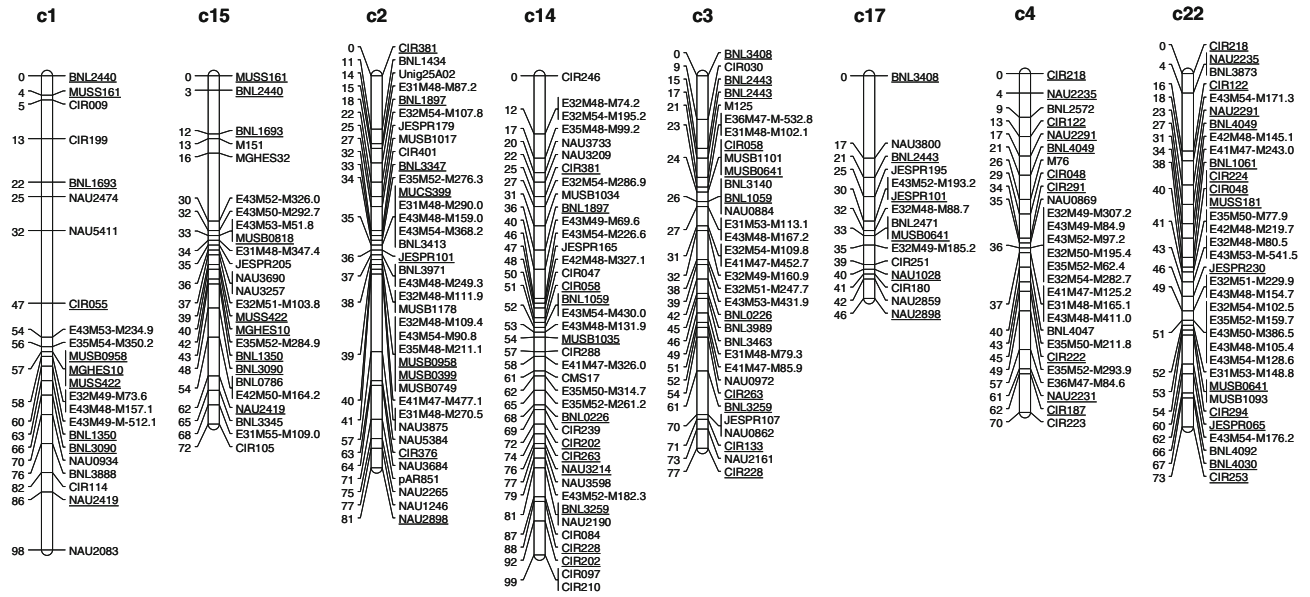

Fig. 1 Genetic map of the Guazuncho 2 × VH8–RIL population. The 26 chromosomes are arranged according to the 13 homoeologous pairs of A (c1–c13) and D (c14–c26) chromosomes, i.e. c1 (At genome) is followed by its homoeologous c15 (Dt genome), etc. The 670 markers include 281 AFLPs named by Eco/Mse primer pair code (see text for primer nomenclature) followed by the fragment size (bp), 387 SSRs of various sources (BNL, CIR, MU, NAU, JESPR, MGHES and SSR-containing RFLPs noted A, pAR, pGH, Gate, Coau and Unig) all described at http://www.cottonmarker.org, and two morphological markers (R2 and pet). Underlined loci are duplicated loci. Map distances are given in centimorgans (cM)

throughout their length, except for chromosome 21 which emerged with two unbridgeable linkage groups. A gap distance of 25 cM was extrapolated for c21 after comparisons with other reference maps. The average distance between 2 loci was 2.55 cM in the RIL map. The map comprised 24 bridgeable gaps of at least 10 cM, of which 7 were larger than 15 cM.

The RIL genetic map contained 281 AFLP loci, 517 SSR loci and two morphological marker loci. The two morphological loci both mapped to the expected chromosomal location: petal spot (*R2*) on c7 and petal color (*Y1* or *pet*) on c13. The 517 SSR loci of the RIL map derived from 387 primer pairs, originating from 5 SSR collections BNL (158), CIR (138), NAU (89), MU (76), and JESPR (36). Marker groupings and loci order on the RIL map were validated by comparative alignments based on shared SSR loci with the CIRAD BC_1 map and other public maps.

The A and D subgenomes (Table 1) contributed very similarly to marker composition (408 and 392 loci, respectively) and map distance (980 and 1,064 cM, respectively).

Segregation distortion and residual heterozygosity

The RIL population displayed broad genetic distortion towards the *G. hirsutum* (Guazuncho) parent. For SSR markers the allelic contribution of both parents could be directly calculated from the observed marker genotypes. For dominantly scored AFLP markers in ambiguous genotypic classes "C" and "D", the allelic contribution of the two parents was calculated by extrapolation from SSR data. On average, any given RIL carried Guazuncho 2 alleles at 70.8% of all loci and VH8 alleles at 29.2% of all loci (total 800 loci). The contribution of Guazuncho 2 and VH8 ranged from 95 and 5 to 32 and 68%, respectively. Only 15 of the 140 RILs contained over 50% of *G. barbadense* alleles. Nearly all loci (99%) deviated significantly from the expected 1:1 parental contribution (χ^2 test at $P < 0.05$, corresponding to 59/41% is met by only 6 of the 800 loci). The parental allele ratio for individual markers ranged from 44/56 to 87/13% (*G. hirsutum/G. barbadense)* and the distortions are evenly distributed across chromosomes, without apparent "hotspots".

