TARGETED INTROGRESSION OF COTTON FIBER QUALITY QTLs USING MOLECULAR MARKERS

J.-M. Lacape, T.-B. Nguyen, B. Hau, and M. Giband
CIRAD-CA, Programme Coton, TA 70/03, Avenue Agropolis, 34398 Montpellier Cedex 5, France (Author for correspondence: J.-M. Lacape : lacape@cirad.fr)

Summary
Within the framework of our cotton breeding program, we are using molecular markers to improve the efficiency of the introgression of fiber quality traits into a favorable genetic background. This effort relies on i) the development of a saturated genetic map of the cotton genome, ii) the identification of fiber quality QTLs, and iii) the marker-assisted introgression of favorable genomic regions into an adequate genetic background.

Introduction

Upland cotton, *Gossypium hirsutum*, dominates the world's cotton fiber production, accounting for approximately 90% of the total world production. The second most cultivated species, *G. barbadense*, includes superior extra-long, strong and fine cottons. Attempts in utilizing deliberate interspecific *G. hirsutum*/*G. barbadense* recombination by conventional breeding have had a limited impact on cultivar development. Although the use of DNA markers for marker-assisted selection (MAS) has received considerable attention among plant and animal breeders in the past 10-15 years, the application of this technology in breeding programs is still scarce. In particular, efforts devoted to cotton and dedicated at a better understanding of the molecular bases of cotton fiber quality have lagged behind those devoted to other crop species. It is only recently that reports on the identification of fiber quality-related QTLs based on the study of interspecific *G. hirsutum* X *G. barbadense* populations have been published (Jiang *et al*., 1998; Kohel *et al*., 2001; Paterson *et al*., 2003). Nevertheless, the implementation of MAS in cotton has not been reported to date.
Within the framework of a marker-assisted backcross introgression scheme aimed at transferring fiber quality traits from *G. barbadense* (donor variety) into *G. hirsutum* (recipient background), we first developed a saturated genetic map of tetraploid cotton (Lacape *et al.*, 2003). In this report, we describe the use of molecular markers to identify and select for QTL-rich regions involved in determining fiber quality.

Material and methods

The major milestones in the marker-assisted selection process include the construction of 2 genetic maps from the BC1 and BC2 populations, the detection of fiber quality QTLs from 3 phenotyping data sets, and the actual marker-based selection in the BC3 and BC4 generations.
The initial interspecific cross involved the *G. hirsutum* variety Guazuncho 2 and the *G. barbadense* variety VH8-4602. Guazuncho 2, a modern pure-line *G. hirsutum* variety created in Argentina was chosen as a recipient genome in the backcross generations for its overall good agronomic performance. VH8-4602, a *G. barbadense* variety of Sea Island type, was the donor parent for superior fiber quality, in particular for length, strength, and fineness; conversely its fiber color indexes are of low value (Table 1).
The BC1 (75 individuals) and BC2 (200 individuals) maps were constructed separately using the MapMaker 3.0 software before being combined to constitute a consensus framework map comprising RFLP, AFLP and SSR markers. The consensus map then served for the QTL analysis of fiber quality components by composite interval mapping (CIM) using the QTL

Cartographer software. The fiber quality measurements (11 traits) were conducted on an HVI (High Volume Instrument) line at Cirad/Montpellier.
Details on the plant material used and the types of analyses undertaken during the different steps of the MAS process are given in Figure 1.

