Arabidopsis thaliana as a source of candidate genes for cotton fiber quality

M. Giband[1], S. Pagant[2], C. Pannetier[1, 2] and H. Höfte[2]
[1]CIRAD-CA, Programme Coton, Montpellier FRANCE
[2]Laboratoire de Biologie Cellulaire, INRA-Centre de Versailles, Versailles France
Correspondence author marc.giband@cirad.fr

ABSTRACT

Cotton fibers are single-celled structures that differentiate from epidermal cells of the outer integument of the seed coat that elongate, fill with secondary cell wall cellulose, mature and die. Studies, which include the analysis of gene expression throughout fiber development and the detection of fiber quality QTLs, are underway in many laboratories to identify loci and/or genes of interest that might explain the underlying mechanisms that determine cotton fiber development and quality. The model plant Arabidopsis thaliana *is being used to investigate the molecular basis of many important biological functions, some of which, such as cell differentiation, cell elongation, and cellulose biosynthesis and deposition are relevant to the development of cotton fiber cells and orthologous genes active in these processes might provide useful candidates for genes that determine fiber quality traits. Cellulose synthesis, in particular, could play an important role in determining certain aspects of cotton fiber quality, such as fiber length, strength, or fineness. Recent advances in understanding the synthesis of cellulose have come from the analysis of* A. thaliana *cellulose-deficient mutants, and the isolation of a number of genes involved in this process, including the 10 members of the catalytic sub-unit of the cellulose synthase (CESA). Database searching has identified a number of cotton EST CESA orthologs that cluster with* A. thaliana *sequences rather than together, allowing the identification of orthologs for five of the 10* A. thaliana *CESA isoforms. Furthermore, one cotton EST formed a branch on its own, possibly indicating a specialized function for this gene in fiber development. Other genes involved in cellulose biosynthesis, as well as genes involved in cell elongation or in maintaining polarized cell expansion have been identified in* A. thaliana.*and for most of these genes, cotton orthologs can be found among fiber ESTs. These orthologs represent useful candidate genes to develop molecular markers for genetic mapping, fiber quality QTL analysis, and diversity studies in cotton and could be the targets for transformation experiments aimed at understanding their respective contribution to fiber quality, thus opening the way to the genetic engineering of cotton for the modification of fiber traits.*

Introduction

Cotton fibers are specialized single-celled structures that undergo four different but overlapping developmental stages: cell differentiation, cell elongation (primary cell wall synthesis), cell thickening (secondary cell wall synthesis), and maturation (cell death and dehydration) (Basra and Malik, 1984). Each of these stages will have an impact on the quality of the end-product: the rate and duration of the stage of cell elongation will determine fiber length, while the stages of fiber initiation and cell wall thickening, or the combination of both, will impact on the fineness of the fibers. Despite the interest in understanding the molecular bases of fiber quality to optimize cotton (fiber) improvement, little is known on the molecular events underlying the various parameters determining fiber quality.

A first approach used to understand the genetic bases of fiber quality and to identify loci involved in determining such parameters consists in the molecular mapping of fiber quality QTLs. Recent advances in this field have led to the identification of molecular markers linked to such QTLs in cotton (Jiang *et al.*, 1998; Paterson *et al.*, 2003; Zhang *et al.*, 2003), which should facilitate the manipulation of chromosomal regions involved in determining important quantitative traits through the use of molecular markers in a marker-assisted selection (MAS) program. Though useful for breeding programs aimed at improving fiber quality, this approach does not allow the identification of the specific genes involved in determining the trait of interest, nor does it allow the understanding of how these genes contribute to the trait.

The isolation of genes expressed in cotton fibers, and the study of their expression throughout fiber development could contribute to be a better understanding of the molecular mechanisms regulating fiber growth and in fine, involved in determining fiber quality. Recent years have seen advances in the characterization of such genes and of their expression profiles (for examples, see Wilkins and Jernstedt, 1999; Harmer *et al.*, 2002; Li *et al.*, 2002, and references therein). Although these types of experiments do not constitute the proof of the involvement of a particular gene in the elaboration of fiber quality, nor do they permit an insight in the complexity of gene regulation and interaction, they are most useful in identifying important genes. Indeed, such fiber-specific (or fiber–enriched) genes constitute good candidate genes that could be used to develop molecular markers for fiber quality QTL analyses. Such analyses could constitute a first step towards linking a specific gene (or allele thereof) to a phenotype.