The RIL population displayed a higher-than-expected level of residual heterozygosity: 4.9%, instead of the expected 0.8% for an F_8 population derived by SSD. Residual heterozygosity in individual lines ranged from 0.1 to 29.7%. Residual heterozygosity was similar for all chromosomes, ranging from 4.5% (c22) to 6.6% (c6).

Genotypic content and population structure

The frequency and distribution of recombination crossovers (CO) was estimated from examination of the graphical genotypes for all RILs. In case of dominant marker codings (AFLP data), the most conservative interpretation of the coding was used in order to minimize CO estimates. For

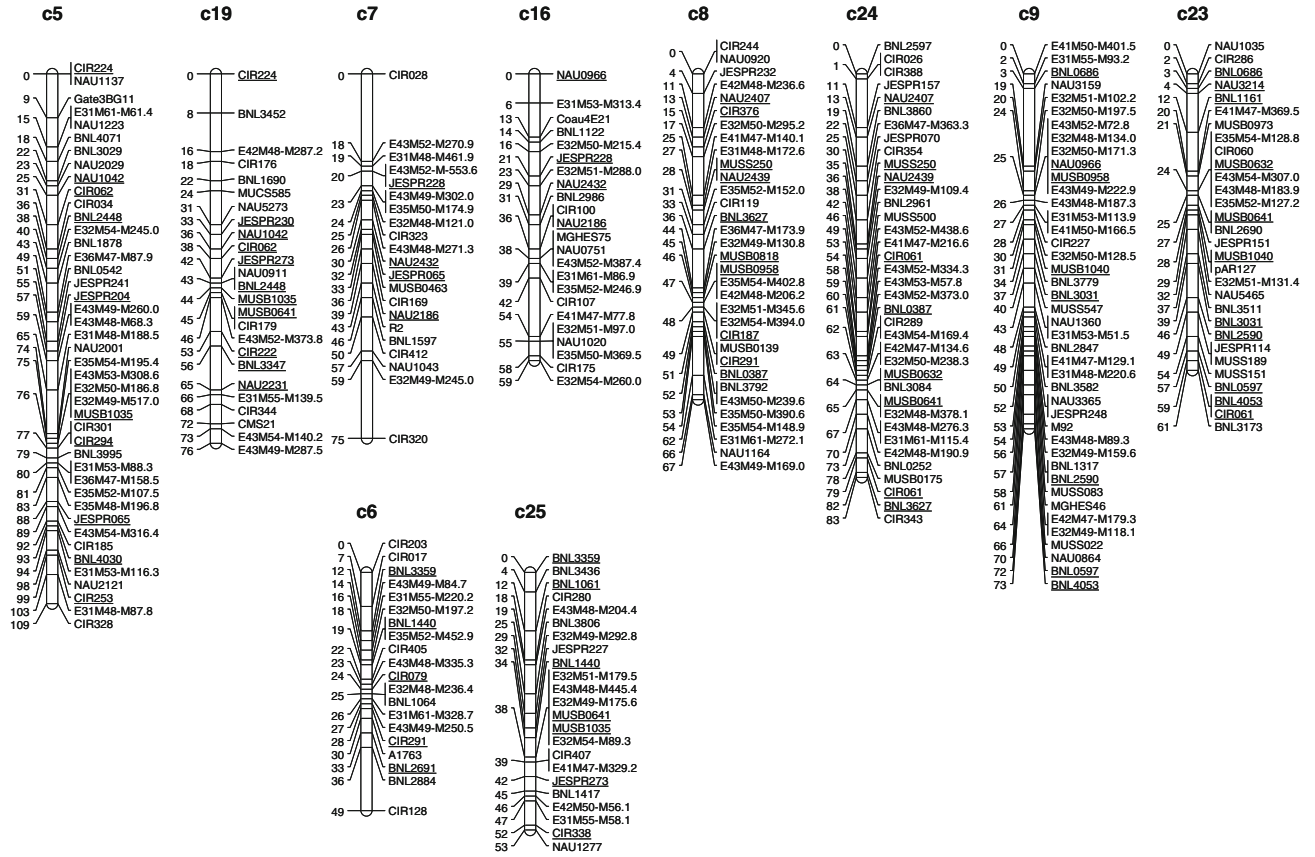

Fig. 1 continued

example, if three successive loci were coded "ADA" or "HDH", the "D" coding would be assumed to be an A and an H, respectively, and no CO would be inferred.