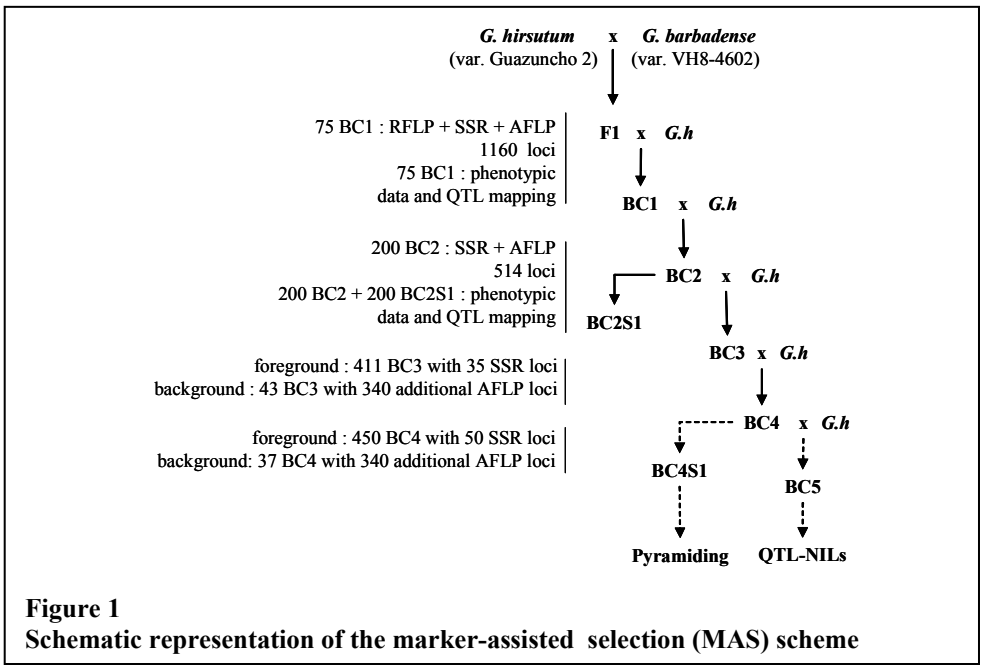

Figure 1
Schematic representation of the marker-assisted selection (MAS) scheme

Results

Genetic mapping

The first development of our program was the construction of genetic maps of tetraploid cotton combining RFLP, SSR, and AFLP markers, generated separately for the first two backcross generations (BC1 and BC2). The initial BC1 map, comprising 888 loci grouped in 37 linkage groups, and spanning 4400 cM (Lacape *et al*., 2003), benefited from the development and integration of new additional microsatellite markers (Nguyen *et al.*, submitted). This updated saturated BC1 map now spans 5500 cM and comprises a total of 1160 loci ordered along 26 chromosomes or linkage groups. On the other hand, the BC2 map (512 loci in total) constructed using AFLP and SSR markers had 393 loci in common with the BC1 map. The two maps agreed for loci order, thus allowing their merging into a combined map. This new consensus BC1/BC2 map then served for 3 separate QTL analyses (Lacape *et al.*, in preparation) of fiber quality components and as a support for the MAS program.

QTL detection

Three separate QTL analyses were conducted by composite interval mapping using 2 molecular data sets (BC1 and BC2) and 3 sets of fiber measurements carried on a single plant basis for the BC1 and BC2 populations, and on a per-line basis (with 2 replicates) for the BC2S1 population. The fiber measurements, initially including 11 traits, were reduced into 6 groups of strongly correlated traits after analyzing the correlation between traits. The fiber characteristics that were retained for measurement include length, (length) uniformity, strength, elongation, fineness or maturity, and color.
For the 6 fiber quality components studied, we globally identified 50 QTLs that met permutation-based LOD thresholds (LOD scores between 3.2 and 4 for most of the traits). Though suggestive, thirty additional QTLs having a LOD value above 2.5 were also taken