Another avenue for the identification of fiber quality candidate genes is to take advantage of the knowledge gained from studies using Arabidopsis thaliana. Indeed, this model plant is being used to investigate

the molecular bases of important biological functions, some of which may be relevant to the development of cotton fiber cells. Among the important biological functions of plant cells, cellulose biosynthesis and deposition, cell elongation, and polarized cell expansion are those that seem the most relevant for the study of cotton fiber development. For each of these functions, *A. thaliana* mutants have been identified, and the corresponding genes isolated. Examples of such genes include those encoding the catalytic sub-unit of the cellulose synthase (CESA genes), other genes involved in cellulose biosynthesis (KORRIGAN, Nicol *et al.*, 1998; KOBITO1, Pagant *et al.*, 2002), COBRA (Schindelman *et al.*, 2001, Roudier *et al.*, 2002), a gene involved in maintaining the orientation of cell expansion, and genes involved in maintaining the orientation of microtubules (BOTERO1, Bichet *et al.*, 2001; TONNEAU1, Traas *et al.*, 1995) (for reviews, see Delmer, 1999; Williamson *et al.*, 2002; Doblin *et al.*, 2002).

As a first step in applying the *A. thaliana* candidate gene approach to cotton fiber quality QTL analysis, we have identified the cotton ortholog(s) of some of these genes among the fiber ESTs available in public databases. These sequences were then positioned on a genetic map of tetraploid cotton, which was derived from an interspecific cross (*G. hirsutum* var. Guazuncholl x *G. barbadense* var. VH8, Lacape *et al.*, this volume).

Experimental procedure

Phylogenic analysis of cotton and A. thaliana *CESA sequences*

The phylogenic analysis of the CESA sequences from cotton and *A. thaliana* was conducted as described in Fagard *et al.* (2000). The unrooted tree was built after aligning the second hypervariable (HVR2) domain (corresponding to amino acids 690 to 743 in the Arabidopsis CESA6 member (PROCUSTE 1 (PRC1) Fagard *et al.* (2000)) of the *A. thaliana* and cotton CESA sequences. The HVR2 region was chosen to build the phylogenic tree due to sequence availability (the cotton ESTs correspond to 3'-end sequencing), but also to maximize the chances of distinguishing different family members, which otherwise are very conserved.

Identification of the cotton orthologs of the A. thaliana genes of interest

In order to identify the cotton ortholog(s) of the *A. thaliana* genes of interest, the latter sequences were used to search the NCBI EST database (dbEST) using the "BLAST" function. In some cases (as for the search for the ortholog of the KOR gene), no significant homology with cotton sequences was found. In such a case, the "TBLASTX" (translated query-translated db) function was used. Only the cotton sequence showing the highest homology with the *A. thaliana* sequence was used for further analysis.

Development of molecular markers from candidate gene sequences

The cotton sequences of the candidate genes were used to design adequate primers. In the case of the CESA sequences, primers were chosen as to amplify the HVR2 domain. The specificity of the primers was checked through sequencing of the PCR amplification products. The amplification products were then used as RFLP probes. Alternatively, the same primer pairs were used to generate CAPS markers or as primers for a modified AFLP protocol (M. Giband, unpublished).

Mapping of the candidate genes

The mapping population is composed of 75 BC1 plants originating from an interspecific cross between *G. hirsutum* var. Guazuncholl and *G. barbadense* var. VH8 (backcrossed to the *G. hirsutum* parent). The Mapmaker software was used to position the markers for which polymorphism could be detected between the parental varieties. The mapping population, and the parameters used, are described in more detail in Lacape *et al.* (this volume).

