



## Genes expressed in cotton (*Gossypium hirsutum*) buds isolated with a subtractive library

M.P.N. Pinheiro<sup>1</sup>, V.G.L. Batista<sup>2</sup>, N.F. Martins<sup>3</sup>, R.C. Santos<sup>2,4</sup>,  
P.A. Melo Filho<sup>5</sup>, C.R.C. Silva<sup>1</sup> and L.M. Lima<sup>2,4</sup>

<sup>1</sup>Departamento de Pós-Graduação em Biotecnologia,  
Rede Nordeste de Biotecnologia, Universidade Estadual do Ceará,  
Fortaleza, CE, Brasil

<sup>2</sup>Departamento de Pós-Graduação em Ciências Agrárias,  
Universidade Estadual da Paraíba, Campina Grande, PB, Brasil

<sup>3</sup>Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil

<sup>4</sup>Embrapa Algodão, Campina Grande, PB, Brasil

<sup>5</sup>Departamento de Agronomia, Universidade Federal Rural de Pernambuco,  
Recife, PE, Brasil

Corresponding author: L.M. Lima  
E-mail: liziane.lima@embrapa.br

Genet. Mol. Res. 12 (1): 37-43 (2013)

Received July 10, 2012

Accepted November 26, 2012

Published January 16, 2013

DOI <http://dx.doi.org/10.4238/2013.January.16.7>

**ABSTRACT.** A subtractive cDNA library from cotton buds was constructed to prospect for differentially expressed genes related to early bud development. A library was constructed and 768 cDNA sequences were obtained, comprising 168 clusters, with 126 contigs and 42 singlets. Both the *Gossypium* as well as *Arabidopsis* databases were utilized for the *in silico* analysis, since some genes identified in cotton have not yet been studied for functionality, although they have homology with genes from other species. The transcriptome revealed a large number of transcripts, some of them with unknown function, and others related to pollen development, pollen tubes, ovules, and fibers at different stages. The most populated contig was identified as fiber

from 0-10 days after anthesis, with 12 reads. The success and novelty rates generated from the library were 67 and 51%, respectively. The information obtained here will provide a framework for research on functional cotton genomics.

**Key words:** Fiber; Ovule; Genes; Sequencing

## INTRODUCTION

Cotton (*Gossypium hirsutum* L.) is one of the most important commodity crops and a significant employment generator worldwide. Cotton fibers, the main product of the plant, are an excellent raw material in both the textile industry and manufacturing (Park et al., 2010). However, the management of this crop is expensive owing to its production costs, of which 40% support machinery and pest and disease control (Hashemi et al., 2009; McKiniona et al., 2009).

Cotton crops are susceptible to attack by various pests, which cause direct losses in yield and therefore to fiber trading. Among these insects, coleopteran and lepidopteran are the most damaging and difficult to control biologically (Grossi-de-Sá et al., 2007; Showler, 2008). In these cases, genetic transformation has contributed substantially to the improvement of cotton, introducing genes that provide resistance to several biotic and abiotic factors (James, 2010).

Recent genomic studies of several vegetal species have generated a wealth of information and developed useful DNA sequences and complementary DNA (cDNA) databases for biotechnological research (Rao et al., 2011). *Gossypium* spp along with *Arabidopsis thaliana* have expressive and available expressed sequence tag (EST) databases synthesized from several plant tissues at various physiological stages. The Cotton Genome Database and the Arabidopsis Information Resource are specific to cotton and *A. thaliana*, respectively, and are therefore important tools for quick searching and selection of potential genes of agronomic importance.

The Cotton Genome Database contains several genes that have already been functionally characterized, including those related to expression in reproductive structures, knowledge of which is essential for further studies related to the development of transgenic plants resistant to tissue-specific pests (Zhou et al., 2008). Such genes, often regulated by semi-constitutive or specific promoters, have various levels of expression (Shelenkov and Korotkov, 2009). Studies involving isolation of tissue-specific promoters are necessary in genetic engineering given the arsenal of genes currently available to control several biotic and abiotic problems that affect the yield of major crops and the possibility of using such promoters to improve the expression of the target genes (Hsu et al., 2005; Wroblewski et al., 2005; Chen et al., 2007). Herein, we report a subtractive cDNA library constructed to search for structural and regulatory genes in cotton buds to acquire knowledge of some of the genes involved in the regulation of growth and development.