into consideration after comparing the results between our 3 populations, or between our results and those reported in the literature (Jiang *et al*., 1998; Kohel *et al*., 2001; Paterson *et al*., 2003). Table 1 summarizes the data generated from the QTL analysis for the 6 traits of interest, and the phenotypical effects of the detected QTLs. In general, the contribution of each QTL, measured as a percentage of explained variation of a given trait, was variable and in most cases fairly low. For example, for traits of economical importance, individual contributions varied from 4.8 to 14.8% in the case of fiber length, 4.4 to 21.3% for fiber strength, and 4.6 to 29.1% for color reflectance. Overall, we observed that these 80 QTLs partitioned as expected from the phenotypic values of the *G. hirsutum* and *G. barbadense* parents: a majority of positive alleles for length (12 of the 15 QTLs), strength (8 of the 12 QTLs), and fineness (13 of the 21 QTLs) derived from the *G. barbadense* parent, while a majority of positive alleles for fiber color (13 of the 16 QTLs) derived from the *G. hirsutum* parent (Table 1). Furthermore, the QTLs detected for the various traits often co-localized within QTL-rich regions (see Table 2). In some cases, QTL detection and mapping were in agreement between generations (BC1 and BC2), and interestingly, in 25 cases, they confirmed the results reported in the literature, both for the position of a QTL and for the sign of its phenotypic effect.

Table 1
Mean parental values (average of the 3 sets of data) of fiber technological parameters. The number of QTLs for each trait and range of observed positive (*Gb* +) and negative (*Gb* -) phenotypic effects conferred by the *G. barbadense* alleles, detected over the 3 populations (BC1, BC2 and BC2S1), are given.

	Gb **mean**	***Gh*** **mean**	**QTL** ***Gb*** **+**	**Range Phenotypic effects**	**QTL** ***Gb*** **-**	**Range Phenotypic effects**
Length (mm)*	39,2	29,5	12	+0.7 to +2.1	3	-1.6 to -1.8
Uniformity	85,4	83,3	3	+0.5 to +1.5	3	-1.1 to -3.3
Strength (g/tex)	44,7	28,8	8	+0.8 to +2.8	4	-0.9 to -3.4
Elongation	5,7	5,8	6	+0.2 to +0.5	4	-0.3 to -0.6
Fineness (mtex)**	184,1	222,4	13	-10 to -20	8	+9 to +40
Color reflectance	75,2	74,7	3	+1.8 to +2.5	13	-0.9 to -3.5
Total			**45**		**35**	

* UHML, ** standard fineness

The chromosome regions carrying QTLs (corresponding to a single or to several traits) whose positive alleles derived from the *G. barbadense* donor genome were reduced to a number of 19, which were carried by 15 different chromosomes (Table 2). Altogether, the confidence intervals (1 LOD) of the involved QTLs delimit a total length of 636 cM (20% of the carrier genome) (Table 2), or 11.5 % of the total genome.

Table 2
Identification of the 19 targeted regions mapped on 15 different chromosomes, and comprising 1 or several co-localized fiber quality QTLs from *G. barbadense* for possible introgression into a *G. hirsutum* genetic background. All targeted QTLs show a positive contribution from the *G. barbadense* allele, except for a few negative cases indicated in brackets. The target region is defined as situated between the 2 loci flanking, at a 1 LOD confidence interval, the QTL peak LOD value

Chromosome	Chromosome length (cM)	Target interval (cM)	Target size (cM)	Trait
c14	197	28-57	29	Length
c3	153	32-67	35	Length, fineness
		90-138	48	Length, strength, fineness
c4	190	102-118	16	Fineness
c22	139	112-139	27	Fineness
c5	360	78-101	23	Strength
c6	296	137-144	7	Length, fineness
c25	183	44-73	29	Length, strength
c16	168	65-117	52	Strength, fineness, color
c23	173	45-66	21	Strength (elong -, color -)
		113-135	22	Length, strength
c10	192	0-21	21	Fineness
		78-120	42	Length, fineness, color
c20	268	88-161	73	Elongation, fineness
c26	195	67-143	76	Length (color -)
A01	233	16-54	38	Length
		171-209	38	Strength
c18	158	32-46	14	Fineness
A03	271	209-234	25	Strength, uniformity
Total carrier chromosomes	**3176**	**Total target regions**	**636**	

