Biotechnology and Molecular Biology Reviews Vol. 4 (1), pp. 011-028, February, 2009
Available online at http://www.academicjournals.org/BMBR
ISSN 1538-2273 © 2009 Academic Journals

Standard Review

Emerging trends in enhancement of cotton fiber productivity and quality using functional genomics tools

N. Manikanda Boopathi[1*] and R. Ravikesavan[2]

[1]Department of Plant Molecular Biology and Biotechnology, Coimbatore- 641003, India.
[2]Department of Cotton, Tamil Nadu Agricultural University, Coimbatore- 641003, India.

Accepted 4 January, 2009

Cotton, the most preferred natural fiber in the world, is the mainstay of global economy for several centuries. However, the fiber productivity has reached its plateau in the past decade which forced the research community to develop high-yielding and high quality cotton cultivars. In this genomics era, cotton researches focussed on two aspects: identification of genes for important agronomic traits and manipulation of such genes in view of developing elite cotton cultivar. Despite the complexity of the molecular mechanisms underlying its development, the study of the cotton fiber has become a trait of primary interest besides biotic and abiotic stress resistance. Albeit several strategies, functional genomics approach offers new unprecedented opportunities for identification of complex network of genes involved in fiber productivity and quality. Recent years have witnessed a better understanding of the plethora of genes affecting cotton fibre. Molecular, cellular and developmental changes related to fiber development have been identified through high-throughput EST projects and microarray analysis coupled with cotemporary biological tools. Despite impressive progress, the genomics and post-genomics revolution will be applicable in plant breeding only when they can elucidate the relationship between variation in phenotypic traits and the variation in gene sequences and/or expression. To this end, there is an immediate demand for integration of disciplines such as structural genomics, transcriptomics, proteomics and bioinformatics with plant physiology and breeding. Integration of multidisciplinary approaches is indispensable in upcoming cotton improvement programs since cotton is an important renewable resource that needs to be preserved for future generations.

Key words: Cotton fibre, expressed sequence tags, functional genomics, microarray, transcriptomics.

INTRODUCTION

Cotton, a high valued agricultural commodity for more than 8000 years, has long been recognized as a vital component of the global economy (Arpat et al., 2004). Cotton production provides income for approximately 100 million families and approximately 150 countries are involved in cotton import and export (Chen et al., 2007a). All parts of the cotton plant are useful and it has hundreds of uses. No other fiber comes close to duplicating all of the desirable characteristics combined in cotton. In addition to the fiber used in textile manufacturing, cotton seed is used to produce oil, seed meal (rich in essential amino acids which is lacking in most seed crops) and seed hulls (used for mulch and cattle feed). It has been estimated that 180 million people depend, either directly or indirectly, on the production of cotton for their livelyhood (Benedict and Altman, 2001).

Globally, cotton is grown on 32 million hectares (mha) with approximately 71% of the production in developing countries (FAO, 2000; http://apps.fao.org/page /collections?subset=agriculture). India, USA and China are the main producers of world production. India has larger sown cotton area (9 mha) than any country in the world

*Corresponding author: nmboopathi@tnau.ac.in.

(Clive, 2006) and produces almost three million tonnes of all qualities and staple lengths of cotton per year. The cotton industry in India has 1,543 spinning units, more than 281 composite mills, 1.72 million registered looms and an installed capacity of 36.37 million spindles (Kambhampati et al., 2005). Cotton provides a livelihood to more than 60 million people in India by way of support in agriculture, processing and use of cotton in textiles and also contributes 30% to the Indian agricultural gross domestic product and thus cotton is a very important cash crop for Indian farmers (Barwale et al., 2004). Albeit India's cotton area repre-senting 25% of the global area of cotton, it produced only 12% of world production. Yields of cotton in India are low, with an average yield of 300 kg/ha compared to the world average of 580 kg/ha (Clive, 2006). The major limiting factors to both cotton production and quality in India are biotic and abiotic stresses. As with many cotton growing areas of the world, major damage is due to insect pests, especially the bollworm complex, sucking pests and viruses. The productivity is still worsened by abiotic stress such as drought and heat. It is worth to mention here that most of the cotton in India is grown under rainfed conditions and about a third is grown under irrigation (Sundaram et al., 1999), which also experiences water stress during certain growth periods. Rising production costs to combat biotic and abiotic stresses and stagnant pricing are the additional factors that threaten cotton production. The low pricing of recent years due to poor quality of fiber resulted to biotic and abiotic stresses and has forced many growers to plant alternative crops, even in the face of farm subsidies. Hence, to cope with the growing demand on cotton fiber and by products, genetic enhancement of cotton is indispensable which will ensure competitiveness in the market of this natural-renewable product with petroleum-derived synthetic fibres, given the projected future decline in petroleum reserves. Moreover, modifications to expand the use of seed derivatives for food and feed could profoundly benefit the diets and livelihoods of millions of people in food-challenged economies (Chen et al., 2007b).

The genus *Gossypium* L. has long been a focus of genetic, systematic and breeding research and has a long history of improvement through breeding, with sustained long-term yield gains. *Gossypium* spp. consists of at least 45 diploid and 5 allotetraploid species. The allotetraploid cotton species, which include two commercially important cultivated species, *G. hirsutum* L., and *G. barbadense* L., were generated by A- and D-compound genomes (Fryxell, 1979). The best living models of the ancestral A- and D-genome parents are *G. herbaceum* and *G. raimondii*, respectively (Endrizzi et al., 1985). These four genome groups have received special attention, since they have been domesticated for their abundant seed trichomes. The diploid donor of the allopolyploid A_T genome [where the T subscript indicates the A genome in the tetraploid (AD) nucleus], was a species much like the modern *G. arboreum* or *G. herbaceum*, whereas the allopolyploid D_T genome is derived from a progenitor similar to the modern *G. raimondii* species. These well-established relationships provide a phylogenetic framework to investigate the evolution of gene expression both in terms of domesticated fiber production and polyploidy (Udall et al., 2007). Understanding the contribution of the A and D subgenomes to gene expression in the allotetraploids may facilitate improvement of fiber traits (Chen et al., 2007a).

Interestingly, the A genome species produce spinnable fiber and are cultivated on a limited scale, whereas the D genome species do not (Applequist et al., 2001). More than 95% of the annual cotton crop worldwide is G. *hirsutum* (Upland or American cotton) and *G. barbadense* (the extra-long staple or Pima cotton) which accounts for less than 2% (National Cotton Council, 2006. http://www.cotton.org). Yield and fiber quality of upland cotton varieties have declined over the last decade – a downward trend that has been attributed to erosion in genetic diversity of breeding stocks and an increased vulnerability to environmental stresses (Meredith, 2000). The level of genetic diversity is low in *G. hirsutum*, especially among agriculturally elite types, as revealed by all means of assessment (Lacape et al., 2007a).

Increasing diversity is therefore essential to genetic improvement efforts. Each of the three major approaches to increasing genetic diversity that is, mutagenesis, germplasm introgression and transformation has advantages and disadvantages. Interspecific germplasm introgression is particularly attractive in that it utilizes a broad germplasm base, can be targeted to one or more specific traits/genes or modulated to include thousands of genes/even entire genomes and is readily coupled to marker-assisted genome analysis and selection (Saha et al., 2006). Though, quantitative trait loci (QTL) mapping and marker aided selection has potential application in genetic improvement of cotton for higher productivity, its application is not yet documented in cotton breeding program due to poor knowledge on physiological and genetic nature of fibre quality and productivity traits, low and complex heritability of investigated traits, genotype X environment interactions etc., (Lacape et al., 2007b). Although introgression of genes across species boundaries is difficult, it is quite desirable because the gene pools of cultivated species do not contain all of the desired alleles. Alternatively, mutagenesis and transgenic technology has been proposed. However, currently they have limited applications due to several technical reasons such as non availability of novel genes, lack of efficient method to alter/transfer large genetic element etc., (Wilkins et al., 2000). Thus, the paucity of information about genes that control important traits impedes the genetic improvement of cotton.

Fiber represents over 90% of the total value of the cotton crop and the genetic improvement of fiber properties is certainly a major target trait besides biotic and abiotic

stress resistance since these stresses also ultimately affect the final productivity and quality of fiber. The cotton fiber is a complex biological system that is the net result of the intricate interplay of elaborate developmentally regulated pathways consisting of literally thousands of genes and gene products and as such has been a difficult subject to tackle using conventional approaches, especially as cotton does not lend itself easily to genetic analysis (Taliercio and Boykin, 2007). The high value per hectare of cotton and global textile market demand for increased fiber uniformity, strength, extensibility and quality clearly justify the importance of new and innovative approaches towards evaluating and under-standing genetic mechanisms of fiber qualities (Saha et al., 2006). Gene discovery and strategies to produce the desired pattern of expression and phenotype are major goals that lay ahead for biotechnology programs. In this context, the use of genomic tools and resources to facilitate breeding using molecular approaches (Wilkins and Arpat, 2005) and characterization of the cotton fiber transcriptome (Arpat et al., 2004) are considered as key strategies for the genetic improvement of cotton. Besides its economic importance, cotton fiber is an outstanding model for the study of plant cell elongation, cell wall and cellulose biosynthesis (Kim and Triplett, 2001). The fiber is composed of nearly pure cellulose, the largest component of plant biomass. Compared to lignin, cellu-lose is easily convertible to biofuels. Translational genomics of cotton fibre and cellulose may lead to the improvement of diverse biomass crops (Chen et al., 2007a). Thus, functional genomics of cotton fiber development has several folds of applications.

SCOPE OF FUNCTIONAL GENOMICS

It has been only in the last decade that researchers have begun to focus on studying the underlying developmental mechanisms that control fiber properties as the basis for manipulating biological and cellular processes to improve fiber characteristics. The genetic complexity of the cotton fiber transcriptome is currently estimated to consist of approximately 18000 genes in the genome of cultivated diploid species. The cotton fiber transcriptome in allotetraploid species is similarly estimated at approxima-tely 36000 genes and to include homoeologous loci from both the A_T and D_T genomes (Arpat et al., 2004). The high genetic complexity of the fiber transcriptome in both diploid and tetraploid species accounts for a significant proportion (45 - 50%) of all the genes in the cotton genome (Wilkins and Arpat, 2005). However, despite approximately 1.5 million years of evolution following the polyploidization event, polyploidy has not been accompanied by rapid genome change, as the genome organization and gene sequences of orthologous loci from the A and D genomes of diploid and tetraploid cotton species are highly conserved (Senchina et al., 2003; Rong et al., 2004). Moreover, spatial and temporal expression patterns have been evolutionarily conserved (Cedroni et al., 2003). Thus, fiber gene function is highly conserved in the genomes of wild and cultivated species, as well as diploid and tetraploid species, despite millions of years of evolutionary history. The phenotypic variation in fiber properties therefore is more likely one of the quantitative differences in gene expression as opposed to differences in the genotype at the DNA level (Wilkins and Arpat, 2005). Further studies, hence, are required to understand the genes, their copy number and specific function in fiber development.