Results

Phylogenic analysis of cotton and A. thaliana *CESA sequences*

The complete sequencing of the *A. thaliana* genome (The Arabidopsis Genome Initiative, 2000) has revealed that there are 10 isoforms of the catalytic subunit of the cellulose synthase (CESA) (see *http://www.arabidopsis.org/home.html; http://cellwall.stanford.edu/*). Among the (close to) 50,000 cotton fiber EST sequences present in the NCBI EST database (GenBank dbEST Release 021403 from Feb. 14th 2003 contains close to 39,000 ESTs from *G. arboreum* and close to 10,000 ESTs from *G. hirsutum*), a number of them show homology to *CESA* genes. When considering the *G. hirsutum* sequences (including non-EST sequences), these can be classified into six different "contigs" (*http://cellwall.stanford.edu/*). A phylogenic analysis was used to study the relationship between the 10 *A. thaliana* sequences and those from cotton (*G. hirsutum*). In this study, the HVR2 domain, corresponding to amino acids 690 to 743 close to the C-terminus of the PRC1 sequence, was used to construct the similarity tree. This choice was guided not only for practical reasons (many of the *G. hirsutum* ESTs correspond to 3' single-pass sequences), but also because it maximizes the chance of distinguishing among the different family members (Fagard *et al.*, 2000) which otherwise show a high degree of similarity. The results of this analysis show that the cotton sequences do not cluster together, but with one or more sequences from *A. thaliana* (Figure 1). Thus, the HVR2 domain of the CESA proteins is more conserved among orthologs of the same isoform than among different isoforms from the same species. *G. hirsutum* (Gh) EST-AI728789 (named Gh CesAX in this

study) formed a cluster with *A. thaliana* (At) CesA9, At CesA2, At CesA5 and At CesA6, while Gh CelA2 (U58284) clustered with At CesA4, Gh EST-AI725882 (Gh CesA4) clustered with At CesA1 and At CesA10, Gh CelA3 (AF150630) with At CesA3, and Gh CelA1 (U58283) with At CesA8. Interestingly, Gh EST-AI727450 (Gh CesAN) formed a clade on its own, pointing to a possible specialized function for this gene perhaps in fiber development. No cotton sequence clustered with At CesA7, indicating that, if present in the cotton genome, the ortholog of this gene is not highly, or at all, expressed in fibers.

Identification of cotton orthologs

The analysis of *A. thaliana* mutants has led to the isolation of a number of genes involved in biological functions that are relevant to fiber development. Indeed, the way (and amount) cellulose is synthesized and deposited within fiber cells, the rate and duration of cell elongation (or polarized cell expansion), and other such phenomena will have a repercussion on the (physical) properties of the mature fiber. *A. thaliana* mutants impaired in such functions have been identified and the gene(s) involved characterized (see Table 1). Searches in the NCBI EST database, using the "BLAST" or "TBLASTX" functions, in conjunction with clustering information (Figure 1) led to the identification of at least one cotton ortholog for each of the *A. thaliana* sequences of interest (Table1). This shows that, not surprisingly, the great biological functions are conserved among plant species. In our study, only the cotton sequence showing the closest match with any given *A. thaliana* sequence was retained for further analysis.

Mapping of the candidate genes

Genes involved in the functions that seem relevant to the elaboration of fiber quality represent valuable candidate genes for a QTL analysis. The genes that were considered are those encoding different genes of the *CESA* family (Figure 1 and Table 1), as well as the cotton orthologs of *A. thaliana* genes of interest identified through mutants (Table 1). As a first step towards using these candidate genes for a fiber quality QTL analysis, we sought to position these genes on a map of tetraploid cotton (Lacape *et al.*, this volume). To this end, several techniques were used to detect polymorphism, within the candidate genes, between the parental species of the mapping population. Among the 11 candidate genes tested (6 Gh *CESA* isoforms, and the orthologs of *KORRIGAN*, *KOBITO*, *TONNEAU1*, *COBRA*, and *BOTERO1*), only for two of them (Gh *CELA3*, Gh *KORRIGAN*) were we not able to detect polymorphism between the parental species. The nine candidate genes that showed polymorphism revealed 16 loci, between one and four scorable loci being revealed per gene tested. Figure 2 shows an example of a polymorphism detected by using a candidate gene (Gh *COBRA*) as a RFLP probe, while Figure 3 shows an example of a polymorphism revealed using candidate gene-derived primers to develop CAPS markers. Mapping allowed assigning the 16 polymorphic loci to 13 chromosomes or linkage groups. Locus CesA1-a mapped on chromosome c4, Kob-b and Bot-a both mapped on chromosome c22, CesA1-c was localized on chromosome c5, Kob-a on linkage group D04, Bot-b and CesA4-a are both found on c16, CesAN-a on c9, Bot-c on c23, CesAX-a on c12. *COBRA* was mapped on chromosome c26, while CesA2-a and CesA1-b were on linkage groups A02 and D03 respectively (J. M. Lacape, personal communication). CesA1-d was positioned between microsatellite markers CIR003 and BNL1066 on linkage group A03, while Ton1-a and Ton1-b both co-localized with RFLP markers A1296 and pXP3_26a, and microsatellite markers BNL2895a, CIR398, CIR061a and CIR068 on linkage group D02 (Lacape *et al.*, this volume).