A total of 8,673 CO were counted for 800 loci across all 140 RILs. On average, we counted 62 CO per RIL, representing 2.4 CO per chromosome or 0.031 CO/cM (Table 1). The recombination frequency was fairly stable across most chromosomes, ranging from 0.024 CO/cM on c10 and c21 to 0.046 CO/cM on c9 and c12. Higher values like in c9 and c12, could be indicative of remaining errors in marker coding or map order. On average, any given RIL presented at least one recombination event on 21 of the 26 chromosomes, meaning that on average five chromosomes per RIL were unrecombined, or completely of one single parental origin, with no commonality of the unrecombined chromosomes in different RILs. In the context of trait introgression from *G. barbadense*, it is interesting to note that within the population of RILs it was possible to find most of the 26 *G. barbadense* donor chromosomes in a non-recombined form in at least one RIL. The frequencies of recombination found in the current cotton RIL population are similar to those reported for maize and Arabidopsis RIL populations. In maize, an average of 57 CO per RIL (5.7 CO per chromosome, 0.032 CO/cM) was reported for 91 intermated RILs genotyped with 857 (Fu et al. 2006). In Arabidopsis, 1.13 and 1.86 CO per chromosome or 0.016 and 0.022 CO/cM were reported from two different RIL populations (420 and 423 lines) by Loudet et al. (2002) and Törjék et al. (2006) respectively.

Comparison of the RIL and BC_1 maps from the same parents

Marker order in the current RIL map was very consistent with the marker order of the previously published BC_1 map from the same parents (Nguyen et al. 2004). In contrast, average marker distance and total map size were markedly different between the two maps, with the RIL map being much smaller. One reason for the size differences could be the mapping algorithms used. The BC_1 map (1,275 loci, 5,735 cM) was constructed with MapMaker 3.0, while the RIL map was constructed using JoinMap 4.0. MapMaker and JoinMap use different algorithms and, as a consequence, are known to generate differences in distances, typically shorter distances with JoinMap as compared to MapMaker (Bradeen et al. 2001). The distances of the BC_1 map were recalculated using JoinMap 4.0 with the same mapping parameters (see "Material and methods") as used