The direct or indirect aims of all functional genomics programs focusing on fiber development are to define the function of the genes involved and to find candidate genes to improve fiber productivity and quality. This helps to sieve a smaller subset of genes which can be employed to draw up a final list of candidate genes (or master regulators) that could be used in future innovative breeding program for superior fiber productivity and quality (Chen et al., 2007a). Several efforts have led to the creation of the first fiber model (Wilkins and Jernstedt, 1999) that serves as the current framework for cotton biotechnology programs worldwide to alter the timing, rate, and/or duration of fiber elongation.

As for methods of cloning fiber-related genes, differential screening of the fiber cDNA library (DD-RT PCR) was most popular in the earlier days. At the same time, sequencing randomly selected cDNA clones from the fiber library, PCR amplification using gene-specific probes and suppression subtractive hybridization were shown to be useful in isolating cotton-fiber-related genes. Recently, expressed sequence tag (EST) projects and microarray technology have gained recognition as means for discovering novel genes. The results were confirmed by (reverse) northern blotting and/or quantitative real time PCR techniques. Thus techniques to monitor global gene expression rely either on hybridisation (microarrays) or on PCR (real-time PCR, cDNA-AFLP and differential display) or, more recently on massively parallel signature sequencing (http://mpss.udel.edu) based methods. However, it is noteworthy to mention that a combination of different profiling methods is likely to be most informative (Jansen and Nap, 2001), since no method is perfect and each one of them has its own pros and cons. At the time of this writing, microarrays are considered as a powerful method to simultaneously measure relative expression levels for thousands of genes. Two main types of microarrays have been developed in cotton, namely cDNA-based and oligonucleotide-based arrays. Based on their technical specifications, cDNA and oligonucleotide microarrays may differ in sensitivity and dynamic range for detecting variation in mRNA abundance as well as in power to discriminate between related target sequences (Wilkins and Arpat, 2005).

Limitations and constraints of the different microarray platforms (Mah et al., 2004; Rensink and Buell, 2005) or in comparison with PCR-based techniques have been

emphasized in several instances (Reijans et al., 2003; Tan et al., 2003). The principal limitation for the microarray technique in fiber transcriptome analysis, irrespective of its types, is that only a fraction of genes, those for which DNA sequence are available can be investigated.

As of 1st September, 2008, 369872 *Gossypium* sequences were in GenBank, including 39232 ESTs from *G. arboreum* (A), 63577 from *G. raimondii* (D), 265793 from *G. hirsutum* (AD tetraploid), 1023 from *G. barbadense* and 247 from *G. herbaceum* var. *africanum* (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). Non-redundant ESTs have been used both to develop sequence - specific markers for genetic mapping and to construct microarrays for the identification of candidate genes involved in fiber cell initiation and elongation (Arpat et al., 2004; Lee et al., 2006; Shi et al., 2006; Wu et al., 2006; Udall et al., 2007). For example, Udall et al. (2007) created a long oligonucleotide microarray for cotton because of its low manufacturing cost, flexibility in design, homogeneous melting temperatures (Tm) and relative ease of adding probes. A small EST assembly (~ 45,000 ESTs) was previously used to generate oligonucleotide probes for cotton fiber (Arpat et al., 2004). A larger scale EST assembly (> 150,000 ESTs) was recently produced as a community-wide effort by cotton researchers (Udall et al., 2006). Subsequent additions of cotton ESTs to Genbank (> 210,000 ESTs) have been compiled into a large EST assembly (TIGR Cotton Gene Index: http:// compbio.dfci.harvard.edu/tgi). These two assemblies constitute nearly all of the known genic sequence from cotton. An increasing number of sequence resources (bacterial artificial chromosomes (BACs) and ESTs) in *Gossypium* have been used to design fiber cDNA microarrays for functional studies by another group of scientists (Zhang et al., 2001; Wu et al., 2006). Recent efforts from the cotton community lead to the public release of 2 cotton microarrays (essentially from fiber ESTs), including a 24 K GeneChip® Cotton Genome Array from Affymetrix (http://www. Affymetrix.com /products/arrays/specific/cotton.affx); and a 23 K oligonucleotide (60-70mer) microarray by Udall et al. (2007).

Two essential considerations of microarray quality include the number of targeted genes and the broad utility of the microarray for specific tissues or treatments. Regarding the first consideration, the 22,778 genes described by Udall et al. (2007) include perhaps 46 - 60% of the total genic diversity, given that the total number of genes in the cotton genome may be approximately 40,000 – 50,000 (Hawkins, 2006). Regarding the second consideration, a detailed analysis of the probes revealed that ~ 7,300 probes represented genes expressed in specific tissues or under specific conditions (Udall et al., 2007). These two considerations suggest that the oligonucleotides selected for the cotton oligonucleotide microarray have a broad diagnostic utility while potentially targeting tissue specific transcripts expressed under a variety of conditions. It is worth mentioning that the sequences and annotations of all the probes are publicly available via a web-based query (http://cottonevolution.info) or by request.

Molecular, cellular and developmental aspects of fiber development

Cotton fibers are seed hairs and they originate from the epidermal cells of the ovular surface. It is well known that fiber development is composed of four overlapping stages: fiber cell initiation and enlargement from -3 to 1 day postanthesis (DPA), fiber elongation after anthesis until 25 DPA, secondary cell wall cellulose deposition from 15 DPA to 50 DPA and fiber cell dehydration and maturation after 45 DPA (Basra and Malik, 1984). Thus, the near-synchronous growth of 500,000 terminally differentiated single-type fiber cells per ovule (Havov et al., 2008) is characterized by four major discrete developmental stages – differentiation/initiation, expansion/elongation, primary cell wall (PCW) synthesis, secondary cell wall (SCW) synthesis and maturity (Wilkins and Jernstedt, 1999). Among these developmental stages the productivity and quality of cotton depends mainly on two processes: fiber initiation, which determines the number of fibres present on each ovule and fiber elongation, which determines the final length and strength of each fiber. As all plant cells undergo cell expansion to some degree during growth and development, rapidly elongating cotton fibres offer a unique single-celled model to study;

(i) The molecular and cellular mechanisms that regulate the rate and duration of cell expansion, and hence, govern cell size and shape, and in the case of cotton, important agronomic traits as well (Wilkins and Arpat, 2005)
(ii) Cell wall development.
(iii) Cellulose biosynthesis (John and Crow, 1992).

The type of growth mechanism (diffuse versus tip) associated with the exaggerated growth rate of rapidly elongating cotton fibers has been debated for decades (Seagull, 1990; Wilkins and Jernstedt, 1999). However, structural and physiological data provide compelling evidence for diffuse growth (Tiwari and Wilkins, 1995). Interestingly, no genes known to be specific to tip-growing cell types (for example pollen and root hairs) have been identified in developing cotton fibres, further supporting a diffuse-growth mechanism during rapid polar elongation (Wilkins and Arpat, 2005). Polar elongation of developing cotton fibres via diffuse growth is controlled at the cellular level by the transverse orientation of microtubules, the sites of cell wall loosening, the cell wall deposition in the extracellular matrix and polar vesicular trafficking (Cosgrove, 2001; Tiwari and Wilkins 1995). Fiber length is dictated by the rate and duration of cell ex-

pansion, which is in turn, governed by developmental programs that co-ordinately regulate cell turgor, the driving force of cell expansion and cell wall loosening (Ruan et al., 2001). The transition from PCW to SCW synthesis, which occurs during the latter stage of expansion between ~ 16 and 21 DPA, is distinguished by the re-orientation of microtubules and cellulose micro fibrils to steeply pitched helical arrays (Seagull, 1992) in anticipation of SCW synthesis to produce a thick cell wall consisting of > 94% cellulose. Coincident with the termination of fiber elongation at ~ 21 DPA is an increase in fiber strength (Hsieh, 1999), presumably due to cross-linking of cellulosic and non-cellulosic matrices (Carpita and Gibeau, 1993), that is also accompanied by a major loss (~ 36%) of high molecular weight noncellulosic polymers in the PCW (Shimizu et al.,1997). Developmental programs regulate the temporal synthesis of fiber PCW and SCW, which differ significantly in structure and composition. While the thin PCW (0.2 - 0.4 µm) deposited during fiber elongation contains < 30% cellulose, the thick SCW (8 - 10 µm) is composed of > 94% cellulose (Meinert and Delmer, 1977). In addition, the degree of polymerization of cellulose micro fibrils also varies, being < 5,000 in PCW and ~ 14,000 in SCW (Marx-Figini, 1966). In some domesticated varieties, cotton fibers may attain a final length of 6 cm or about one-third the height of an entire *Arabidopsis* plant (Kim and Triplett, 2001).

Thus, rapid and simultaneous elongation occurs in millions of fiber cells in the cotton boll without concurrence of cell division and multicellular develop-ment. On the day of anthesis (flower opening), approximately one in four epidermal ovular cells has already been destined to become a cotton fiber, initially appearing as a spherical protrusion. Only about a third of all the epidermal cells become fibres (Berlin 1986), although the exact proportion varies between genotypes and in response to hormone levels (Gialvalis and Seagull 2001; Rahman 2006). A greater understanding of the molecular processes that regulate which cells become fibres and molecular and physiological mechanisms of fiber development could increase the ability to either breed for, or engineer, cotton plants with a higher density of fibres and hence, a higher yield.

Functional genomics and fiber development

The cotton fiber transcriptome has attracted a lot of attention in recent years (Wilkins et al., 2007). Many fiber-specific genes involved in fiber cell initiation, fiber elongation or cell wall biogenesis have been identified from the comparisons of normal (wild-type) versus fiber mutants of *G. hirsutum* species. Few reports have also investigated the mechanisms and genes underlying the important developmental differences between *G. hirsutum* and *G. barbadense* (Ruan et al., 1997; Wu et al., 2005; Wu et al., 2006).

Since the first report of John and Crow (1992), who had had cloned the E6 gene through differential screening of a fiber cDNA library in 1992, several reports have shown that many genes were expressed preferentially in cotton fibers (Wilkins and Arpat, 2005). Among these cotton fiber genes, some are highly expressed during early fiber development, some are predominantly expressed during fiber SCW deposition and some show high expression during the entire fiber development. Like E6, H6 and B6, their exact functions are not clear; however, the primary structures of putatively encoded proteins, developmental regulation and tissue specificity suggest that they are likely important for fiber development. On the other hand, some genes have definite functions in cotton fiber development for example, *GhExp1*, which specifically accumulates in developing cotton fibres, encodes a cell wall protein and regulates cell wall loosening by the disruption of non-covalent bonds between wall components (Harmer et al., 2002).

Most of the cotton fiber transcriptome was identified by a single gene discovery project (Arpat et al., 2004; http://cfgc.ucdavis.edu/) and numerous studies targeted the rapidly elongating cotton fibres for a number of the following compelling reasons:

i The rate and duration of fiber elongation governs important agronomic properties, such as yield and fiber length,
ii The exaggerated growth of elongating fibres, underpinned by high levels of metabolic activity was expected to be especially gene rich from a gene discovery perspective
iii Isolation of fibres free of contamination from surrounding complex tissues permit an unambiguous look at the transcriptome of a single cell within a biologically relevant framework (Wilkins and Arpat, 2005).