## MATERIAL AND METHODS

### Germplasm and tissue collection

Seeds of early cotton CNPA 8H were grown in a greenhouse at Embrapa Algodão, in Campina Grande, PB, Brazil (7°13'11"S and 35°52'31"W). Buds at several stages, leaves, stems, and roots were collected for RNA extraction (Table 1).

**Table 1.** Major events related to the biological stage of development in cotton buds.

Bud length (mm)	Event associated
2	Microspore mother cells at the end growth phase; outline of ovules
4	Microspore mother cells in prophase; differentiation of tapetum cells; development ovules
6-8	Microspores in the anthers are completely separate; early differentiation of the macrospore mother cell; degeneration of tapetum; macrospore mother cells in prophase
10	Binucleated pollen granules; macrospore mother cells after the second division
12	Embryo sac with two or four nuclei
14	Anther wall reduced to two layers of cells; complete embryo sac; antipodals present
16	Antipodals degenerate
18-20	Antipodals disappearance and degeneration of their nuclei

Adapted from Quintanilha et al., 1962.

### RNA extraction and cDNA library construction

RNA was isolated from fresh tissue samples (100 mg) using a Plant RNA Mini-Spin Invisorb kit (Invitec, Germany). cDNA was synthesized using a Super SMART PCR cDNA Synthesis Kit (Clontech, USA). A subtractive cDNA library was then constructed using a PCR Select cDNA Subtraction kit (Clontech) and the vector pGEMT-Easy (Promega, USA).

### Cloning and sample sequencing

The library was cloned in *Escherichia coli* strain XL1-blue by electroporation. Then, the cells were restored with LB (Luria-Bertani) liquid medium and further incubated overnight at 37°C. Plasmids were prepared on 96-well plates according to the protocol recommendation (Neto Borges et al., 2005). The cDNA inserts were sequenced using T7 and SP6 primers in automatic sequencer (Applied Biosystems model 3700).

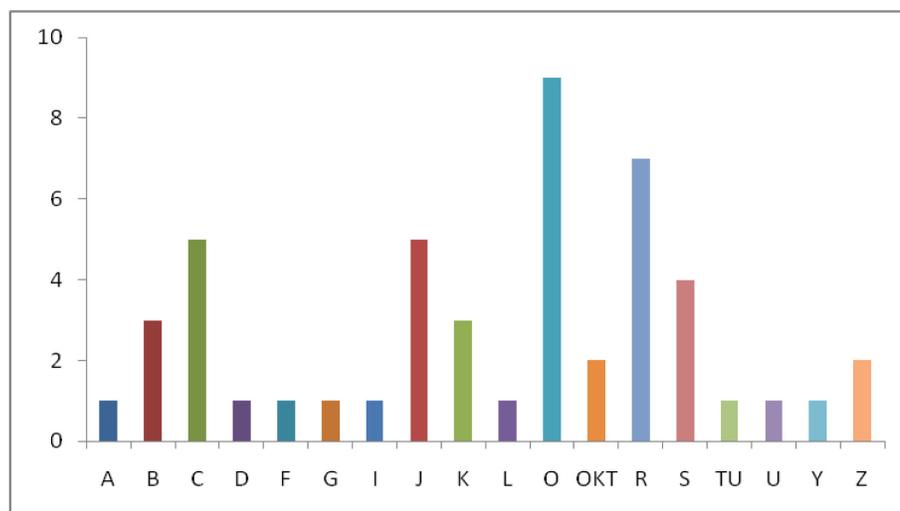
### Clustering and *in silico* analysis

The sequences generated were deposited in the Genome System (SISGEN) Bioinformatics Laboratory at Embrapa Recursos Genéticos e Biotecnologia (<http://genoma.embrapa.br>). The quality of sequences was performed using the PHRED program (Ewing et al., 1998), and the sequences were assembled into clusters using the TGICL program from TIGR (Institute for Genomic Research), both incorporated into Sistema Genoma. The criteria for acceptance in the system were a minimum value of 20 in PHRED and a length of 150 bases per read (Pappas et al., 2008). Automatic annotation followed in the Basic Local Alignment Search Tool BLASTx 2.2.3 program (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov>) (Altschul et al., 1990).