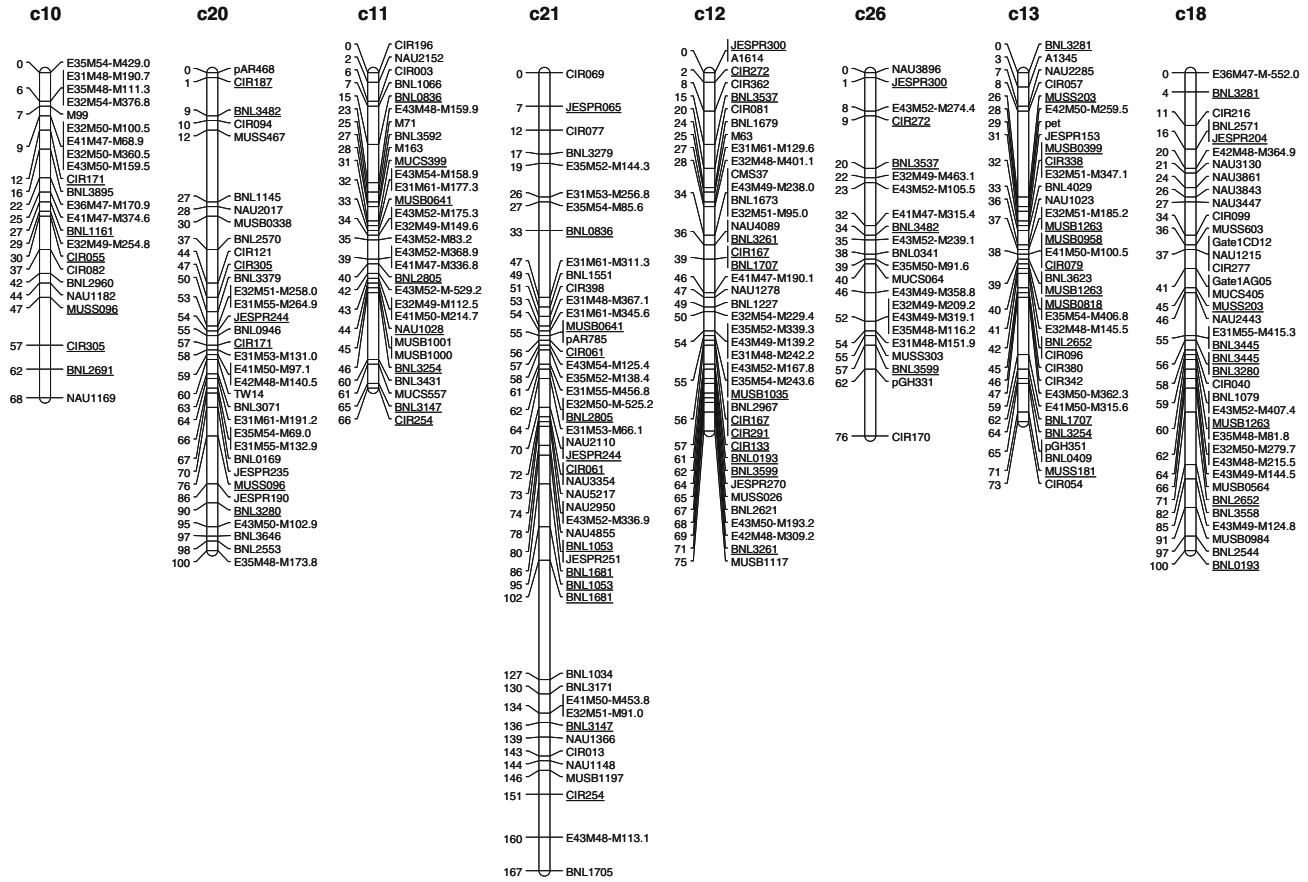

Fig. 1 continued

for the RIL map. This confirmed that map distances generated by JoinMap 4.0 were shorter, by 28% in our case, as compared to the same data processed with MapMaker 3.0, resulting in a total adjusted size of the BC_1 map of 4,104 cM. The comparison of the RIL map with the adjusted BC_1 map indicated that around 16% (670 cM) of the BC_1 was not covered in the RIL map. This difference related to the existence in the BC_1 map of more markers and of markers of different marker systems (RFLP) as compared to the RIL map. The total length of syntenic segments (1,910 of 2,044 cM in the RIL map corresponding to 3,433 of 4,104 cM in the BC_1 map) indicated that the RIL map only represented 56% of the length of the syntenic region of the BC_1 map. Except for chromosomes 5, 13, 21, and 25 which were not significantly different in length between BC_1 and RIL map, the size contraction was fairly similar across all chromosomes (not shown).

Integration of the BC_1 and RIL maps

The nearly perfect conservation of marker order between the RIL and BC_1 maps facilitated their integration into a consensus map based on the combined raw data, using the command "*combine groups for map integration*" of the JoinMap software. The consensus map (Table 1) was built from a total of 1,745 loci, including 255 SSR loci shared between the two maps and 988 and 502 loci unique to the BC_1 and RIL maps, respectively. The number of bridging loci per chromosome (Table 1) was on average ten and ranged from six (c8 and c17) to 14 (c5, c14, c20 and c21). Figure 2 displays a simplified view of the colinearity of locus order between the two component maps (BC_1 and RIL) and the consensus map for three chromosomes c1, c2, and c3 (complete set of 26 chromosomes shown in supplementary Fig. A1).

The 1,745 loci of the consensus map consisted of 781 SSR, 715 AFLP, 190 RFLP, 2 morphological markers, and 57 other markers. 28 BC_1 and 43 RIL loci were excluded by JoinMap and not integrated into the consensus map (Fig. A2). On average, each chromosome contained 67 loci, and the consensus map spanned a total distance of 3,637 cM, with an average distance of 2.1 cM between two loci. Thus, the map size and marker distances in the consensus map are intermediate between the distances covered in the BC_1 and RIL maps. The 3,637 cM of the consensus are similar to the 3,540 cM of a JoinMap-based SSR-rich interspecific