Two general approaches undertaken to maximize gene discovery, included ‘‘deep’’ sampling of a high-quality fiber cDNA library to generate ESTs and sequential rounds of normalization to remove highly redundant gene sequences and thereby identify rare gene transcripts.
However, the judicious selection of a cultivated diploid species (*G. arboreum* L.) as a model for fiber development proved especially rewarding, as the rate of gene discovery was enhanced at least two-fold simply by avoiding redundancy due to polyploidy. To define the cotton fiber transcriptome, more than 46,000 ESTs from rapidly elongating fibers were generated from *G. arboreum* L. (Arpat et al., 2004). The transcriptome of cotton fiber across a developmental time-course, from a few days post anthesis through PCW and SCW and maturation stages were evaluated. This revealed that there were dynamic changes in gene expression between PCW and SCW biogenesis. Among them, transcripts for cell wall structure and biogenesis, the cytoskeleton and energy/carbohydrate metabolism were the three major functioning groups during the rapid elongation of fiber cells (Arpat et al., 2004; Wilkins and Arpat, 2005).

It is also important to identify the genes that are preferentially expressed in fiber tissues of *G. hirsutum* L., which is highly adapted to the present environment and the most widely cultivated species. Further, similar biochemical pathways may have diverged during evolution and thus can create a possible drawback for the identification of *G. hirsutum* fiber specific candidate genes when the knowledge from *G. arboreum* EST projects is translated (Salentijn et al., 2007). Hence, ESTs were developed from fast elongating fiber of *G. hirsutum* (Shi et al., 2006). Results from these studies were useful to co-clude that the genes are very much essential for fiber development in commercial cultivars. Detailed sequence comparisons showed that significant sequence divergence exists between the two species (Shi et al., 2006).

Transcriptomic model for fiber development: A proposal

It has been indicated elsewhere that fiber development is a highly complex metabolic process and it involves the expression of nearly half of the cotton genome for complete development of matured fiber (Havov et al., 2008). Based on the results of several studies the following transcriptomic model for fibre development is proposed. This model may be considered as preliminary since detailed functional and physiological studies yet to be performed to get a lucid understanding on this model. It should be noted that many of the genes up-regulated in the fiber initials have been identified as either unclassified or unknown (Arpat et al., 2004; Wilkins et al., 2007; Havov et al., 2008) could fuel future research and hence a clearer picture on this model with respect to fibre cell initiation, expansion and cell wall growth in cotton can be expected in the near future. More than half of all genes were up-regulated during at least one stage of fibre development. Genes implicated in vesicle coating and trafficking were found to be over expressed throughout all stages of fiber development, indicating their important role in maintaining rapid growth of this unique plant cell (Havov et al., 2008). The following sections describe the developmental stage in the specific expression of genes and their putative role in fibre development.

Initiation

A unique feature of cotton seed development is that ~ 30% of the ovule epidermal cells initiate into fibers from the outermost layer of integument at anthesis (Ruan et al., 2001). On the day of anthesis the cotton fiber initial cells swell out from the ovule surface and so are clearly distinguished from adjacent epidermal pavement cells. Since impressive progress has been made only on the later elongation stage (Ji et al., 2003; Arpat et al., 2004; Shi et al., 2006), it resulted in a rudimentary understanding of the molecular events at the early initiation and cell expansion stage. Fortunately, a small set of cotton mutants that lack seed fibres (Li et al., 2002; Lee et al., 2006; Wu et al., 2006) and contemporary biological dissecttion methods (Wu et al., 2007) have facilitated the discovery of a few genes expressed early during fiber development.

Transcript profiling and ovule culture experiments both indicate that several phytohormones mediate cotton fiber initiation. Auxin and gibberellins appear to promote early stages of fiber initiation. Ovules cultured *in vitro* become competent to produce fiber in response to auxin and giberellic acid (Graves and Stewart, 1988). Fiber initiation also requires brassinosterol production (Sun et al., 2005). Interestingly, the phytohormone-related genes were induced prior to the activation of MYB-like genes, suggesting an important role of phytohormones in cell fate determination (Chen et al., 2007b). However, several attempts have been made to alter the expression of genes involved in auxin and cytokinin biosynthesis in the fibers but no favourable phenotypic changes were observed in the resultant transgenic plants (John, 1999). This indicates complicate interactions of genes and phytohormones during fiber initiation. Fiber differentiation is evident *in vivo* by -1 DPA when microtubules reorient in epidermal cells destined to differentiate into fibers (Ryser, 1999). By 1 DPA, fiber initials bulge from the surface of the ovule. Protein biosynthesis and nucleoli size increase in very young fibers (Van't Hof, 1998). *In vitro* cultured ovules indicated that mRNA synthesis is required for fiber initiation up to 2 DPA and the ovules remained competent to initiate fibers up to 5 DPA (Triplet, 1998). Thus, the period of fiber initiation ends at 2 DPA and may extend to 5 DPA.

Genes that peak in expression during fiber initiation then decrease in expression during elongation would be expected to play a specific role in fiber initiation. Several well annotated genes with a fiber initiation-specific pattern of expression give potentially new insight into fiber initiation. Synchronously differentiating fibers represent a valuable developmental model to determine how developmental signals are integrated to control differentiation and elongation of fiber and how these signalling pathways differ between ovular and leaf trichomes. Fiber initiation requires transcription of several new genes and therefore transcription factors are likely to play an important role in fiber initiation. The Myb109 and Myb2 transcription factors are expressed in fiber initials (Suo et al., 2003). Evaluation of fiber less cotton mutants has identified genes differentially expressed in very young fiber, including transcription factors (Lee et al., 2006). The Myb2 transcription factor is able to comple-ment *Arabidopsis thaliana* trichome mutants and activate expression of R22-like (RDL) gene expressed in fiber initials (Wang et al., 2004). Additionally, the RDL gene along with genes involved in cell structure, long chain fatty acid biosyn-thesis and sterol biosynthesis have been identified as those absent or reduced in a fiber less mutant of cotton (Li et al., 2002). Most of these genes are expressed in 1 DPA ovules. A second round of fiber initiation occurs that produces the

short linters or fuzz fibers. Yang et al. (2006) deposited a large number of ESTs into the Genbank database from cDNA libraries of whole ovules spanning the period from -3 to +3 DPA augmenting those already collected by Wu et al. (2005). They were able to identify a large number of transcription factor genes (as expected) orthologous to genes in *Arabidopsis*, some known to be involved in *Arabidopsis* trichome formation, but have yet to localise expression of any of these genes specifically to cotton fibers or fiber initials.

It has also been documented that the expression of both Myb109 and Myb2 transcription factors were abundant in 1 DPA fibers and persisted into the elongation stage of fiber development (Taliercio and Boykin, 2007). Of the approximately 624 putative tran-scription factors represented on the microarray, five were regulated similarly to Myb109 and Myb2 and therefore were candidates to play a role in controlling fiber initiation. Further, Taliercio and Boykin (2007) have also identified five transcriptt-ion factors with a similar pattern of expression that could play a role in fiber development. One of the notable examples is two genes similar to CAPRICE/TRIPCHON (CPC). CPC acts as a negative regulator of trichome development in *Arabidopsis* (Schellmann et al., 2002). One of the putative CPC genes was down regulated in 1 DPA fiber compared to ovules. The inhibitors described for *Arabidopsis* are not down regulated in trichomes; therefore it is not possible to draw a conclusion based on gene expression about which putative CPC gene in cotton was more likely involved in fiber development. If CPC genes in cotton act as inhibitors of fiber initiation, reducing expression of these genes with interfering RNAs would be expected to increase the number of fibers (Taliercio and Boykin, 2007). Therefore a transgenic cotton line with reduced CPC expression could be agronomically valuable.

Wu et al. (2007) used laser capture micro-dissection coupled with cDNA microarray and found that except for a few regulatory genes, the genes that are up-regulated in the cotton fiber initials relative to epidermal cells predominantly encode proteins involved in generating the components for the extra cell membrane and PCW, carbohydrates and lipids needed for the rapid cell expansion of the initials. Overall, there were few genes that differed markedly between the fiber initials and the epidermal pavement cells- the most different being threefold to fivefold up- or down-regulated. The majority of differential genes were only about twofold more or less expressed in fiber initial cells. The absence of large differences may have been because the individual cell types were not completely pure. This has been a criticism of laser capture micro-dissection methodology, particularly with rigid plant tissues (Day et al. 2005), but does not contradict the enormous power of this technology.

Amongst the classes of genes expressed in the initials, many of the genes were involved in DNA metabolism despite cell division having ceased in these differentiated cells (Wu et al., 2007). Various authors have suggested that cotton fiber cells may undergo endoreduplication to amplify their DNA content to support their specialised function like *Arabidopsis* leaf trichomes (Szymanski and Marks 1998). Endoreduplication in cotton fiber development, however, remains a contentious topic. Some earlier reports suggested a limited increase in DNA content in fiber nuclei (Van't Hof, 1999) while others could detect no difference in DNA content of 14 - 25 DPA fibers (Taliercio et al. 2005). Wu et al. (2006) concluded that fiber initials at 0 DPA undergo at least one round of endoreduplication, so the abundance of DNA metabolism genes is not surprising. Others have suggested that the increase in DNA content may be associated with enlargement of the nucleolus (Kim and Triplett 2001).

In another study, Taliercio and Boykin (2007) had confirmed that an increase in the endoplasmic reticulum occurred in fiber initials on the day of anthesis (Ryser, 1999) and persisted through 3 DPA. In addition to that, they have identified consistent increase in membrane associated component proteins which played a role in early fiber development. They have also noticed genes associated with novel regulation of brassinosterols, GTP mediated signal transduction (which play roles in vesicle trafficking) and cell cycle control and components of Ca^{2+} mediated signalling pathway. They argued that the presence of more Ca^{2+} in fiber initials than other ovule cells and the differential expression of calmodulin and calmodulin binding proteins indicated a role for Ca^{2+} in fiber development. It is likely that the marked and long lasting increase in endoplasmic reticulum in fiber initials was unique to the ovular trichomes, indicating an early departure between the developmental programs that give rise to ovular and leaf trichomes. This increase in endoplasmmic reticulum was consistent with the increase in golgi bodies reported in fiber initials (Taliercio and Boykin, 2007). Abundant endoplasmic reticulum may play a role in biosynthesis and transport for components of the rapidly expanding cell membrane, cell wall and cuticle. Indeed, analysis of genes differentially regulated during fiber initiation and elongation identify numerous genes associated with these developmental pathways. Hence an increase in the endoplasmic reticulum may represent the first stages of fiber elongation since increase demands for cell membrane, PCW, and cuticle production which persist through the elongation phase of fiber development.

Genes other than transcription factors can have profound affects on expression of other genes. Expression of some other types of regulatory genes increased in 1 DPA fibers and persisted in 10 DPA fibers. Examples include receptor kinases, calmodulin, calmodulin binding proteins and lumen receptors. Increased expression of a calmodulin gene unique to fibers and differential expression of calmodulin binding proteins were also observed (Taliercio and Boykin, 2007) as indicated above. It seems likely that a calmodulin mediated signaling pathway exists that ei-

ther causes or responds to the redistribution of calcium into the endoplasmic reticulum. Interestingly, deesterified pectins increase in fiber initials (Turley and Vaughn, 1999). Deesterified pectins bind calcium; therefore it is likely that the cell walls may also compete for Ca^{2+}. Manipulating expression of the calmodulin or manipulating calcium levels *in vitro* may determine whether a calcium mediated pathway exists that causes or responds to the increase in endoplasmic reticulum and what role a calmodulin mediated response to Ca^{2+} plays in fiber development.