## RESULTS AND DISCUSSION

After sequencing and assembly, 768 cDNA sequences were obtained that comprised 168 clusters with 126 contigs and 42 singlets. Using the Eukaryotic Orthologous Groups classification, we grouped the ESTs into 18 functional categories. Among them, 5 were more representative: 1) Energy conversion and production (5 ESTs/cluster), 2) translation, ribosomal

structure, and biogenesis (5 ESTs/cluster), 3) post-translational modification, chaperone proteins (9 ESTs/cluster), 4) prediction of general function (7 ESTs/cluster), and 5) proteins with unknown functions (4 ESTs/cluster; Figure 1).



**Figure 1.** Frequency of 18 functional categories identified from the database of functional proteins (KOG - Eukaryotic Orthologous Groups). A = modification and processing of RNA; B = structure and dynamics of chromatin; C = energy production and conversion; D = cell cycle control, cell division, compartmentalization of chromosomes; F = nucleotide transport and metabolism; G = carbohydrate transport and metabolism; I = lipid transport and metabolism; J = translation, ribosomal structure and biogenesis; K = transcription; L = replication, recombination and repair; O = post-translational modification, protein chaperones; R = general function prediction; S = function unknown; T = mechanism of signal translation; U = intracellular traffic, vesicular transport and secretion; Y = nuclear structure; Z = cytoskeleton.

The functions of several genes clustered in these categories are described in the literature for some *Gossypium* species. For example, genes associated with energy conversion and production act in secondary metabolism, and they are more expressed in cotton fiber at the initial development phase [0-5 days post-anthesis (dpa)] (Iqbal et al., 2008). Genes with functions related to translation, ribosomal structure, and biogenesis have been described by Taliércio and Boykin (2007). Structural constituents of ribosomes (ribonucleoprotein) also have an important role in the early stage of cotton fiber development, especially 0-10 dpa. Another interesting category is related to the synthesis of chaperon proteins in cotton. Studies of differentially expressed proteins in a mutant-upland cotton (*G. hirsutum* L.) carried out by Zhao et al. (2010) have shown that these proteins are involved in the process of fiber stretching.

Based on data obtained from our library, the success and novelty indices were estimated at 67 and 51%, respectively. These values are close to those found in the literature in studies of other species (Takahashi, 2005).

The analysis in the cotton and *A. thaliana* databases showed homology between some contigs (Table 2). *G. hirsutum*-CL2Contig2 and *G. hirsutum*-CL28Contig1 refer to fiber development at 0-10 and 1-3 dpa, respectively. However, in *A. thaliana*, these genes

are associated with pollen differentiation (Trionnaire et al., 2009), suggesting that they may be involved in the regulation of large event cascades related to microsporogenesis until after fruit formation. Ito et al. (2004) have reported that in *Arabidopsis*, these genes are required for the development of archesporium cells after all processes of sporogenesis until the final stage of gametogenesis. According to other authors, these genes are homologous to those associated with male sterility and are essential for normal anther development, in which absence or malfunction prevents pollen grain viability (Wilson et al., 2001; Sorensen et al., 2003; Yang et al., 2003).