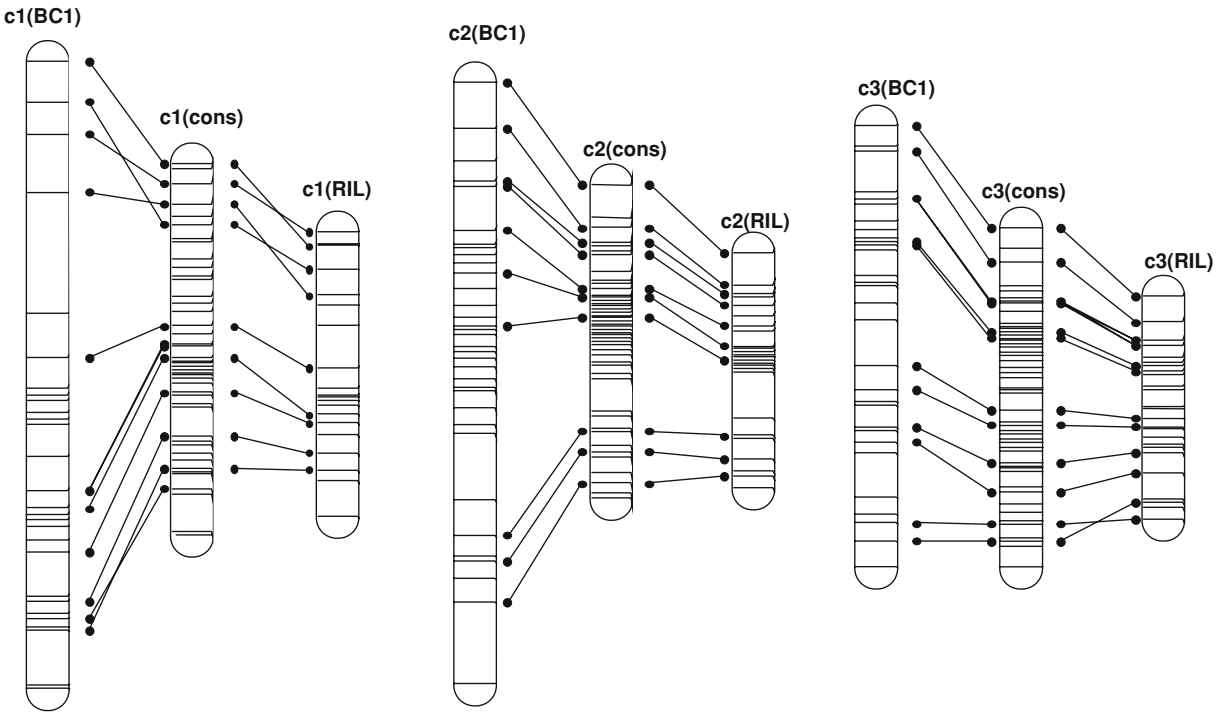

Fig. 2 Simplified representation of the BC_1 and RIL maps originating from the Guazuncho 2 × VH8 cross aligned to an integrated consensus map (cons). RIL and consensus maps were produced using JoinMap. For consistency with published data, the BC_1 map is unchanged from earlier reports (Nguyen et al. 2004), i.e., obtained with MapMaker 3.0. Three chromosomes (c1, c2 and c3) are shown; complete series of 26 chromosomes available from supplementary Fig. A1. Presence of a locus is denoted by a horizontal line within the chromosome bar and only loci in common between BC_1 and RIL maps are connected

genetic map recently reported by Guo et al. (2007) from 1,790 loci and 138 BC_1 progeny.

Locus order is globally well conserved between the consensus and component maps, with a high overall consistency for the majority of the 26 chromosomes (Fig. 2, Fig. A1 and A2). However, a few notable differences in locus order were observed:

c4. several loci in the bottom region of the BC_1 map were not integrated into the consensus map because of loose linkages (these include pAR574, CIR223, CIR142),
c5. an important shrinkage in the corresponding top regions of the BC_1 and RIL maps was caused by uncertainties in the BC_1 map (2 loci not integrated into the consensus map, NAU1223 and NAU1042), and in the RIL map (5 SSR loci not integrated, NAU1137, Gate3BG11, NAU1223, BNL4071, BNL3029, and NAU1042),
c6. three loci of the bottom region of the BC_1 map were not integrated into the consensus map because of loose linkages (these include 2 SSRs, CIR128, CIR179, and the RFLP, M16-147),
c7. the stretching in the top region of the BC_1 map was caused by the wrong position of the top distal locus CIR028 correctly relocated in the consensus map,
c8. 20 loci of the RIL map were not integrated into the consensus map by JoinMap, including 16 AFLPs and 4 SSRs (NAU0920, NAU2407, NAU0250, NAU2439),
c9. the stretching in the top region of the BC_1 map was caused by three AFLP loci and two SSRs (BNL1423 and BNL1162); an overall re-arrangement in the consensus map was proposed in agreement with other published maps,
c21. four SSR loci of the RIL map were not integrated into the consensus map by JoinMap (BNL1034, NAU1366, NAU1148, and MUSB197); length of consensus c21 may be overestimated because of the existence of an extrapolated 25 cM gap in the RIL map,
c23. two loci (CIR198 and CIR019) positioned 42 cM apart in the top region of the BC_1 map were repositioned in the consensus map in agreement with other published maps,
c25. the RIL map was lacking 1/3rd in the bottom region because of a lack of scored markers; as a consequence length of consensus c25 may be overestimated.

Variation of locus density along chromosomes was assessed by counting loci in 5-cM intervals and visualized in Fig. A3. The distribution of loci along chromosomes was clearly not random. A single region of high marker density

flanked by regions of lower density was observed on each of the 26 chromosomes except for c9, c10, c16, and c19 which displayed more uniform marker distribution.

Discussion

Segregation distortion

The cotton RIL population described here displays severe, genome-wide allelic distortion towards the *G. hirsutum* parent (71% *G. hirsutum* alleles and 29% *G. barbadense* alleles). Biased transmission of parental alleles is a common feature of mosaic interspecific plant populations (Zamir and Tadmor 1986). In cotton, a comparable biased transmission was observed in interspecific (*G. hirsutum* × *G. barbadense*) BC_1 (Jiang et al. 2000) and RIL populations (He et al. 2008 and Mauricio Ulloa, personal communication), with in all cases *G. hirsutum* alleles being favored. The most probable explanation for the observed distortion relates to an (unintended) selection bias towards the *G. hirsutum* parent during the SSD process. It is known that interspecific *Gh* × *Gb* progenies suffer from fertility breakdown, and this was also true for the current RIL population. Moreover, the *G. barbadense* parent used in this study has some less favorable reproductive features, such as lateness in flowering and reduced seed numbers per boll as compared to the *G. hirsutum* parent. We assume that our attempts to produce progeny seed within a reasonable time frame has resulted in inadvertent selection against individuals bearing *G. barbadense* alleles that negatively influence flowering time and fruiting properties. Distorted segregation at a genome-wide scale has also been reported for interspecific RIL populations in a number of other species, including chickpea (Cobos et al. 2006), tomato (De Vicente and Tanksley 1993; Paran et al. 1995) and rice (Wang et al. 1994; Lee et al. 2005; Jie et al. 2006). In all these cases, the percent of skewed loci ranged from 54 to 80% and displayed a scattered distribution throughout the genome. The segregation distortion, however, should not affect the utility of this population for mapping QTLs for significant agronomic and fiber quality traits observed in the RILs.