Documentation of common set of genes for fiber initials across studies has been shown only in few cases and this may be partly due to different type of experimental materials and protocols used in various studies. Fiber cell determination is believed to occur a few days before there are any visual changes in the fiber initials (Ramsey and Berlin 1976), so ideally microarray comparison should be carried out at -2 to -3 DPA, but at this stage all the cell types look identical. As more early fiber genes and promoters are continuously characterised, a combination of laser capture micro-dissection and transgenic plant technologies to tag initiating fiber cells with non-destructive visual markers, such as green fluorescent protein (GFP) will eventually allow the profiling of gene expression changes occurring up to 3 days before anthesis when there are no outwardly visible differences between cells starting to differentiate into fibres and their adjacent non-fiber epidermal cells (Wu et al., 2007).

Elongation

Many important fiber traits such as the length, shape, structure and composition of the fiber cell, are determined during this stage of fiber development. Genetic control of fiber size and shape, and hence agronomic properties is governed by the rate and duration of fiber expansion and elongation which in turn, positively correlates to fiber length (Wilkins and Arpat, 2005). Although some progresses have been made in recent decade (Xu et al., 2007), little is known on how these genes regulate fiber elongation. During the most active elongation period (5 to 20 DPA), vigorous cell expansion with peak growth rates of > 2 mm/day are observed in upland cotton (Ji et al., 2003).

As discussed earlier, fiber elongation occurs by a diffuse growth mechanism (Tiwari and Wilkins, 1995). Many genes expressed during the elongation stage of fiber differentiation relate to cell expansion, cell wall loosening and osmoregulation. Ovule culture studies confirmed a role for brassinosterols during fiber elongation in addition to fiber initiation (Sun et al., 2005). Based on genomic, genetic, molecular, biological and physiological studies it was found that ethylene plays a major role in promoting cotton fiber elongation. The role for ethylene in fiber elongation was confirmed when longer fibers were obtained with the addition of ethylene to ovule culture.

Furthermore, ethylene may promote cell elongation by increasing the expression of sucrose synthase, tubulin and expansin genes (Shi et al., 2006). Although a review on cotton fiber (Kim and Triplett, 2001) states that cytokinins, abscisic acid and ethylene inhibit fiber development, this statement was based on experiments that did not in fact include ethylene in the ovule culture studies (Shi et al., 2006). However, the role of brassinosteroid in fiber cell elongation appears to be less prominent than that of ethylene, for it was only modestly effective. Although the actions of brassinosteroid and ethylene on fiber elongation are not interdependent, it appears that they do not act completely independently either and each hormone positively modulates the synthesis of the other. Such positive interactions between the two hormones potentially contribute to the extreme elongation of fiber cells (Shi et al., 2006).

Each cotton single fiber cell elongates from 10 to 15 mm up to 2.5 to 3.0 cm by ~ 16 DPA before it switches to SCW cellulose synthesis (Tiwari and Wilkins, 1995). The rapid fiber elongation is believed to be driven by high turgor with a highly extendable PCW (Ruan et al., 2001). Cell turgor in plants is achieved largely through the influx of water driven by a relatively high concentration of osmoticum within a cell (Cosgrove, 1997). The accumulation of osmoticum into fibers may be coupled with the transmembrane proton gradient, because the plasma membrane H^+-ATPase gene is expressed strongly during the rapid phase of fiber elongation (Smart et al., 1998). This H^+ pump could also acidify the apoplast for cell wall loosening (Cosgrove, 1997). At this time, the expansin gene of major importance in mediating cell wall extension (Cosgrove, 1997), is expressed in fibers, although its temporal expression pattern over the elongation period is not clear (Orford and Timmis, 1998). The results of Ruan et al. (2001) provide a remarkable explanation of how the gating plasmodesmata and the expression of genes for solute import and cell wall loosening are developmentally coordinated to potentially control single-cell elongation. Fiber plasmadesmata are initially opened (0 - 9 DPA) but closed at ~ 10 DPA and reopened at 16 DPA and it is accompanied by a gradual switch from simple to branched forms of plasmadesmata. However, it is difficult to assign functional implications for such a structural change and mechanism responsible for the reversible gating of fiber plasmodesmata (Ruan et al., 2001). Coincident with the transient closure of the plasmadesmata, the sucrose and K^+ transporter genes are expressed maximally in fibers at 10 DPA with sucrose transporter proteins predominantly localized at the fiber base. Consequently, fiber osmotic and turgor potentials were elevated, due to increased accumulation of soluble sugars, driving the rapid phase of elongation. The higher turgor can be maintained in fibers due to the closure of plasmadesmata. The expansin gene was highly expressed at the early phase of elongation (6 to 8 DPA) and decreased rapidly afterwards. Given the increased rigidity of the fiber

cell wall, indicated by the decreased expression of expansin, it is almost certain that the higher turgor in fibers at this stage plays a critical role in driving the rapid-fiber elongation (Ruan et al., 2001). This shows that fiber elongation is initially achieved largely by cell wall loosening and finally terminated by increased wall rigidity and loss of higher turgor. This has also been shown in another study that an increase in fiber K^+ concentrations, most probably leading to higher turgor, has yielded longer fibers with higher quality (Cassman et al., 1990).

In an another study, a group of genes related to turgor regulation and the cytoskeleton such as the plasma membrane proton-translocating ATPase, vacuole-ATPase, proton-translocating pyrophosphatase (PPase), phosphoenolpyruvate carboxylase, major intrinsic protein (MIP) and α-tubulin were found to be involved in cotton fiber elongation (Smart et al.,1998). Similarly, four genes viz; putative gibberellin-regulated protein, putative tonoplast intrinsic protein, putative plasma membrane intrinsic protein and putative membrane protein were identified in the 10 DPA fiber subtracted library and they were found to be expressed during early fiber development (Liu et al., 2006). The role of these genes in fiber elongation is yet to be tested albeit their biochemical activity is established in some other studies.

The translation elongation factor 1A, *eEF1A*, plays an important role in protein synthesis, catalyzing the binding of aminoacyl-tRNA to the A-site of the ribosome by a GTP-dependent mechanism. To investigate the role of *eEF1A* for protein synthesis in cotton fiber development, nine different cDNA clones encoding eukaryotic translation elongation factor 1A were isolated from *G. hirsutum* fiber cDNA libraries (Xu et al., 2007). The isolated genes (cDNAs) were designated as nine cotton elongation factor 1A genes viz., *GhEF1A1* - *GhEF1A9* and all of them share high similarity at nucleotide level (71 - 99% identity) and amino acid level (96 – 99% identity) with conserved GTP-binding domain in the deduced amino acid sequence. The different members of *eEF1A* gene family in the plant may originate from a series of gene duplications during evolution. Of the nine *GhEF1A* genes, five are expressed at relatively high levels in young fibers. Further analysis indicated that expressions of the *GhEF1As* in the fibre are highly developmental-regulated suggesting that protein biosynthesis is very active at early fiber elongation to meet the rapid elongation of fiber cells in cotton (Xu et al., 2007).

Determining protein sequences and their expression changes in defined growing stages will further enhance understanding on fiber development mechanisms. In a proteomics study, cytoskeleton-related proteins, enzymes involved in flavonoid biosynthetic pathway, putative S-adnosylmethionine synthases, allergin like proteins and annexin were identified besides few unidentified proteins (Yao et al., 2006). Among them, the cytoskeleton related proteins are responsible for directing polar expansion during the rapid elongation period which contributes significantly to fiber shape (Arpat et al., 2004).

When cotton fiber cells underwent maximal expansion (12 - 15 DPA), transcripts of several genes except for PPase accumulated to the highest levels, then declined at the beginning of SCW deposition. *GhGLP1* transcripts coding for a germin-like protein, accumulated to their highest levels during the period of fiber expansion, followed by a sharp decline when the rate of cell expansion decreased (Kim and Triplett, 2001). While germins and GLPs appear to be involved in defence mechanisms in some plants, both biotic and abiotic stresses down-regulated the expression of *GhGLP1*. Expression of two other genes associated with fiber cell elongation, expansin and α-tubulin 1, also declined when the same abiotic and biotic stresses were applied (Kim and Triplett, 2001). Numerous functions have been proposed for dicot GLPs. However, to date, there is little direct evidence for how these proteins function *in vivo*. The association of maximal *GhGLP1* expression with stages of maximal cotton fiber elongation suggests that some GLPs may be important for cell wall expansion (Kim and Triplett, 2001).

In developing cotton fibers, there is one major expansin and several minor isoforms, suggesting that the isoforms may perform specialized roles during fiber elongation (Arpat et al., 2004). Indeed, suppression of expression of all isoforms results in a shorter, coarser fiber (Wilkins and Arpat, 2005). An interesting possibility for such specialized roles is the potential for discrete subcellular localization of the various isoforms, such that the major isoform is preferentially targeted to the tip of the cell and functions primarily in polar elongation, while the minor species are differentially targeted to regions of the fiber likely to promote lateral expansion of the fiber. It would therefore be very interesting to determine if the various expansin isoforms are localized to defined regions of elongating fibers at the subcellular level. Moreover, major isoforms for functionally important genes may well account for the bulk of genetic variability asssociated with major QTL for fiber quality and as such provide candidate genes for genetic mapping and marker-assisted selection (Wilkins and Arpat, 2005).

Expression profiles that compare transcript abundance during and after fiber elongation revealed that ~ 20% of the fibre transcriptome, including large numbers of metabolism related genes and cell wall proteins, is down-regulated coincident with termination of cell expansion (Arpat, 2004). Besides a large group of fiber genes that perform basic functions during elongation, the switch in developmental programs from PCW to SCW synthesis and the termination of cell expansion is accompanied by dynamic changes in gene expression. About 2,500 highly and moderately expressed fiber genes were down-regulated in terminal stages of fiber elongation which function selectively or preferentially during cell expansion (Arpat et al., 2004). An increase in cellulose and expression of genes encoding cellulose synthase marks the end of the rapid elongation stage of fiber development.

Primary and secondary cell wall synthesis

In addition to a large diverse group of constitutively expressed genes, expression profiling of the transcriptome revealed two developmentally regulated stage-specific expression patterns that follow rapid cell elongation: PCW and SCW biogenesis (Wilkins and Arpat, 2005). The PCW is laid down during the elongation phase, lasting up to 25 DPA. Synthesis of the secondary wall commences prior to the cessation of the elongation phase and continues to 40 DPA, forming a wall (5 - 10 µm) of almost pure cellulose (Wilkins and Arpat, 2005) and cell wall biosynthesis is a major synthetic activity in fiber cells. The cell wall components are synthesized and transported by a functionally integrated membrane system of endoplasmic reticulum, golgi complex and plasmalemma. Newly synthesized cell wall polypeptides are released into the endoplasmic reticulum lumen before transportation and incorporation into cell walls. Glycosylation of structural proteins, as well as polymerization of hemicelluloses and pectin, takes place in the golgi complex from which the products are released into the plasmalemma through the fusion of vesicles (Basra and Malik, 1984).