Marker-assisted selection in the BC3 and BC4 generations

The molecular markers, essentially SSRs, that framed the 19 targeted regions of interest (Table 2) were used for the genotyping of BC3 and BC4 plants (Fig.1) at the seedling stage. This early marker-based selection allowed to choose those plants that showed the appropriate allelic constitution (heterozygous – *Gb*/*Gh* – for the locus of interest), and to carry forward the introgression process only for these plants. Thus, only 47 BC3 plants out of 411 (11.4%), and 37 BC4 plants out of 450 (8.2%) were backcrossed to the recurrent parent (and self-pollinated in the case of the BC4 plants). Following the SSR-based foreground selection of targeted regions, the selected BC3 and BC4 plants were genotyped with a subset of AFLP primers that covered the rest of the genome, in order to analyze the allelic composition of the unselected genomic regions (background selection) (see Fig.1). At both the BC3 and BC4 generations, the early selection of the carrier chromosome(s) and the verification of the status of the non-carrier chromosomes allowed us to follow the amount and the quality of donor genome that was retained throughout the backcrossing stages. As an illustration of the selection pressure applied through the use of molecular markers for foreground selection, Figure 2 shows the graphical genotype of a BC4 individual (Fig.2b), as well as that of the BC1 plant from which this individual is derived (Fig.2a). This figure shows that, at least in some cases, the process used is efficient in selecting for chromosomal regions of interest (foreground selection), while letting the rest of the genome return towards that of the recurrent parent. In this particular example, the BC4 plant (Fig.2b) has retained, at the heterozygous state, genomic regions carrying favorable alleles on chromosomes/linkage groups c6, c25 and

c26. The other regions carrying QTLs on c3, c23, c20, A01 and A03, which were heterozygous on the BC1 plant (Fig.2a), have partly or completely returned to the homozygous *Gh/Gh* state. Most of the non-carrier genome (91% of the genotyped loci) of the BC4 plant has returned to the homozygous state (Fig.2b, non-carrier chromosomes).