Map distances in the RIL map

Despite the marker distortion it was possible to develop a coherent genetic map from the genotype data of 140 RILs. The RIL map provided good coverage of all 26 cotton chromosomes. However, after eliminating the effect of the mapping software, we observed that the map size and marker distances of the RIL map were markedly reduced (56%) compared with the size and distances of syntenic regions in the published BC_1 map from the same cross. The size reduction, in theory, should not relate to the difference in population type (BC_1 vs. RIL) and consequential difference in inferred numbers of meioses, because the mapping software compensates for these effects. We have identified two factors that may have contributed to the observed differences in map distances: (1) different rates of missing data and coding errors between the BC_1 and RIL data sets and (2) the allelic imbalance in the RIL data. Hackett and Broadfoot (2003) have shown that coding errors increase map size estimates, while missing data reduce map size estimates. AFLP codings are generally thought to be more prone to coding errors (false positives, homoplasy) than other marker systems. Since the BC_1 map contained more AFLP loci than the RIL map (465 vs. 281) and since AFLP markers for the RIL map were more stringently selected, we expect that more coding errors were included in the BC_1 map, which may have resulted in overestimation of map distances in this map. Conversely, the presence of more missing data in the RIL data set (6.0%) compared to the BC_1 data set (2.2%), may have resulted in an underestimation of map distances in the RIL map. The RIL population displayed strong distortion towards the *G. hirsutum* parent, while the BC_1 population was more balanced. The bias towards one parent may have lowered the number of "effective" (detectable) recombinations in the RIL population, while this would not be the case in the BC_1 population. Differences in map distances between various maps from the same species (even with identical population types), as well as a strong influence of the mapping software were previously reported (Qi et al. 1996; He et al. 2001; Armstead et al. 2002; Knox and Ellis 2002). In cotton, the reported sizes of the seven most saturated interspecific genetic maps range from 3,500 to 5,300 cM or even more (Nguyen et al. 2004; He et al. 2007; Guo et al. 2008; Zhang et al. 2008). Thus, the current RIL map and Guazuncho 2 × VH8 consensus map are positioned at the lower end of cotton map sizes. We believe, together with Wang et al. (1994), that map distances can be expected to fluctuate among crosses, while marker assignation and order along chromosomes are expected to be repeatable between studies.

The Guazuncho 2 × VH8 consensus map

Comparative analysis of genetic maps and map-associated information (such as QTL data) is a complex and laborious process, especially if data originating from different laboratories must be compared. In principle, two approaches may be considered: (1) map comparison based on final individual map versions, using one selected map as a stable reference (Truco et al. 2007), or (2) a true integration of the raw data and re-computation of an integrated map (through a software package that uses the base data). Two recent

Table 2 Distribution of shared loci between 8 whole genome cotton genetic maps: Guazuncho × VH8-consensus (GV-cons) and its two component (GV–BC_1 and GV–RIL) maps, and five interspecific *G. hirsutum* × *G. barbadense* maps, CH–F_2, HP–F_2, T3–RIL, TH–BC_1, and PK–F_2

Population[a]	No loci[b]	GV-cons	GV–BC_1	GV–RIL	CH–F_2	HP–F_2	T3–RIL	TH–BC_1	PK–F_2
GV-cons	1,745								
GV-BC_1	1,271	1,226							
GV-RIL	800	546	255						
CH-F_2	1,097	296	260	184					
HP-F_2	1,029	243	235	136	195				
T3-RIL	433	116	75	83	89	60			
TH-BC_1	1,790	313	249	205	236	180	85		
PK-F_2	2,584	203	202	50	68	68	28	64	
Total common		657	–	–	399	302	162	400	221

[a] Populations compared are: Guazuncho 2 × VH8-consensus (GV-cons) in this article; Guazuncho 2 × VH8–BC_1 (GV–BC_1) reported by Nguyen et al. (2004); Guazuncho 2 × VH8–RIL (GV–RIL) in this article; CRI36 × Hai7124–F_2 (CH–F_2) by Yu et al. (2007); Handan 208 × Pima 90–F_2 (HP-F_2) by He et al. (2007); TM1 × 3-79–RIL (T3–RIL) by Frelichowski et al. (2006); TM1 × Hai 7124–BC_1 (TH–BC_1) by Guo et al. (2007); palmeri × K101 F_2 (PK–F_2) by Rong et al. (2004)

[b] The number of loci is the one reported in referenced publication

A locus is common between two maps when it is (1) derived from the same marker and (2) mapped on the same chromosome on two maps (only one locus considered per chromosome/linkage group in case of paralogous duplication on the same LG)

publications reported on the integration of cotton maps using the first approach (Rong et al. 2007; Xu et al. 2008). In both cases, the race palmeri × K101-F_2 map of Rong et al. (2004) was considered as a basis for performing either meta-analysis of QTLs from different experiments (Rong et al. 2007) or for the integration of genetic and physical maps of chromosomes 12 and 26 (Xu et al. 2008). In this paper we describe map integration based on raw data. We show that, with exception of a few minor inconsistencies in locus order, the consensus map faithfully represents the locus order of the two component maps. We believe that this approach will help in providing a stronger basis for future comparative mapping and meta-QTL analysis.