In addition to cellulose, cotton fiber also contains smaller amounts of pectins, hemicellulose, waxes, proteins and inorganic salts. The mechanism by which cellulose micro fibrils are produced and assembled along with the other components is not fully understood. Microtubules that are situated in the cytoplasm directly adjacent to the developing cell wall may participate in microfibril organization (Seagull, 1992). Some of the protein constituents of cotton fiber (enzymatic, structural or regulatory) are unique to fiber cells and are likely to influence the development and properties of cotton. Evaluation of proteins from fibers of different development mental stages and other cotton plant tissues by two-dimensional gel electrophoresis has revealed fiber-specific proteins that are developmentally regulated (Yao et al., 2006).

Although it is clear that no gross changes occur in the RNA population, the protein content may change during cell wall development. Earlier measurement of protein content in the cotton cell wall (weight percent) has shown to decrease from a high of 40% in 5 DPA fibers to < 2% in 18 DPA fibers (Mienert and Delmer, 1977). Measured as weight per unit length, the fiber wall protein content peaks in ~ 16 DPA fibers before rapidly declining (Mienert and Delmer, 1977). The level of steady-state E6 RNA was high in 20 DPA fibers and persisted beyond 24 DPA fibers. However, there was a sharp decline of E6 proteins in 15 DPA and older fibers (John and Crow, 1992). Thus, the decline of E6 protein in 15 DPA fibers and the near absence of E6 protein in 24 DPA fibers as shown by John and Crow, (1992) can be interpreted to be due to a protein degradation occurring in fiber cells during this time period. This degradation must be selective since a major biosynthetic activity, cellulose synthesis/deposition, occurs in fibers 24 DPA and older and protein components necessary for this event would therefore be expected to be preserved. As revealed by *in silico* expression analysis, cytoskeleton and cell wall-related genes are by far, among the most abundant gene transcripts during PCW synthesis in elongating fibers, while metabolism-related genes account for the vast majority of moderately expressed genes (Wilkins and Arpat, 2005).

PCW and SCW are markedly different in terms of structure and composition (Carpita and Gibeau 1993). In cotton fibers, a thin (0.2 - 0.4 mm) PCW deposited during fiber elongation contains < 30% cellulose, whereas the thick (8 - 10 mm) SCW is composed of > 94% cellulose, no lignin and few proteins (Meinert and Delmer, 1977). A gene encoding H^+-pyrophosphatase is constitutively expressed but enzymatic activity is temporally regulated; suggesting that this particular proton pump plays a functional role during the developmental switch from PCW to SCW synthesis (Smart et al., 1998). Temporal regulation of vacuolar H^+-ATPase gene expression, which closely parallels fiber growth rate, is accompanied by a corresponding change in protein abundance and enzymatic activity (Smart et al., 1998). In addition to that, many cell wall and carbohydrate metabolism related genes are temporally regulated during SCW synthesis (Arpat et al., 2004) in a manner typified by the developmental regulation of *GhCesA1* and *GhCesA2* genes encoding the catalytic subunit of cellulose synthase. Expression of *GhCesA* genes initially detected at low levels during the latter stages of fiber expansion, signals the onset of SCW synthesis (Wilkins and Arpat, 2005). Expression dramatically increases coincident with the rate of cellulose synthesis to reach peak levels at approximately 24 DPA (Meinert and Delmer, 1977, Pear et al., 1996).

CelA1 and *CelA2*, two homologs of the bacterial celA genes that encode the catalytic subunit of cellulose synthase, were cloned from cotton fiber, and they showed expression in developing cotton fibers at the onset of SCW cellulose synthesis (Pear et al., 1996). The polypeptide encoded by the *CelA1* DNA fragment with four highly conserved subdomains had the ability to bind the UDP-glucose (UDP-Glc) *in vitro* and UDP-Glc is the substrate to synthesize cellulose (1,4-β-D-glucan) (Pear et al., 1996). Similarly, screening of 20 DPA fiber subtracted library has shown that arabinogalactan protein and fiber glycosyl hydrolase family 19 protein, were found to be highly expressed in fibers with the maximal transcription level during the developmental switch from elongation to cellulose deposition (Liu et al., 2006).

Termination of elongation

Termination of fiber elongation is accompanied by a corresponding decrease in growth rate, transcriptional activity (Kosmidou-Dimitripoulou, 1986) and protein complexity (Graves and Stewart, 1988). At this stage, cross-linking of cellulose micro fibrils and non-cellulosic matrices presumably ‘‘fix’’ the structure of the PCW (Wilkins

and Jernstedt, 1999), resulting in the first significant increase in fiber strength (Hsieh 1999). Although it has been speculated that programmed cell death plays a role in fiber maturity, virtually nothing is known about the latter stages of fiber development as molecular studies are hindered by the inability to isolate RNA from fibers much past 25 DPA when cellulose is being deposited (Wilkins and Arpat, 2005).

Cotton fiber cells do not synthesize SCW cellulose until their elongation process stops at ~ 16 to 18 DPA (Basra and Malik, 1984). The undetectable level of expansin gene indicates that the initial highly extendable PCW of elongating fiber has become quite a rigid synthesis (Ruan et al., 2001). Consistent with this change is the suberization of fiber SCW (Ryser, 1999).

The deposition of hydrophobic suberin in the basal part of fibers excludes the apoplastic pathway for solute import (Ryser, 1999). This could be the structural basis for the reopening of fiber plasmodesmata (Ruan et al., 2001). The dramatically reduced expressions of sucrose and K^+ transporter genes have been documented at 16 DPA (Kühn et al., 1997).

This reduction in transporter expression is in agreement with the shift back to the symplastic pathway of solute import into fibers. Although symplastic sucrose import into fibers is sustained by the activity of sucrose synthase in the cytosol of fibers, the import of K^+ into fibers was greatly reduced after 15 DPA (Ruan et al., 1997). This may contribute to the decrease of osmotic and turgor potentials in fibers and slow down the elongation (Cassman et al., 1990). Furthermore, the reopening of fiber plasmodesmata at ~ 16 DPA would release higher turgor in fibers, if any, to a level similar to that present in the seed coat (Ruan et al., 2001). Together, these studies propose that fiber elongation could be terminated by the combination of increased cell wall rigidity and loss of higher turgor.

Until now, the total mechanism of fiber growth and development has been unclear, even though the expression patterns and putative function of array of genes have been analysed in details as shown above. Thus fiber cells, though devoid of any critical functions in the cotton plant except seed dispersal, contain large numbers of active genes common to other cell types along with fewer active genes unique to fiber.

A second point that became clear during transcriptome analysis was that no major changes occurred in the mRNA population during PCW and SCW synthesis stages and no subset of genes that are exclusively expressed during a given developmental stage was detected. Thus, it seems likely that most of the genes in the fiber are active throughout its development. Alternatively, gene transcription ceases early in fiber development, but differential mRNA utilization occurs during growth and thus the protein content may change during later fiber developmental stages. It is believed that advances in 'omics' studies may open up new avenues in this area in the near future.

Functional genomics of biotic and abiotic stress resistances

It is noteworthy that under field conditions, biotic and abiotic stresses are the major reasons for poor seed set which leads to both yield loss and lower fiber quality. Hence, it is important to understand the impact of these stresses on fiber development. Though there has been no such direct studies made in cotton, the molecular response of cotton with respect to external environmental stresses were documented. For example, during the process of *Verticillium dahliae* infection, the resistant and susceptible plant varieties respond differently. Molecular cloning of the transcripts related to the cotton *Verticillium* wilt response have shown that several defense genes such as chitinase, β-1,3-glucanase and PR-10, phenylalanine ammonialyase and those encoding several metabolic and signaling enzymes were activated and their expressions changed radically (Zhang and Klessing, 2001). In an another study, suppressive subtractive hybridization method was employed to identify differentially expressed ESTs in the *V. dahliae* infection process and many up-regulated and down-regulated novel ESTs, including those involved in synthesizing signal molecules, oxidative metabolism and those related to stress tolerance were isolated (Zuo et al., 2005). Cap-turing these transcripts and characterizing their roles helped in explaining the molecular resistance mechanism of cotton *Verticillium* wilt.

In a similar transcriptome study of cotton- *Xanthomonas campestris pv. malvacearum* interaction, clones from a cDNA library were used to identify host genes expressed in upland cotton leaves following inoculation (Patil et al., 2005). Microarray analysis indicated that 98% of the analysed genes (which were enriched) that are involved in response to inoculation were significantly up-regulated at one or more of the sampling times. Of these, 63% had sequence similarity to plant genes associated with defence responses, that is, to genes that function in disease resistance/defence, protein synthesis/turnover, secondary metabolism, signalling, stress induced/ programmed cell death or code for pathogenesis-related proteins or retrotransposon-like proteins. Some of the genes in this study (17%) exhibited a maximum in differential transcript abundance at an early time. Hence, Patil et al. (2005) proposed a working hypothesis that it is quickness of up-regulation after inoculation that matters rather than the final intensity of up-regulation. Therefore, identification of early up-regulated genes would have potential applications in developing host resistance to bacterial blight.

Global expression profiling during abiotic stress has been documented in rice, wheat, maize and other crops (Vij and Tyagi, 2007). However, very scarce expression profiles are available with cotton abiotic stress response. Such large scale EST projects are useful in cotton to aid in the identification of the genes involved in abiotic stress. Water stress has a very critical impact in flower develop-

ment, boll formation and fiber maturity. Identification of genes involved in drought tolerance at the flowering and boll formation phases would be useful to enhance fiber productivity and quality under water limited environments. This has been exemplified by Lan et al. (2005) who found that more than one-half of the pollination/fertilization-related genes in rice were regulated by dehydration stress, indicating that water stress may be a crucial factor during pollination and fertilization.

Domestication of cotton involved selection for fiber-related traits, typically under irrigated conditions. This selection pressure may have inadvertently narrowed the genetic base and diluted the alleles that once enabled cotton to survive arid conditions in the wild (Rosenow et al., 1983). Drought tolerance may be re-introduced into elite lines with the aid of DNA markers for QTL associated with the trait. Saranga et al. (2004) identified drought-related QTLs for several physiological and crop productivity parameters in cotton. The next step to characterize osmotic adjustment and other drought-related QTLs in cotton is to identify the genes that correspond to these QTLs. The candidate gene approach is one method to achieve this goal (Pflieger et al., 2001). Generally, the main limitation hindering utilization of this method is the availability of genes associated with a particular trait. Research utilizing microarray technology offers a solution in that a list of candidate genes related to the phenotype of interest can be produced. However, in cotton, only fiber-based microarray chips are currently available. Fiber development and abiotic stress resistance in cotton has several similarity in gene expression as discussed below and so the available chips can be used to characterize abiotic stress responsive genes in cotton.