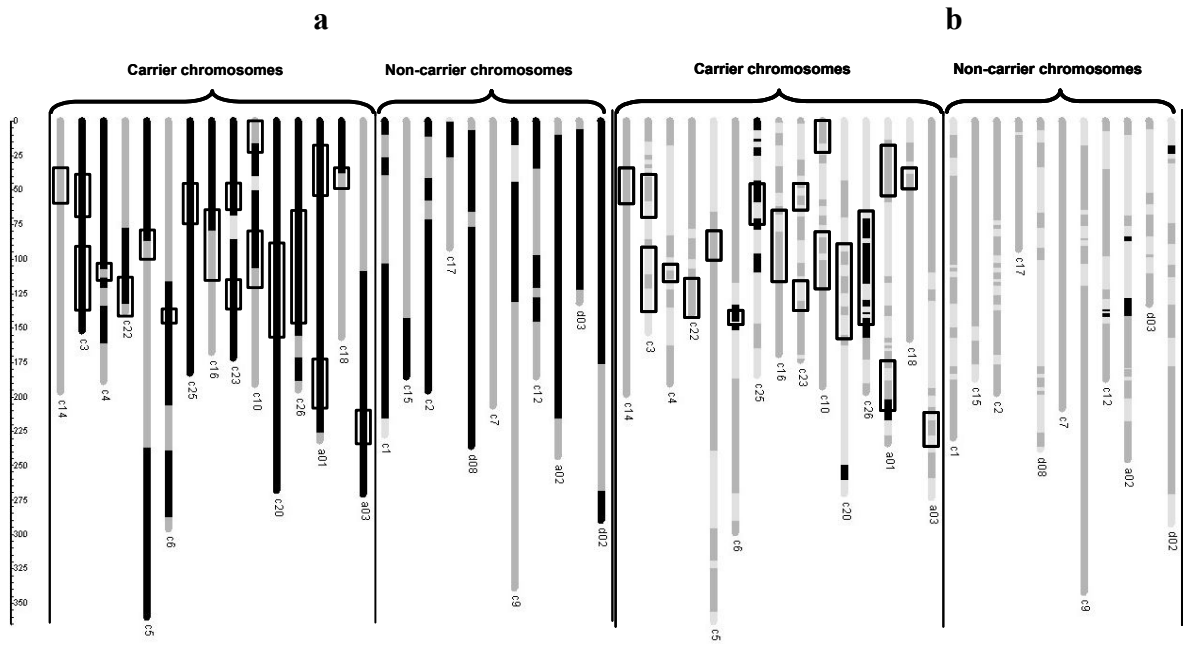

Figure 2
Graphical genotype (26 chromosomes) of a BC1 plant (n° BC1/16) (a), and of a selected BC4 plant (n° BC4/419) (b), derived from it. The 2 possible allelic forms, homozygous *Gh*/*Gh* and heterozygous *Gh*/*Gb*, are denoted in grey and black respectively. Light grey areas indicate portions of unknown allelic composition. Boxed areas represent the localization of QTLs on carrier chromosomes.

Discussion

In an attempt to overcome the limitations of conventional breeding for the improvement of cotton fiber quality through the use of interspecific hybridization, we decided to use molecular markers in a MAS scheme to improve the efficiency of the introgression of fiber quality traits. We decided to use the Advanced Backcross-QTL (AB-QTL) analysis strategy (Tanksley and Nelson, 1996), since this method allowed us to concomitantly develop a genetic map of the cotton genome, carry out a fiber quality QTL analysis, and attempt the introgression of favorable alleles in an adequate recipient genetic background (Fig.1).

Cotton fiber quality is a complex concept, which involves a number of fiber traits or characteristics. Each of these traits is under the influence of numerous QTLs, indicating a complex genetic determinism. Indeed, in our hands, at least 6 QTLs govern fiber uniformity, and up to 21 QTLs influence fiber fineness. When considering 6 traits that can account for fiber quality, a total of 80 QTLs were detected (Table 1). This figure falls in the same range as that found by Paterson *et al.* (2003). Due to the fact that some of these QTLs co-localized within the same chromosome region, and by choosing those QTLs whose positive allele derived from the donor parent and had the strongest effect on economically-important fiber characteristics, we reduced the number of target regions to be introgressed down to 19 (Table 2). Nevertheless, this number of QTLs remains too high to allow the identification of a single

plant that would carry them all. Indeed, in our experience, at the BC3 stage, single plants carried a maximum of 5 regions of interest, while at the BC4 stage this number was reduced to 4. At this stage of the MAS process, 2 routes can be envisaged (Fig.1). The first avenue consists in identifying the best BC4 plants, i.e. those showing the highest amount of favorable QTL introgression, and fixing the favorable allele by self pollination. Such BC4S1 plants can then be crossed with other BC4S1 plants of different ascent in order to pyramid as many QTLs as possible (each contributing to different traits) within the same genome. Similarly, BC4S1 plants can be used to pyramid various QTLs responsible for a given trait ("selective pyramiding"). This latter strategy could especially apply to traits of commercial importance, such as fiber strength or fineness. The second avenue consists in carrying further the backcrossing process until the development of near isogenic lines differing only at a given QTL (QTL-NILS). Such plant material could prove useful not only to study the effect of a single given QTL on the phenotypical value of a plant harboring it, but also in the case the introgressed QTL proves to contribute significantly to the improvement of a given trait. Furthermore, QTL-NILS could also be used as donor material for QTL pyramiding. Whatever the case, only the phenotypical analysis of the plant material stemming from the MAS process will allow the validation of this procedure.

References

Jiang C., Wright R., El-Zik K.M., and Paterson A.H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). P Natl Acad Sc USA **95**: 4419-4424

Kohel R.J., Yu J., Park Y.-H., and Lazo G. (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica **121**: 163-172

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

Paterson A.H., Saranga Y., Menz M., Jiang C., and Wright R.J. (2003) QTL analysis of genotype x environment interactions affecting cotton fiber quality. Theor Appl Genet **106**: 384-396

Tanksley S.D., and Nelson J.C. (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet **92**: 191-203