Compared with the component maps, the Guazuncho × VH8 consensus map has improved genome coverage at two levels: (1) higher marker density (one locus per 2 cM) and (2) reduced number of gaps (13 gaps exceed 15 cM of which 5 span 20–25 cM). The distribution of loci along individual chromosomes is still unequal in the consensus map. A single region of higher locus density is typically observed on 20 of 26 chromosomes (Fig. A3). AFLP markers are slightly over-represented in these locus-dense regions (56 vs. 41% overall). Clustering of *Eco*RI/*Mse*I AFLP markers is expected to occur preferentially in centromeric regions (Young et al. 1999). Thus, the marker-dense regions could indicate centromeres.

Alignment of the Guazuncho × VH8 consensus map with other genetic maps provided additional evidence (more duplicated loci) to substantiate the 13 known homoeologous A/D relationships of tetraploid cotton (Fig. A2), bridged by 104 pairs of homoeologous duplications (86 SSR and 18 RFLP), and also confirmed the known reciprocal translocations involving the two pairs of A chromosomes, c2 and c3 (12 duplicated loci), and c4 and c5 (6 loci).

The consensus map as a community resource

The Guazuncho × VH8 consensus map reported here is of particular interest to the cotton research community due to the fact that it contains the highest number of "common" markers when compared with other public cotton maps (Table 2 and CMap display from http://tropgenedb.cirad.fr/en/cotton.html).

Common markers were calculated for six genome-wide genetic maps of cotton, containing (after exclusion of paralogous duplications) 7,397 genetic loci. The following six maps were included:

- Guazuncho 2 × VH8-consensus (GV-cons) and its two component maps, GV–BC_1 and GV–RIL,
- CRI36 × Hai7124–F_2 (CH–F_2) reported by Yu et al. (2007),
- Handan 208 × Pima 90–F_2 (HP-F_2) reported by He et al. (2006),
- TM1 × 3-79–RIL (T3–RIL) reported by Frelichowski et al. (2007),
- TM1 × Hai 7124–BC_1 (TH–BC_1) reported by Guo et al. (2004),
- Race palmeri × K101–F_2 (PK–F_2) reported by Rong et al. (2004).

The 7,397 loci identified on the six maps represent 6,046 different loci: 5,254 are unique to one specific map and 792 (13%) are shared between two and six maps (6 loci shared

between all 6 maps, 49 between 5 maps, 104 between 4 maps, 148 between 3 maps, and 477 between 2 maps). Except for the comparison of PK–F_2 with GV-cons which is partially based on RFLP markers, all other cross-population comparisons rely exclusively on SSR loci. SSR markers from the BNL collection are shared between the highest numbers of maps. This is easy to understand, as this was the first SSR library developed in cotton. The GV-consensus map (by far) contains the highest numbers of bridge loci in a single map (Table 2), with 657 loci (38% of its loci) shared with at least one of the five other framework maps. This establishes the GV-consensus map as an excellent tool for future map integration.

Acknowledgments The French National Research Agency, ANR, through its program specialized in plant genomics, Génoplante, has sponsored this research (project nr ANR-06-GPLA-018) that was otherwise supported by Bayer CS and CIRAD funding.

References

Armstead I, Turner L, King I, Cairns A, Humphreys M (2002) Comparison and integration of genetic maps generated from F2 and BC1-type mapping populations in perennial ryegrass. Plant Breeding 121:501–507

Bradeen JM, Staub JE, Wye C, Antonise R, Peleman J (2001) Towards an expanded and integrated linkage map of cucumber (*Cucumis sativus* L.). Genome 44:111–119

Burr B, Burr F, Thompson KH, Alberston MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Chee P, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. Theor Appl Genet 111:757–763

Cobos MJ, Rubio J, Strange RN, Moreno MT, Gil J, Millan T (2006) A new QTL for Ascochyta blight resistance in an RIL population derived from an interspecific cross in chickpea. Euphytica 149:105–111

De Vicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:586–596

Dellaporta SL, Wood J, Hicks JN (1983) A plant DNA minipreparation: Version II. Plant Mol Biol Rep 1(4):19–21

Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M (2006) Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol Genet Genomics 275:479–491

Fu Y, Wen TJ, Ronin YI, Chen HD, Guo L, Mester DI, Yang Y, Lee M, Korol AB, Ashlock DA, Schnable PS (2006) Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics 174:1671–1683

Guo W, Cai C, Wang C, Han ZG, Song XL, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007) A microsatellite-based, gene-enriched 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 (2008) A preliminary analysis of genome structure and composition in *Gossypium hirsutum*. BMC Genomics 9:314

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity 90:33–38

He P, Li JZ, Zheng XW, Shen lS, Lu CF, Chen Y, Zhu LH (2001) Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. Crop Sci 41:1240–1246

He DH, Lin Z, Zhang X, Nie YC, Guo XP, Zhang Y, 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 153:181–197

He DH, Lin ZX, Zhang XL, Zhang YX, Li W, Nie YC, Guo XP (2008) Dissection of genetic variance of fibre quality in advanced generation from an interspecific cross of *Gossypium hirsutum* and *G. barbadense*. Plant Breed 127:286–294

Jiang C, Chee P, Draye X, Morrell P, Smith CW, Paterson AH (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (Cotton). Evolution 54(3):798–814