Growth is possible if turgor pressure is greater than a minimum threshold. Turgor pressure, in turn, is related, to the osmotic potential and to the transport coefficient for water uptake (Cosgrove, 1997). From the previous sections it is clear that the cotton fibre is a cell that elongates rapidly within a period of approximately 21 DPA. The cell faces unique challenges in maintaining proper balance between turgor pressure and cell wall extensibility due to the dilution of cellular osmoticum that results from water influx during rapid polar expansion. To cope, cotton fibers up-regulate a suite of genes that aide in solute accumulation and cell wall loosening (Ruan et al., 2001). Root cells face similar turgor versus cell wall extensibility challenges during water deficit stress. For growth to occur in well-watered conditions, cell turgor pressure must again be higher than threshold turgor pressure but unlike fiber cells, the main force that root cells must counter to maintain turgor pressure is the soil water potential (Hendrix et al., 2004). In a drought situation, the water potential of the soil is further reduced and becomes a greater hindrance to growth. To counter this, plants increase the accumulation of cellular solutes and cell wall loosening enzymes (Cosgrove, 1997; Ruan et al., 2001). These actions serve to increase turgor pressure and extensibility and, thus enable continued cell growth despite the low water potential of the surrounding soils. Though the above situations are driven by different stimuli (that is, fiber cells: rapid expansion versus root cells: soil water potential), cellular physiology demands the accumulation of solute and loosening of cell walls. Accordingly, gene expression profiles may be similar.

Nevertheless, a proof of concept is needed to address the validity of utilizing fiber-based microarrays to study drought tolerance in other tissues. Both *in silico* and biological expression analyses were used to assess the genetic similarity of these two phenomena (Hendrix et al., 2004). The data from the *in silico* analysis were supplemented with direct comparisons of the root and fiber transcriptomes using fiber-based microarray chips. Of the 5,506 genes, 95.9% (5,282) were expressed in both tissues. However, 169 fiber and 55 root genes were identified as tissue specific genes. Thus, utilization of new fiber-based microarrays for gene discovery during water-deficit stress is a feasible approach. However, expanding the fiber-based chips to include genes from other tissues would further facilitate gene discovery related to water-deficit stress and may eventually lead to complete characterization of drought-related QTLs in cotton (Hendrix et al., 2004).

It must be kept in mind that the basic task of the identification of key gene(s), whose manipulation will ultimately affect crop performance in response to abiotic stress, is highly complex and difficult to decipher because of the polygenic nature of the stress response. In addition, the plant's response to each stress is unique, and thus the response to multiple stresses will also be different. For example, global expression profiling of a plant's response to abiotic stress conditions has shown that, although overlap may occur for different abiotic stresses such as cold, salt, dehydration, heat, high light and mechanical stress, a set of genes unique to each stress response is also seen (Vij and Tyagi, 2007). However, most of the studies carried out to investigate the performance of plants under abiotic stress conditions have not focused on this aspect, making it an important area of concern especially as it is known that plants are exposed to multiple environmental stresses in the field. Further, the response to abiotic stress is also developmentally regulated (Vinocur and Altman, 2005). For instance, in plant species such as rice, wheat, tomato, barley and corn, salt tolerance increases with an increase in plant age. Moreover, it has been shown that QTLs associated with salt tolerance in the germination stage in barley, tomato and *Arabidopsis* are different from QTLs associated with the early stage of growth (Vij and Tyagi, 2007). Furthermore, in transgenic studies on crop plants such as rice, the majority have not evaluated the effect of stress on grain yield. It is apparent that an understanding of the abiotic stress responsive network will require a considerable amount of time and resources, but a systematic and concerted effort will ensure that only the most suitable genes

are identified for crop improvement. The task can be shortened by integrating the information already available and by avoiding the repetition of effort or branching away from the main focus. The final list of candidate genes and their alleles identified through this approach must be subjected ultimately to field trials to determine their efficacy.

Databases

Comprehensive study of any genome depends on the availability of a saturated and fully integrated genetic and physical map of cotton. Hence, all the information should be collected and made publicly available. There are several genome databases that are exclusively developed to serve the cotton research community. They are: The International Cotton Genome Initiative (ICGI; http://icgi.tamu.edu/), The Cotton Genome Database (CottonDB; http://www.cottondb.org;), The Cotton Microsatellite Database (CMD; (http://www.cottonssr.org), Comparative Evolutionary Genomics of Cotton (http://cottonevolution.info/), TropGENE Database (http://tropgenedb.cirad.fr/en/cotton.html), Cotton Genome Centre (http://cottongenomecenter. ucdavis.edu), The Cotton Diversity Database (http://cotton.agtec.nga.edu), the Cotton Portal (http://gossypium.info), National Center for Biotechnology Information (http://ncbi.nlm.nih.gov) for EST resources and BACMan resources at Plant Genome Mapping Laboratory (www.plantgenome.uga.edu). CottonDB provides genomic, genetic and taxonomic information including germplasm, markers, genetic and physical maps, trait studies, sequences and bibliographic citations. The cotton portal offers a single port of entry to participating cotton web resources. One participating resource, the Cotton Diversity Database (Gingle et al., 2006), provides for an interface relating to performance trial, phylogenetic, genetic and comparative data and is closely integrated with comparative physical, EST and genomic (BAC) sequence data, expression profiling resources and the capacity for additional integrative queries. Cotton oligo-gene microarrays consisting of approximately 23,000 70-mer oligos designed from 250,000 ESTs can be found at the web site http://cottonevolution.info/microarray. There is a great need to expand bioinformatic infrastructure for managing, curating and annotating the cotton genomic sequences that will be generated in the near future. The cotton sequence database of the future should be able to host and manage cotton information resources in cotton using community accepted genome annotation, nomenclature and gene ontology. Some existing databases may be upgraded to effectively handle a large amount of data flow and community requests, but additional resources will be sought to support key bioinformatic needs.

Future perspectives

Advances in technologies for harvesting specific cell types and in amplifying mRNA pools for expression profiling have stimulated studies of the transcriptome at the cellular level in plants (Galbraith and Bimbaum, 2006). However, these experiments have a common obstacle of sample preparation and single cell-type isolation that could impact interpretations of gene expression. Measuring gene expression in a single, abundant cell type will not have much experimental induced error since the developmental stages are overlapping. The problem is more aggravated when different developmental stages of fiber tissues are studied. Thus, the sampling time and state of the material are crucial for the experiment's results. Methods have been developed to eliminate these obstacles at each developmental time-point of fiber tissues (Taliercio and Boykin, 2007). However, simple and easy protocols which can be routinely used in laboratories need to be developed.

Furthermore, progress in the systematic survey of the genes crucial for fiber development is hampered by several factors including non availability of complete cotton genome, common method to assess microarray data quality, control measures to avoid false positives and poor performance in evolving common strategy to analyze the high throughput gene expression profiles *etc*. Careful experimental design and exploitation of suitable experimental samples along with the following considerations may fuel future research in cotton fiber development at molecular level.

Most of the studies in the literature were conducted in controlled green house experiments (of course, in field conditions in some cases) and several experimental controls such as internal, positive and negative controls, calibration spike-in controls, transgene and vector controls, ratio spike in controls, blank and buffer controls interspersed among the cotton oligoNTs and biological and technical replications were included to avoid experimental error. Thus, care has been taken to ensure the quality of the research (or genes that are expressed preferentially in fiber tissues). However, the impact of biotic and abiotic factors in the developmental process of the fiber, which has a demonstrated influence on fiber quality, is not discussed in the available literature and remains to be tested.

Given that the total number of genes in the cotton genome may be approximately 40,000 (Chen et al., 2007a), the available cotton DNA chips represents ~ 40% of the cotton genome. Analysing all the cotton genes in a single experiment may give a different picture. Based on the estimate of ~ 14,000 genes in the fiber transcriptome of a cultivated progenitor species and evidence of homoeologous genes from the A_T and D_T genomes in allotetraploid species, the fiber transcriptome increases to an estimated 28,000+ genes, making redundancy an issue in gene discovery projects (Arpat et al., 2004). Until now, the absolute mechanism of fiber elongation and cellulose biosynthesis has been unclear, even though the expression patterns and putative function of the reported genes

have been analyzed in details. Although so many related genes were isolated from varied plants, the precise function of the above said genes is not clear. This is mainly because of the recalcitrant nature of matured fibers which do not lend themselves for mRNA isolation. Hence, a proper method has to be identified to grasp the novel genes involved in matured fiber development.

Though proteomics help to get a complete picture of proteins and their role in fiber development, it is hard to grasp fiber proteome due to several reasons such as: presence of interfering material such as cell walls, phenolic compounds and other secondary metabolites that will severely disturb protein separation and analysis (Yao et al., 2006). In addition to that, labile proteins can be lost during the preparation of experimental samples and therefore, must be extracted from tissues by non-destructive techniques such as vacuum infiltration or recovered from liquid culture media from cell suspension cultures or seedlings. As yet, there is no efficient procedure to release proteins that are strongly bound to the extracellular matrix (Jamet et al., 2006). Structural proteins, for instance, extensins or proline-rich proteins, can be cross-linked via di-isodityrosine bonds and until now, extensins have been eluted with salts before their insolubilization from cell suspension cultures. Another difficulty is the separation of polypeptides by classical two-dimensional gel electrophoresis. For instance, basic glycoproteins are poorly resolved by this technique (Jamet et al., 2006). Such kinds of methods require specific methods of isolation and separation since the results are highly dependant on reliability of the purify-cation methods. Yet another limitation is that biochemical function of only a small proportion of the identified proteins have been demonstrated and/or determined based on the assumptions that proteins sharing conserved domains have the same activity. Hence, the remaining proteins (domains of unknown function) remain as a challenge for elucidation of their biological function. In addition to that quantitative data on proteome is still in its infant stage and protein-protein interactions and protein with other metabolites remains to be revealed. All these data combined with genetics, biochemistry and molecular biology can lead to a better understanding of roles of genes/proteins in fiber development.

The fundamental challenge in transcriptomic and proteomic studies is that precise prediction of structure and function of genes. The databases used for annotation are not complete though they have robust data on metabolic pathway (Taliercio and Boykin, 2007). Pathways associated with these may have not been identified due to poor representation of the pathway on the microarray and lack of annotated genes associated with these pathways etc. Jamet, (2004) has provided some examples of misleading annotations that were owing to error in experimental design or in data interpret-tation. A careful and critical bioinformatic analysis of DNA and/or protein sequences therefore appears to be an absolute requirement before starting a transcriptome analysis or discussing the results from such analysis and the relevance of bioinformatic predictions to biological data should be checked whenever possible to prevent mistakes. Certainly, there is a long way to go in determining how these genes work during fiber development. Kim and Triplett, (2004) also concluded that sequence similarity is insufficient evidence to predict accurately how proteins work in plant cells and thus supporting the need for biochemical assay and other related studies.

In addition to cDNA and oligonucleotide microarrays, tiling-path arrays have been used to study gene expression in plants (Vij and Tyagi, 2007). The advantage of tiling-path arrays over conventional microarrays is that they are not stuck-up with the gene structure and hence provide unbiased and more accurate information about the transcriptome. In addition, they provide information about transcriptional control at the chromosomal level. The use of tiling-path arrays could help to provide novel information about fiber transcriptome at the genome-wide level.

The quickly expanding knowledge on gene function and the availability of whole genome sequences of model plants is expected to offer new perspectives to solve the problems encountered in genetically and physiologically complex traits in commercial crops - which is referred to as ‘‘plant translational genomics’’ (Salentijn et al., 2007). The most promising tool for quick implementation of this knowledge is the candidate gene approach. The candidate gene approach is based on the assumption that genes with a proven or predicted function in a ‘model’ species (functional candidate genes) or genes that are enriched from a particular tissue or develop-mental pathway or genes that are co-localized with a trait-locus (positional candidate genes) could control a similar function or trait in an arbitrary crop of interest / target crop (Salentijn et al., 2007). In this respect, the recent progress in high-throughput profiling of the proteomics and metabolomics in *Arabidopsis* trichome development may enable the investigation of the concerted expression of thousands of genes and their possible role in fiber development. The multidisciplinary approaches are expected to contribute novel information toward a more comprehensive understanding of regula-tion of fiber development.