**Table 2.** Genes identified from cDNA cotton bud.

Contigs	Reads	Description/GenBank accession	Homologue organism	E-value
CL2Contig2	12	Fiber (0-10 dpa)/GenBank ID: gb ES835808.1 ES835808 Pollen maturation/TAIR ID: AT5G48810	<i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	0.0 8e <sup>-15</sup>
CL1Contig2	8	Fiber (7-10 dpa)/GenBank ID: gb BF275093.2 BF275093 Catabolism of D-lactic acid methylglyoxal/TAIR ID: AT1G06130	<i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	e <sup>-150</sup> 5e <sup>-13</sup>
CL7Contig1	4	Fiber library/GenBank ID: gb CB350511.1 CB350511 Unknown protein/TAIR ID: AT4G08600	<i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	5e <sup>-34</sup> e <sup>-47</sup>
CL13Contig1	3	Ovule and fiber/GenBank ID: gb DT053700.1 DT053700 Metabolic process (beta galactosidase)/TAIR ID: AT5G01075	<i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	0.0 2e <sup>-07</sup>
CL24Contig1	2	Fiber (0-10 dpa)/GenBank ID: gb ES827796.1 ES827796 Actin depolymerizing factors/TAIR ID: AT3G46010	<i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	0.0 2e <sup>-54</sup>
CL23Contig1	2	Ovule library/GenBank ID: gb DW231944.1 DW231944 Pectinesterase inhibitor/TAIR ID: AT1G02550	<i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	0.0 5e <sup>-41</sup>
CL19Contig1	2	Immature ovule (-3 a 3 dpa) with or without fiber/ GenBank ID: gb DT048152.1 DT048152	<i>Gossypium hirsutum</i>	0.0
CL3Contig2	2	Unknown protein/TAIR ID: AT1G09610 Ovule and fiber/GenBank ID: gb DT049104.1 DT049104	<i>Arabidopsis thaliana</i> <i>Gossypium hirsutum</i>	8e <sup>-13</sup> e <sup>-160</sup>
CL3Contig1	2	Unknown protein/TAIR ID: AT1G11806 Ovule and fiber/GenBank ID: gb DT049104.1 DT049104	<i>Arabidopsis thaliana</i> <i>Gossypium hirsutum</i>	6e <sup>-9</sup> e <sup>-157</sup>
CL28Contig1	2	Unknown protein/TAIR ID: AT1G01115 Fiber (1-3 dpa)/GenBank ID: gb CB350511.1 CB350511 Pollen maturation/TAIR ID: AT2G28180	<i>Arabidopsis thaliana</i> <i>Gossypium hirsutum</i> <i>Arabidopsis thaliana</i>	9e <sup>-19</sup> 2e <sup>-24</sup> e <sup>-68</sup>

dpa = days post-anthesis.

*G. hirsutum*-CL13Contig1 was identified as a gene related to both ovule and fiber formation. These associations were not identified in *A. thaliana*, but in  $\beta$ -galactosidase metabolism, the contig is related to an enzyme essential to cell expansion and signaling of biochemical events in cells during plant development (Pauly et al., 2001). Gantulga et al. (2008) have isolated and characterized the genes At1g45130 and At3g52840, which encode Gal-5 and Gal-2  $\beta$ -galactosidases that belong to the 35-glycosyl hydrolase family, which is involved in the modification of polysaccharides from the cell wall and expressed mainly in leaves, stems, and flowers.

*G. hirsutum*-CL24Contig1 is related to fiber development at an early stage (0-10 dpa); however, in *A. thaliana*, this gene is associated with actin depolymerizing factors (Cheung et al., 2002). Studies carried out by Chen et al. (2002) in tobacco (*Nicotiana tabacum*) to examine the regulation of actin organization by actin depolymerizing factors have shown that these proteins are essential for the growth of the pollen tube and pistil. Therefore, similar to CL2Contig2 and CL28Contig1, CL24Contig1 may be associated with the regulation of pre- and post-gametogenesis events.

*G. hirsutum*-CL23Contig1 was identified in a cotton ovule library, but no specific gene was reported. In *A. thaliana*, this gene has been reported as a pectinesterase inhibitor, the expression of which can affect several physiological processes such as cell wall extension during pollen germination and tube growth (Jiang et al., 2005). *G. hirsutum*-CL1Contig2 was identified in fiber development (7-10 dpa) and the *A. thaliana* database showed that this gene plays a role in detoxifying methylglyoxal catabolism through the glyoxalase system mediated by glyoxalases I and II. *G. hirsutum*-CL1Contig2 also controls cellular differentiation and proliferation (Mustafiz et al., 2010). Studies with tobacco and rice (*Oryza sativa*) have shown that the glyoxalase pathway operates under abiotic stress - specifically, by increasing salt tolerance (Quan et al., 2010).