Discussion

Recent years have seen the unprecedented development of efforts aimed at understanding the molecular bases of cotton fiber quality. These efforts range from the mapping of fiber quality QTLs (Jiang *et al.*, 1998; Paterson *et al.*, 2003; Zhang *et al.*, 2003) to the identification, isolation, and study of genes expressed in developing fibers (for examples, see Wilkins and Jernstedt, 1999; Harmer *et al.*, 2002; Li *et al.*, 2002, and references therein; see also *http://cottongenomecenter.ucdavis.edu/*). Nevertheless, due to the complexity of the processes that underlie the elaboration of traits such as fiber quality, it is most probable that not one single approach will be sufficient to gain a better understanding of the molecular events involved. Instead, an integrated effort combining various approaches (genetics, genomics, biochemistry...) has a much better chance of succeeding in unraveling such a complex system. In addition to the abovementioned efforts undertaken in cotton, the transposition of knowledge gained from the study of model systems (in this case, *A. thaliana*) to species of agronomical interest represents an attractive avenue. Indeed, this model plant is being widely used to study biological functions that are relevant to the development of cotton fiber cells, and thus to the elaboration of fiber quality. Not only has the genome of *A. thaliana* been completely sequenced, but numerous mutants, impaired in important functions, have also been identified (see *http://www.arabidopsis.org/home.html*). This has led to the identification of key genes involved in these functions, and in some cases, has allowed the biochemical characterization of the mutants, making it possible to link together gene expression data, biochemical data, and phenotype.

As a first step in demonstrating the usefulness of such an approach, we have isolated the cotton orthologs of genes of interest identified in *A. thaliana*, and used them to develop molecular makers, allowing these genes to be positioned on the genetic map of tetraploid cotton. The next step will consist in analyzing if

these genes tightly co-localize with fiber quality QTLs. If so, the molecular markers derived from these genes could be used in marker-assisted selection programs aimed at improving fiber quality, or to conduct studies on the genetic diversity of genes of interest. The final proof of the involvement of a candidate gene in the elaboration of a given trait will come from the analysis of mutants affected in this gene, and/or through cotton transformation experiments aimed at understanding its contribution to fiber quality.

The cotton orthologs of genes of interest identified in *A. thaliana* represent useful candidate genes to study the molecular bases of fiber quality. Such genes can be used to develop molecular markers for genetic mapping, fiber quality QTL analysis, and diversity studies. Furthermore, these genes could be the targets for transformation experiments aimed at understanding their respective contribution to fiber quality, thus paving the way to the genetic engineering of cotton for the modification of fiber traits.

Acknowledgements

We thank our colleagues J.M. Lacape, T.B. Nguyen and B. Bojinov for their valuable help and suggestions throughout this work.