Jansen and Nap (2001) proposed the merger of genomics and genetics into the novel concept of genetical genomics: the expression levels of genes or cluster of genes are analysed within a segregating population. In genetical genomics, gene expression profiles are quantitatively assessed within a segregating population, and expression quantitative trait loci (eQTL) can be mapped like classical QTLs (Lacape et al., 2007b). The approach provides a novel way of discove-ring, at a genetic level, regulators of gene expression acting either in cis or trans relative to the target gene. The eQTL position may coincide with the gene itself displaying cis regulation (Kirst et al., 2005) or be different, thus revealing transacting fac-

tors controlling expression. A common feature of eQTL studies is the detection of 'hotspots' of trans-acting eQTL (Keurenjes et al., 2007), interpreted as regions rich in regulatory genes that co-regulate many downstream targets. Population-and genome-wide expression analyses also provides novel opportunities for correlating expression data to phenotypic/functional consequences. However, it should be noted that QTL regions appear often quite complex and approximate and may contain hundreds of genes. Consequently, the actual involvement of the candidate gene in most cases remains to be confirmed by genetic and physical mapping, positional cloning, expression analysis or genetic transformation experiments (Salentijn et al., 2007). Cost-saving alternatives to large genome-wide and population-wide analyses with minimal loss of informativeness have been proposed: analysing pooled samples of phenotypically extreme members of the population (Borevitz et al., 2003), or concentrating on genotypically-selected individuals (Xu et al., 2005).

Decoding cotton genomes will be a foundation for improving understanding of the functional and agronomic significance of polyploidy and genome size variation within the *Gossypium* genus. Sequenced cotton genomes will ultimately stimulate fundamental research on genome evolution, polyploidization and associated diploidization, gene expression, cell differentiation and development, cellulose synthesis, cell growth, molecular determinants of cell wall biogenesis, and epigenomics which will be useful in the sustainable production of high-yielding and high-quality fiber, seed and biomass (Chen et al., 2007a). Future characterization and utilization of cotton genome sequence information should integrate functional and structural genomic resources at the molecular and *in silico* levels, sequence full-length cDNAs for genome annotation and expression assays, perform detailed annotation of the cotton genome sequence to support gene discovery and map-based cloning in this species, implement a large-scale platform for identifying DNA sequence diversity (single nucleotide polymorphisms and genome specific polymorphisms), facilitate high-resolution whole-genome association studies, develop genomic tiling arrays to support gene expression and epigenomic analysis of biological and agronomic traits and sequence and annotate small RNAs and microRNAs and identify their targets (Chen et al., 2007a).

It should also be noted that there is no single cotton cultivar that has been grown globally. Hence, findings made in one variety/accession should not be generalized for the rest of the varieties since each cultivar differ in their character, habit and performance etc. When there is no similarity among the transcriptome and proteome of cell types, then identical functional roles for the same gene or protein among the cultivars cannot be expected. Hence, before making any final conclusion a comprehensive comparison is required.

In brief, characterization of the fiber transcriptome using genomic approaches has provided a development-tal framework to design strategies for the genetic improvement of yield and fiber quality and therefore has immediate applications in agricultural biotechnology. The next major task at hand will be the functional analysis of unannotated genes using reverse genetic approaches, which is much promising in light of recent advances in cotton transformation and regeneration technology (Wilkins et al. 2000; Wilkins and Arpat, 2005). In addition to the potential for bioengineering fiber properties in the future, significant headway should be made to exploit the fiber's ESTs to genetically map the fiber transcriptome as a step toward marker-assisted selection by molecular breeding programs (Lacape et al., 2007b).

Conclusion

Global gene expression analysis will be an important tool for unravelling genetic architecture and the connections between genotypic and phenotypic variation, but the results of such studies require careful interpretation (Holland, 2007). Despite the wealth of efforts in genetic mapping, transcriptomics and DNA sequencing to decipher the molecular determinants for the quality of the cotton fiber is difficult to predict and in what aspects these studies will eventually impact breeding processes. To directly relate gene action to expression phenotypes will require genetic mapping approaches, such as eQTL mapping, although this will remain challenging because of the need to assay expression levels in large numbers of genotypic samples. Even so, eQTL mapping is rapidly gaining recognition as a valuable approach for closing the gap between (structural) genetics and (functional) genomics (Lacape et al., 2007b). The compilation through meta-analysis of fiber QTL data from various studies (since the majority of the markers used are cross referenced in other populations) and the integration of QTL data with expression data (eQTL) would help to identify chromosomal regions important for fiber quality as well as important candidate genes influencing fiber quality and ultimately facilitate the breeding of superior genotypes (Chen et al., 2007a). A good parallel approach may be to search for candidates in commercial cultivars that are having naturally superior fiber qualities. Several studies performed to compare the structural differences in the genomes have shown that the difference is in the expression pattern, rather than in the presence or absence of particular genes (Gingle et al., 2006). Hence, the comparison of gene expression profiling between contrasting genotypes with respect to fiber quality can be extended to transcription profiling at the QTL level, and the genes identified at such QTLs may potentially be better candidates for superior fiber quality. Research is in progress in a recently sanctioned project on genetic improvement of cotton through marker aided selection using a mapping population developed from commercial cotton cultivars adapted to target the environment. Harnessing the full potential of functional genomics

will require a multidisciplinary approach and integrated know-ledge of the molecular and other biological pro-cesses of fiber development. This information has immediate applications in breeding programs geared towards genetic improvement of cotton yield and fiber quality which is the ultimate aim and outreach of these efforts.

ACKNOWLEDGEMENT

This work is supported by the Department of Biotechnology, Ministry of Science and Technology, government of India under the Program Support for Research and Development in Agricultural Biotechnology at TNAU.

REFERENCES

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

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main DO, Wood T, Leslie A, Wing RA, Wilkins TA (2004). Functional genomics of cell elongation in developing cotton fibers. Plant Mol. Biol. 54: 911-929.

Barwale RB, Gadwal VR, Zehr U, Zehr B (2004). Prospects for Bt cotton technology in India. AgBioForum *7*(1&2): 23-26.

Basra AS, Malik CP (1984). Development of the cotton fiber. Int. Rev. Cyt. 89: 65-113.

Benedict JH, Altman DW (2001). Commercialization of transgenic cotton expressing insecticidal crystal protein. Science Publishers Inc, Enfield, NH.

Berlin JD (1986). The Outer Epidermis of the Cottonseed, The Cotton Foundation, Memphis.

Borevitz JO, Liang D, Plouffe D., Chang H, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res. 13: 513-523.

Carpita NC, Gibeau DM (1993). Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3: 1-30.

Cassman KG, Kerby TA, Roberts BA, Bryant DC, Higashi SL (1990). Potassium nutrition effects on lint yield and fibrequality of Acala cotton. Crop Sci. 30: 672-677.

Cedroni ML, Cronn RC, Adams KL, Wilkins TA, Wendel JF(2003). Evolution and expression of MYB genes in diploid and polyploid cotton. Plant Mol. Biol. 51: 313-325.

Chen ZJ, Lee, JJ, Woodward AW, Han Z, Ha M, Lackey E (2007a). Functional genomic analysis of early events in cotton fibredevelopment, World Cotton Research Conference 4, Lubbock, Texas, USA.

Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, Arioli T, Brubaker C, Cantrell R, Lacape JM, Ulloa M, Chee P, Gingle AR, Haigler CH, Percy R, Saha S, Wilkins T, Wright RJ, Deynze AV, Zhu Y, Yu S, Abdurakhmonov I, Katageri I, Kumar PA, Rahman M, Zafar Y, Yu J, Kohel RJ, Wendel JF, Paterson AH (2007b). Towards sequencing cotton (*Gossypium*) genome. Plant Physiol. 145: 1303-1310.

Clive J (2006). Global status of commercialized biotech/GM crops: 2006, ISAAA Brief Number, ISAAA, Ithaca, New York.

Cosgrove DJ (1997). Relaxation in a high-stress environment: The molecular bases of extensible cell walls and cell enlargement. Plant Cell 9: 1031-1041.

Cosgrove DJ (2001). Wall structure and wall loosening: a look backwards and forwards. Plant Physiol. 125: 131-134.

Day RC, Grossniklaus U, Macknight RC (2005). Be more specific! Laser-assisted microdissection of plant cells. Trends Plant Sci. 8: 397-406.

Endrizzi JE, Turcotte EL, Kohel RJ (1985). Genetics, cytology, and evolution of Gossypium. Adv. Genet. 23: 271-375.

Fryxell PA (1979). The Natural History of the Cotton Tribe, Texas A&M University Press, College Station, TX.

Galbraith DW, Birnbaum K (2006). Global studies of cell type-specific gene expression in plants. Ann. Rev. Plant Biol. 57: 451-475.

Gialvalis S, Seagull RW (2001). Plant hormones alter fibreinitiation in unfertilized, cultured ovules of Gossypium hirsutumI. J. Cotton Sci. 5: 252-258.

Gingle AR, Yang H, Chee PW, May OL, Rong J, Bowman DT, Lubbers EL, Day JL, Paterson AH (2006). An integrated web resource for cotton. Crop Sci. 46: 1998-2007.

Graves AD, Stewart MJ (1988). Chronology of the differentation of cotton (*Gossypium hirsutum* L.) fibrecells. Planta 175: 254-258.

Harmer SE, Orford SJ, Timmis JN (2002). Characterization of six α-expansin genes in Gossypium hirsutum (upland cotton). Mol. Genet. Genom. 268: 1-9.

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res. 16: 252-261.

Hendrix B, Stewart JMcD, Wilkins TA (2004). Gene expression in developing cotton fibers as a model of water deficit stress in cotton roots. AAES Research Series. 533: 134-140.

Holland JB (2007). Genetic architecture of complex traits in plants. Current Opinion Plant Biol. 2: 3-15.

Havov R, Udall JA, Hovav E, Rapp R, Flagel L, Wendel JF (2008). A majority of cotton genes are expressed in single-celled fiber. Planta 227: 319-329.

Hsieh YL (1999). Structural development of cotton fibers and linkages to fibrequality, Hawthorne Press Inc., NY.

Jamet E (2004). Bioinformatics as a critical prerequisite to transcriptome and proteome studies. J. Expt. Bot. 55: 1977-1979.

Jamet E, Canut H, Boudart G, Pont-Lezica F (2006). Cell wall proteins: a new insight through protemics. Trends Plant Sci. 11: 33-39.

Jansen RC, Nap JP (2001). Genetical genomics: the added value from segregation. Trends Genet. 7: 388-391.

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC, Zhu YX (2003). Isolation and analysis of genes preferentially expressed during early cotton fibredevelopment by subtractive PCR and cDNA array. Nucleic Acids Res. 31: 2534-2543.

John ME (1999). Genetic engineering strategies for cotton fibremodification. Food Products Press, New York.