Although the functionality of many of these genes has been studied, that of others has not yet been defined - for example, CL19Contig1 related to immature ovules (-3 to 3 dpa), with or without fiber; CL7Contig1 related to fiber, and CL3Contig1 and CL3Contig2, both related to ovule and fiber development (see Table 2).

We constructed a subtractive library from cotton buds that efficiently identified differentially expressed genes with a success index of 67% and a novelty index of 51%. Most contigs obtained were involved in reproductive functions, such as ovule, pollen, and fiber development. These findings can be used in the identification of tissue-specific upstream regulatory sequences and thus assist in breeding programs for cotton. Additional genes with unknown function were also found and represent a valuable resource for further investigations related to buds in pre- and post-development.

## ACKNOWLEDGMENTS

Research supported by Monsanto and Conselho Nacional de Desenvolvimento Científico e Tecnológico.

## REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Chen CY, Wong EI, Vidali L, Estavillo A, et al. (2002). The regulation of actin organization by actin-depolymerizing factor in elongating pollen tubes. *Plant Cell* 14: 2175-2190.
- Chen X, Wang Z, Wang J, Wang M, et al. (2007). Isolation and characterization of Brittle2 promoter from *Zea mays* and its comparison with Ze19 promoter in transgenic tobacco plants. *Plant Cell Tissue Organ Cult.* 88: 11-20.
- Cheung AY, Chen CY, Glaven RH, de Graaf BH, et al. (2002). Rab2 GTPase regulates vesicle trafficking between the endoplasmic reticulum and the Golgi bodies and is important to pollen tube growth. *Plant Cell* 14: 945-962.
- Ewing B, Hillier L, Wendl MC and Green P (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175-185.
- Gantulga D, Turan Y, Bevan DR and Esen A (2008). The *Arabidopsis* At1g45130 and At3g52840 genes encode beta-galactosidases with activity toward cell wall polysaccharides. *Phytochemistry* 69: 1661-1670.
- Grossi-de-Sá MF, Quezado de MM, Silva MS, Silva SM, et al. (2007). Susceptibility of *Anthonomus grandis* (cotton boll weevil) and *Spodoptera frugiperda* (fall armyworm) to a cryIIa-type toxin from a Brazilian *Bacillus thuringiensis* strain. *J. Biochem. Mol. Biol.* 40: 773-782.
- Hashemi SM, Hosseini SM and Damalas CA (2009). Farmers' competence and training needs on pest management practices: participation in extension workshops. *Crop Prot.* 28: 934-939.
- Hsu CY, Jenkins JN, Saha S and Ma DP (2005). Transcriptional regulation of the lipid transfer protein gene LTP3 in cotton fibers by a novel MYB protein. *Plant Sci.* 168: 167-181.
- Iqbal S, Bashir A, Naseer HM, Ahmed M, et al. (2008). Identification of differentially expressed genes in developing cotton fibers (*Gossypium hirsutum* L) through differential display. *Electron. J. Biotechnol.* 11: 4-10.