References

- Arioli, T., Peng, L., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Höfte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J. and Williamson, R.E. (1998). Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science,* ***279****: 717-720.*
- Basra, A.S. and Malik, C.P. (1984). Development of the cotton fiber. *International Review of Cytology,* ***89****: 65-113.*
- Bichet, A., Desnos, T., Turner, S., Grandjean, O. and Höfte, H. (2001). BOTERO1 is required for normal orientation of cortical microtubules and anisotropic cell expansion in Arabidopsis. *The Plant Journal,* ***25****: 137-148.*
- Bouquin, T., Mattsson, O., Naested, H., Foster, R. and Mundy, J. (2003). The *Arabidopsis* lue1 mutant defines a katanin P60 ortholog involved in hormonal contyrol of microtubule orientation during cell growth. *Journal of Cell Science,* ***116****: 791-801.*
- Burk, D.H., Liu, B., Zhong, R., Morrison, W.H. and Ye, Z.H. (2001). A katanin-like protein regulates normal cell wall biosynthesis and cell elongation. *The Plant Cell,* ***13****: 807-827.*
- Cano-Delgado, A.I., Metzlaff, K. and Bevan, M. (2000). The eli1 mutation reveals a link between cell expansion and secondary wall formation in *Arabidopsis thaliana*. *Development,* ***127****: 3395-3405.*
- Delmer, D.P. (1999). Cellulose biosynthesis: exciting times for a difficult field of study. Annual Review of *Plant Physiology and Plant Molecular Biology,* ***50****: 245-276.*
- Desprez, T., Vernhettes, S., Fagard, M., Refrégier, G., Desnos, T., Aletti, E., Py, N., Pelletier, S. and Höfte, H. (2002). Resistance against herbicide Isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. *Plant Physiology,* ***128****: 482-490.*
- Doblin, S.D., Kurek, I., Jacob-Wilk, D. and Delmer P.D. (2002). Cellulose biosynthesis in plants: from genes to rosettes. *Plant and Cell Physiology,* ***43****: 1407-1420.*
- Fagard, M., Desnos, T., Desprez, T., Goubet, F., Refrégier, G., Mouille, G., McCann, M., Rayon, C., Vernhettes, S. and Höfte, H. (2000). PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of *Arabidopsis*. *The Plant Cell,* ***12****: 2409-2423.*
- Harmer, S.E., Orford, S.J. and Timmis, J.N. (2002). Characterization of six a-expansin genes in *Gossypium hirsutum* (Upland cotton). *Molecular Genetics and Genomics,* ***268****: 1-9.*
- 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). *Proceedings of the National academy of sciences of the United States of America,* ***95****: 4419-4424.*
- Li, C.H., Zhu, Y.Q., Meng, Y.L., Wang, J.W., Xu, K.X., Zhang, T.Z. and Chen, X.Y. (2002). Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR. *Plant Science,* ***163****: 1113-1120.*
- Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H. and Höfte, H. (1998). A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. *The EMBO Journal,* ***17****: 5563-5576.*
- Pagant, S., Bichet, A., Sugimoto, K., Lerouxel, O., Desprez, T., McCann, M., Lerouge, P., Vernhettes, S. and Höfte, H. (2002). KOBITO1 encodes a novel plasma membrane protein necessary for normal synthesis of cellulose during cell expansion in *Arabidopsis*. *The Plant Cell,* ***14****: 2001-2013.*
- 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. *Theoretical and Applied Genetics,* ***106****: 384-396.*
- Roudier, F., Schindelman, G., DeSalle, R. and Benfey, P.N. (2002). The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellow fellowship in expansion. *Plant Physiology,* ***130****: 538-548.*
- Scheible, W.R., Eshed, R., Richmond, T., Delmer, D. and Somerville, C. (2001). Modifications of cellulose synthase confer ressistance to isoxaben and thiazolidinones herbicides in *Arabidopsis* Ixr1 mutants. *Proceedings of the National Academy of Sciences of the United States of America,* ***98****: 10079-10084.*

- Schindelman, G., Morikami, A., Jung, J., Baskin, T.I., Carpita, N.C., Derbyshire, P., McCann, M.C. and Benfey, P.N. (2001). COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes and Development,* ***15****: 1115-1127.*
- Taylor, N.G., Laurie, S. and Turner, S.R. (2000). Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. *The Plant Cell,* ***12****: 2529-2539.*
- Taylor, G.N., Howells, R.M., Huttly, A.K., Vickers, K. and Turner, S.R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences of the United States of America,* ***100****: 1450-1455.*
- The *Arabidopsis* Genome Initiative (2000). Analysis of the genome of the flowering plant *Arabidopsis thaliana*. *Nature,* ***408****: 796-815.*
- Traas, J., Bellini, C., Nacry, P., Kronenberger, J., Bouchez, D. and Caboche, M. (1995). Normal differentiation patterns in plants lacking microtubular preprophase bands. *Nature,* ***375****: 676-677.*
- Wilkins, T.A. and Jernstedt, J.A. (1999). Molecular genetics of developing cotton fibers. *In* Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing, Basra, A.S (Ed), Food Product Press, Binghamton, N.Y., pp231-269.
- Williamson, R.E., Burn, J.E. and Hocart, C.H. (2002). Towards the mechanism of cellulose synthesis. *Trends in Plant Science,* ***7****: 461-467.*
- Zhang, T., Yuan, Y., Yu, J., Guo, W. and Kohel, R.J. (2003). Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. *Theoretical and Applied Genetics,* ***106****: 262-268.*