Jie C, Bughio H, Chen DZ, Liu G, Zheng K, Zhuang J (2006) Development of chromosomal segment substitution lines from a backcross recombinant inbred population of interspecific rice cross. Rice Sci 13:15–21

Knox MR, Ellis THN (2002) Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. Genetics 162:861–873

Lacape J-M, Nguyen TB (2005) Mapping quantitative trait loci associated with leaf and stem pubescence in cotton. J Hered 96(4):441–444

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

Lacape J-M, Nguyen TB, Courtois B, Belot J-L, Giband M, Gourlot J-P, Gawryziak G, Roques S, Hau B (2005) QTL analysis of cotton fiber quality using multiple *G. hirsutum* × *G. barbadense* backcross generations. Crop Sci 45:123–140

Lacape J-M, Dessauw D, Rajab M, Noyer JL, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* gene pool: assembling a highly informative genotyping set of cotton SSRs. Mol Breed 19:45–58

Lee SJ, Oh CS, Suh JP, McCouch S, Ahn SN (2005) Identification of QTLs for domestication-related and agronomic traits in an *Oryza sativa* × *O. rufipugon* BC_1F_7 population. Plant Breed 124:20–219

Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-O × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. Theor Appl Genet 104:1173–1184

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

Nguyen TB, Giband M, Brottier P, Risterucci A-M, Lacape J-M (2004) Wide coverage of tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Genet 109:167–175

Paran I, Goldman S, Tanksley SD, Zamir D (1995) Recombination inbred lines for genetic mapping in tomato. Theor Appl Genet 90:542–548

Park Y-H, Alabady MS, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG, Ulloa M (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line (RIL) cotton population. Mol Genet Genomics 274:428–441

Paterson AH, Saranga Y, Menz M, Jiang C, Wright RJ (2003) QTL analysis of genotype × environment interactions affecting cotton fiber quality. Theor Appl Genet 106:384–396

Qi XP, Stam P, Lindhout P (1996) Comparison and integration of four barley linkage maps. Genome 39:379–394

Risterucci A-M, Grivet L, N'Goran JAK, Pieretti I, Flament MH, Lanaud C (2000) A high density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee P, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C-h, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder N, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (2004) A 3,347 locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee P, Draye X, Saranga Y, Wright RJ, Wilkins TA, May LO, Smith CW, Gannaway JR, Wendel JF, Paterson AH (2007) Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577–2588

Shen X, Guo W, Lu Q, Zhu X, Yuan Y, Zhang T (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155:371–380

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Stam P, Van Ooijen JW (1995) JOINMAP version 2.0: software for the calculation of genetic linkage maps

Törjék O, Witucka-Wall H, Meyer RC, von Korff M, Kusterer B, Rautengarten C, Altmann T (2006) Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theor Appl Genet 113:1551–1561

Truco MJ, Antonise R, Lavelle D, Ochoa O, Kosik A, Witsenboer H, Fort SB, Jeuken MJW, Kesseli RV, Lindhout P, Michelmore RW, Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). Theor Appl Genet 115:735–746

Van Berloo R (2008) GGT 2.0: versatile software for visualization and analysis of genetic data. J Hered 99:232–236

Van Ooijen JW (2006) JoinMap® 4.0, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V. Wageningen, Netherlands

Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77–78

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23(21):4407–4414

Wang G-L, Mackill DJ, Bonman JM, McCouch S, Champoux MC, Nelson RJ (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durable resistant rice cultivar. Genetics 136:1421–1434

Wang B-H, Guo W, Zhu X, Wu Y, Huang N, Zhang T (2006a) QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton. Euphytica 152:367–378

Wang K, Song XL, Han ZG, Guo W, Yu J, Sun J, Pan J, Kohel RJ, Zhang TZ (2006b) Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybrisation mapping. Theor Appl Genet 113:73–80

Wang B, Wu Y, Guo W, Zhu X, Huang N, Zhang T (2007) QTL analysis and epistasis effects dissection of fiber qualities in an elite cotton hybrid grown in second generation. Crop Sci 47:1384–1392

Wendel JF, Olson PD, Stewart JM (1989) Genetic diversity, introgression, and independent domestication of old world cultivated cottons. Am J Bot 76(12):1795–1806

Xu Z, Kohel RJ, Song G, Cho J, Yu J, Yu S, Tomkins JP, Yu J (2008) An integrated genetic and physical map of homologous chromosomes 12 and 26 in Upland cotton (*G hirsutum* L.). BMC Genomics

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. Theor Appl Genet 99:785–790

Yu J, Yu S, Lu C, Wang W, Fan S, Song M, Lin Z, Zhang X, Zhang J (2007) High-density linkage map of cultivated allotetraploid cotton based on SSR, TRAP, SRAP and AFLP markers. J Integr Plant Biol 49(5):716–724

Zamir D, Tadmor Y (1986) Unequal segregation of nuclear genes in plants. Bot Gaz 147:355–358

Zhang YX, Lin ZX, Xia QZ, Zhang MJ, Zhang XL (2008) Characteristics and analysis of simple sequence repeats in the cotton genome based on a linkage map constructed from a BC1 population between *Gossypium hirsutum* and *G. barbadense*. Genome 51:534–546