- Ito T, Wellmer F, Yu H, Das P, et al. (2004). The *Arabidopsis* homeotic selector protein AGAMOUS controls a gene essential for microsporogenesis. *Nature* 460: 356-360.
- James C (2010). Global Status of Commercialized Biotech/GM Crops: 2010. ISAAA Brief 42-2010: Executive Summary.
- Jiang L, Yang SL, Xie LF, Pua CS, et al. (2005). VANGUARD1 encodes a pectin methyltransferase that enhances pollen tube growth in the *Arabidopsis* style and transmitting tract. *Plant Cell* 17: 584-596.
- McKinion JM, Jenkins JN, Willers JL and Zumanis A (2009). Spatially variable insecticide applications for early season control of cotton insect pests. *Comput. Electron. Agr.* 67: 71-79.
- Mustafiz A, Sahoo KK, Singla-Pareek SL and Sopory SK (2010). Metabolic engineering of glyoxalase pathway for enhancing stress tolerance in plants. *Methods Mol. Biol.* 639: 95-118.
- Neto Borges CR, Oliveira M, Labuto LBD and Castro AS (2005). A Plataforma de Genômica Funcional Utilizada como Suporte em Programas de Inovação Biotecnológica. Comunicado Técnico, Brasília.
- Pappas GJ, Miranda RP, Martins NF, Togawa RC, et al. (2008). SisGen: A CORBA based data management program for DNA sequencing projects. *Lect. Notes Comput. Sci.* 5109: 116-123.
- Park W, Scheffler BE, Bauer PJ and Campbell BT (2010). Identification of the family of aquaporin genes and their expression in upland cotton (*Gossypium hirsutum* L.). *BMC Plant Biol.* 10: 142.
- Pauly M, Qin Q, Greene H, Albersheim P, et al. (2001). Changes in the structure of xyloglucan during cell elongation. *Planta* 212: 842-850.
- Quan S, Switzenberg R, Reumann S and Hu J (2010). *In vivo* subcellular targeting analysis validates a novel peroxisome targeting signal type 2 and the peroxisomal localization of two proteins with putative functions in defense in *Arabidopsis*. *Plant Signal Behav.* 5: 151-153.
- Quintanilha A, D'êça LS and Cabral A (1962). Desenvolvimento do Botão Floral do Algodoeiro em Função do Tempo. In: Boletim da Sociedade Broteriana (Fernandes A and Neves JB, eds.). Universidade de Coimbra, Coimbra, 189-217.
- Rao MN, Soneji JR and Sahijram L (2011). Citrus: Genomics Resources Developed. In: Wild Crop Relatives: Genomic and Breeding Resources, Tropical and Subtropical Fruits (Chittaranjan K, ed.). Springer Berlin Heidelberg, New York, 50-59.
- Shelenkov A and Korotkov E (2009). Search of regular sequences in promoters from eukaryotic genomes. *Comput. Biol. Chem.* 33: 196-204.
- Showler AT (2008). Relationships of abscised cotton fruit to boll weevil (*Coleoptera: Curculionidae*) feeding, oviposition, and development. *J. Econ. Entomol.* 101: 68-73.
- Sorensen AM, Krober S, Unte US, Huijser P, et al. (2003). The *Arabidopsis* ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *Plant J.* 33: 413-423.
- Takahashi D (2005). Análise de Sequências Expressas em Raízes de Cana-de-Açúcar Colonizadas por *Gomus clarum*. Doctoral thesis, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- Taliercio EW and Boykin D (2007). Analysis of gene expression in cotton fiber initials. *BMC Plant Biol.* 7: 22.
- Trionnaire GL, Grant-Downton R, Hafidh S, Schmid R, et al. (2009). MicroRNA profiling of *Arabidopsis thaliana* mature pollen. *Plant Mol.* 1: 1-13.
- Wilson ZA, Morroll SM, Dawson J, Swarup R, et al. (2001). The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant J.* 28: 27-39.
- Wroblewski T, Tomczak A and Michelmore R (2005). Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis*. *Plant Biotechnol. J.* 3: 259-273.
- Yang SL, Xie LF, Mao HZ, Pua CS, et al. (2003). *Tapetum determinant1* is required for cell specialization in the *Arabidopsis* anther. *Plant Cell* 15: 2792-2804.
- Zhao PM, Wang LL, Han LB, Wang J, et al. (2010). Proteomic identification of differentially expressed proteins in the *Ligon lintless* mutant of upland cotton (*Gossypium hirsutum* L.). *J. Proteome. Res.* 9: 1076-1087.
- Zhou T, Zhang R, Meng ZG, Sun GQ, et al. (2008). Molecular biology research advance on cotton reproductive organ development. *J. Cotton Sci.* 3: 223-230.