Table 1. *List of genes of interest identified in* A. thaliana*, with the function (and/or role) of the protein they encode. When available, the name of the corresponding mutant is given. The GenBank accession numbers of the* A. thaliana *genes and of their respective cotton orthologs are shown.*

Gene of interest (*A. thaliana* mutant)	Function (role)	*A. thaliana* GenBank accession number	Reference	Cotton ortholog GenBank accession number (Blast E-value)
CesA1 (*rsw1*)	Cellulose synthase (I^{ary} wall synthesis)	AF027172	Arioli *et al.* (1998)	AI725882 (3 e-33)
CesA2	Cellulose synthase (I^{ary} wall synthesis)	AF027173	Unpublished	AI728789 (8 e-26)
CesA3 (*ixr1, eli1*)	Cellulose synthase (I^{ary} wall synthesis)	AB018111	Scheible *et al.* (2001) Cano-Delgado *et al.* (2000)	AF150630 (0.0)
CesA4 (*irx5*)	Cellulose synthase (II^{ary} wall synthesis)	AB006703	Taylor *et al.* (2003)	U58284 (1 e-50)
CesA5	Cellulose synthase (I^{ary} wall synthesis)	AB016893	Unpublished	AI728789 2 e-14)
CesA6 (*prc1, ixr2*)	Cellulose synthase (I^{ary} wall synthesis)	AF062485	Fagard *et al.* (2000) Desprez *et al.* (2002)	AI728789 (< 1 e-05)*
CesA8 (*irx1*)	Cellulose synthase (II^{ary} wall synthesis)	AL035526	Taylor *et al.* (2000)	U58283 (1 e-33)
CesA9	Cellulose synthase (I^{ary} wall synthesis)	AC007019	Unpublished	AI728789 (4 e-18)
CesA10	Cellulose synthase (I^{ary} wall synthesis)	AC006300	Unpublished	AI725882 (2 e-57)
KORRIGAN (*kor*)	Endo-1,4-β-glucanase (Cellulose synthesis)	AF073875	Nicol *et al.*(1998)	AW729802 ** (2 e-16)
KOBITO (*kob1*)	(Cellulose biosynthesis)	AV548752	Pagant *et al.* (2002)	BF275186 (4 e-41)
TONNEAU (*ton1*)	(Orientation of microtubules)	AF280059	Traas *et al.* (1995)	AW187391 (3 e-32)
COBRA (*cobra*)	(Orientation of cell expansion)	AB008269.1	Schindelman *et al.* (2001)	AI730765 (2 e-08)
BOTERO (*bot1, fra2, lue1*)	Katanin (orientation of microtubules)	AF048706	Bichet *et al.* (2001) Burk *et al.* (2001) Bouquin *et al.* (2003)	AI729407 (4 e-75)

* For CesA6, Blast searches did not return highly significant hits; in this case, the identity of the cotton ortholog was deduced from the clustering analysis (see Figure 1).

**Note added in proof: since submission of the manuscript, cotton accession number AW729802 was withdrawn from GenBank, and is replaced by accession number BG444933 which shows 100% homology with it.

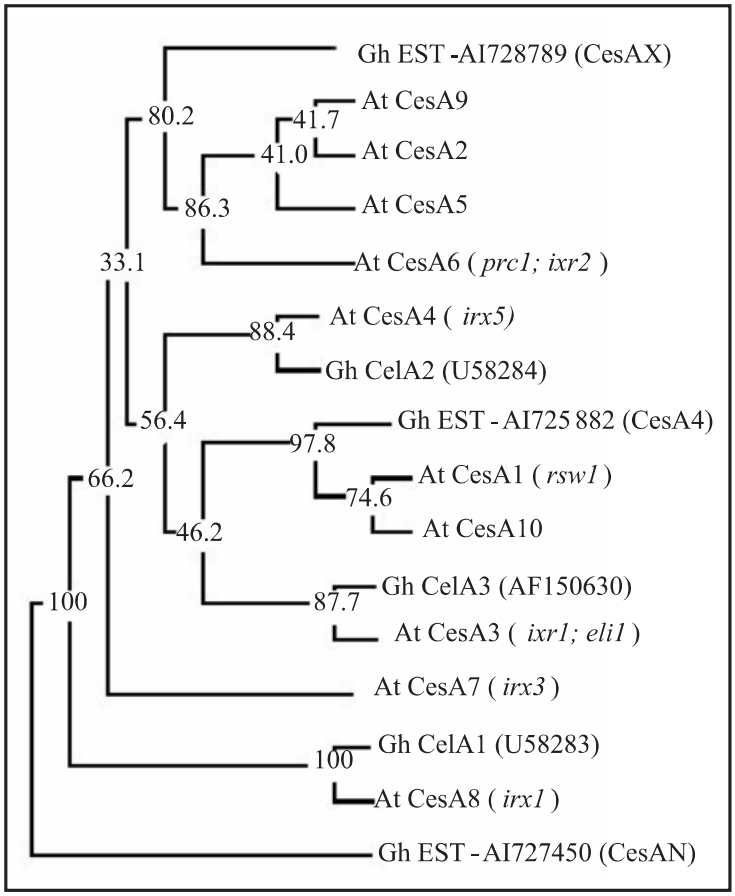

Figure 1.
Phylogenic tree based on the HVR2 domain of the A. thaliana *and cotton CESA protein sequences. The cotton sequences cluster with one or more sequences from* A. thaliana*, rather than together, indicating better conservation between orthologs than between isoforms (adapted from Fagard et al., 2000).*

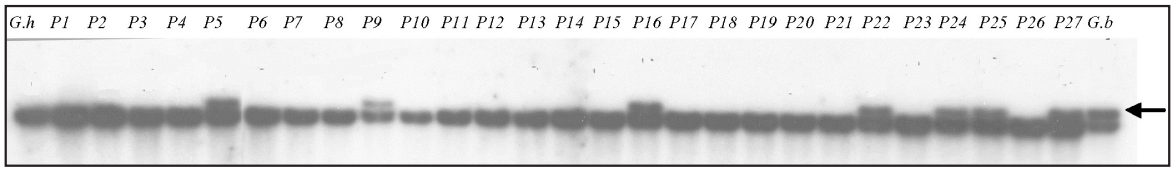

Figure 2.
An example of polymorphism revealed by using a candidate gene (COBRA) as a RFLP probe. The arrow points to the G. barbadense*-specific band (upper band), which segregates among progenies of the BC1 mapping population. G.h:* G. hirsutum*; G.b:* G. barbadense*; P1 to P27 represent different BC1 plants originating from a cross between the parental species.*

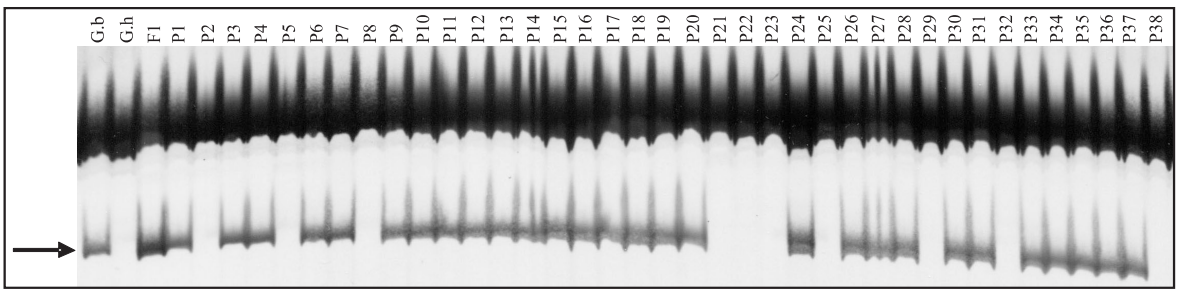

Figure 3.
An example of polymorphism detected using candidate gene-derived CAPS markers. The arrow points to the G. barbadense*-specific band, which is absent from the* G. hirsutum *parent, and present in the F1 and in some plants of the segregating BC1 population. G.b:* G. barbadense*; G.h:* G. hirsutum*; F1: F1 plant originating from the interspecific cross; lanes P1 to P38 represent different BC1 plants.*