John ME, Crow LJ (1992). Gene expression in cotton (*Gossypium hirsutum* L.) fiber: Cloning of the mRNAs. PNAS, (USA). 89: 5769–5773.

Kambhampati U, Morse S, Bennet R, Ismael Y (2005). Perceptions of the impacts of genetically modified cotton varieties: a case study of the cotton industry in Gujarat, India. AgBioForum 8(2&3): 161-171.

Keurenjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJM, Vreugdenhill D, Koorneef M, Jansen RC (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. PNAS (USA) 104: s1708-1713.

Kim HJ, Triplett BA (2001). Cotton fibregrowth in planta and *in vitro*. Models for plant cell elongation and cell wall biogenesis. Plant Physiol. 127: 1361-1366.

Kirst M, Basten CJ, Myburg AA, Zeng ZB, Sederoff RR (2005). Genetic architecture of transcript-level variation in differentiating xylem of a Eucalyptus hybrid. Genet. 169: 2295-2303.

Kosmidou-Dimitripoulou K (1986). Hormonal Influences on FibreDevelopment, Cotton Foundation, Memphis.

Kühn C, Franceschi V, Schulz A, Lemoine R, Frommer WB (1997). Macromolecular trafficking indicated by localization and turnover of sucrose transporters in enucleate sieve elements. Sci. 275: 1298-1300.

Lacape JM, Claverie M, Jacobs J, Llewellyn D, Arioli T, Derycker R, Chiron NF, Giband M, Jean J, Viot C (2007b). Reconciliation of genetic and genomic approaches to cotton fibrequality improvement, World Cotton Research Conference 4, Lubbock, Texas, USA.

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

Lan L, Li M, Lai Y, Xu W, Kong Z, Ying K, Han B, Xue Y (2005). Microarray analysis reveals similarities and variations in genetic programs controlling pollination/fertilization and stress responses in rice (*Oryza sativa* L.). Plant Mol. Biol. 59: 151-164.

Lee JJ, Hassan OSS, Gao W, Wei NE, Kohel RJ, Chen XY, Payton P, Sze SH, Stelly DM, Chen ZJ (2006). Developmental and gene expression analyses of a cotton naked seed mutant. Planta 223: 418-432.

Li CH, Zhu YQ, Meng YL, Wang JW, Xu KX, Zhang TZ, Chen XY (2002). Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR. Plant Sci. 163: 1113-1120.

Liu D, Zhang X, Tu L, Zhu L, Guo X (2006). Isolation by suppression-subtractive hybridization of genes preferentially expressed during early and late fibre development stages in cotton. Mol. Biol. 40: 741-749.

Mah N, Thelin A, Lu T, Nikolaus S, Kühbacher T, Gurbuz Y, Eickhoff H, Klöppel G, Lehrach H, Mellgard B, Costelio CM, Schreiber S (2004). A comparison of oligonucleotide and cDNA-based microarray systems. Physiol. Genomics. 16: 361-370.

Marx-Figini M (1966). Comparison of the biosynthesis of cellulose *in vitro* and *in vivo* in cotton bolls. Nature 210: 747-755.

Meinert MC, Delmer DP (1977). Changes in biochemical composition of cell wall of cotton fibreduring development. Plant Physiol. 59: 1088-1097.

Meredith JWR (2000). Continued progress for breeding for yield in the USA? in: U. Kechagia, (Ed.), Proceedings of the World Cotton Research Conference II, Athens, Greece. pp. 97-101.

Orford SJ, Timmis JN (1998). Specific expression of an expansin gene during elongation of cotton fibers. Biochem. Biophys. Acta 1398: 342-346.

Patil BA, Pierce ML, Phillips AL, Venters BJ, Essenberg M (2005). Identification of genes up-regulated in bacterial blight resistant upland cotton in response to inoculation with *Xanthomonas campestris pv. malvacearum*. Physiol. Mol. Plant Patho. 67: 319-335.

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996). Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. PNAS (USA) 93: 637-642.

Pflieger S, Lefebvre V, Causse M (2001). The candidate gene approach in plant genetics: A review. Mol. Breed. 7: 275-291.

Rahman H (2006). Number and weight of cotton lint fibres: variation due to high temperatures in the field. Australian J Agri. Res. 57: 583-590.

Ramsey JC, Berlin JD (1976). Ultra structure of early stages of cotton fibredifferentiation. Bot. Gaz. 137: 11-19.

Reijans M, Lascaris R, Groeneger AO, Wittenberg A, Wesselink E, van Oeveren J, de Wit E, Boorsma A, Voetdijk B, van der Spek H, Grivell LA, Simons G (2003). Quantitative comparison of cDNA-AFLP, microarrays, and GeneChip expression data in *Saccharomyces cerevisiae*. Genomics 6: 606-618.

Rensink WA, Buell CR (2005). Microarray expression profiling resources for plant genomics. Trends Plant Sci. 10: 603-609.

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park CH, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (2004). A 3347-locus genetic recombination map of sequence tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166: 389-417.

Rosenow DT, Quisenberry JE, Wendt CW, Clark LE (1983). Drought tolerant sorghum and cotton germplasm. Agric. Water Manage. 7: 207-222.

Ruan YL, Chourey PS, Delmer PD, Perez-Grau L (1997). The differential expression of sucrose synthase in relation to diverse patterns of carbon partitioning in developing cotton seed. Plant Physiol. 115: 375-385.

Ruan YL, Llewellyn DJ, Furbank RT (2001). The Control of Single-Celled Cotton Fibre Elongation by Developmentally Reversible Gating of Plasmodesmata and Coordinated Expression of Sucrose and K^+ Transporters and Expansin. Plant Cell, 13: 47-57.

Ryser U (1999). Cotton FibreInitiation and Histodifferentiation, Hawthorn Press.

Saha S, Jenkins JN, Wu J, McCarty JC, Gutierrez OA, Percy RG, Cantrell RG, Stelly DM (2006). Effects of chromosome-specific introgression in upland cotton on fibreand agronomic traits. Genetics 172: 1927-1938.

Salentijn EMJ, Pereira A, Angenent CG, van der Linden F, Krens MJM, Smulders M, Vosman B (2007). Plant translational genomics: from model species to crops. Mol. Breed. 20: 1-13.

Saranga Y, Jiang CX, Wright RJ, Yakir D, Paterson AH (2004). Genetic dissection of cotton physiological responses to arid conditions and their inter-relationships with productivity. Plant Cell Env. 27: 263-277.

Schellmann S, Schnittger A, Kirik V, Wada T, Okada K, Beermann A, Jürgens G, Hülskamp M (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. The EMBO J. 21: 5036-5046.

Seagull RW (1990). Tip growth and transition to secondary cell wall synthesis in developing cotton hairs, Academic Press, San Diego.

Seagull RW (1992). A quantitative electron microscopic study of changes in microtubule arrays and wall microfibril orientation during in vitro cotton fibre development. J. Cell Sci. 101: 561-577.

Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003). Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20: 633-643.

Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, Zhu YX (2006). Transcript profiling, molecular biological and physiological studies reveal a major role for ethylene in cotton fibrecell elongation. Plant Cell 18: 651-664.

Shimizu Y, Aotsuka S, Hasegawa O, Kawada T, Sakuno TF, Hayashi T (1997). Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fibrecells. Plant Cell Physiol. 38: 375-378.

Smart LB, Vojdani F, Maeshima M, Wilkins TA (1998). Genes involved in osmoregulation during turgor driven cell expansion of developing cotton fibers are differentially regulated. Plant Physiol. 116: 1539-1549.

Sun Y, Veerabomma S, bdel-Mageed HA, Fokar M, Asami T, Yoshida S, Allen RD (2005). Brassinosteroid Regulates FibreDevelopment on Cultured Cotton Ovules. Plant Cell Physiol. 46: 1384-1391.

Sundaram V, Basu AK, Krishna Iyer KR, Narayanan SS, Rajendran TP (1999). Handbook of cotton in India. Mumbai, India: Indian Society for Cotton Improvement.

Suo J, Liang X, Pu L, Zhang Y, Xue Y (2003). Identification of *GhMYB109* encoding a *R2R3 MYB* transcription factor that expressed specifically in fibreinitials and elongating fibers of cotton (*Gossypium hirsutum* L.). Biochimica et Biophysica acta, pp. 25-34.

Szymanski DB, Marks MD (1998). *GLABROUS1* Overexpression and *TRIPTYCHON* alter the cell cycle and trichome cell fate in *Arabidopsis*. Plant Cell, 10: 2047-2062.

Taliercio EW, Boykin D (2007). Analysis of gene expression in cotton fibreinitials. BMC Plant Biol. 7: 22.

Taliercio EW, Hendrix B, Stewart JM (2005). DNA content and expression of genes related to cell cycling in developing Gossypium hirsutum (Malvaceae) fibers. American J. Bot. 92: 1942-1947.

Tan PK, Downey TJ, Sptiznagel EL, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC (2003). Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. 31: 5676-5684.

Tiwari SC, Wilkins TA (1995). Cotton (*Gossypium hirsutum* L.) seed trichomes expand via diffuse growing mechanism. Canadian J. Bot. 73: 746-757.

Triplett BA (1998). Stage-Specific Inhibition of Cotton FibreDevelopment by Adding a-Amanitin to Ovule Cultures. *In vitro* Cell. Dev. Biol. 34: 27-33.

Turley RB, Vaughn KC (1999). The primary walls of cotton fibers contain an escheating pectin layer. Protoplasma, 209: 237.

Udall JA, Flagel LE, Cheung F, Woodward AW, Hovav R, Rapp RA, Swanson JM, Lee JJ, Gingle AR, Nettleton D, Town CD, Chen ZJ, Wendel JF (2007). Spotted cotton oligonucleotide microarrays for gene expression analysis. BMC Genomics. 8: 81.

Udall JA, Swanson JM, Haller K, Rapp RA, Sparks ME, Hatfield J, Yu Y, Wu Y, Dowd C, Arpat AB, Sickler BA, Wilkins TA, Guo JY, Chen XY,

Scheffler J, Taliercio E, Turley R, McFadden H, Payton P, Klueva N, Allen R, Zhang D, Haigler C, Wilkerson C, Suo J, Schulze SR, Pierce ML, Essenberg M, Kim H, Llewellyn DJ, Dennis ES, Kudrna D, Wing R, Paterson AH, Soderlund C, Wendel JF (2006). A global assembly of cotton ESTs. Genome Res.16: 441-450.

Van't Hof J (1998). Production of micronucleoli and onset of cotton fibregrowth. Planta, 205: 561-566.

Van't Hof J (1999). Increased nuclear DNA content in developing cotton fiber cells. American J. Bot. 86: 776-779.

Vij S, Tyagi AK (2007). Emerging trends in the functional genomics of the abiotic stress response in crop plants. Plant Biotech. J. 5: 361-380.

Vinocur B, Altman A (2005). Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Current Opinion Biotech. 16: 123-132.

Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ, Chen XY (2004). Control of Plant Trichome Development by a Cotton Fibre *MYB* Gene. Plant Cell 16: 2323-2334.

Wilkins T (2007). Cotton biotechnology: current status and future prospects. World Cotton Research Conference 4, held on September 10-14, 2007 (Lubbock, Texas, USA).

Wilkins TA, Arpat AB (2005). The cotton fibretranscriptome. Physiologia Planta. 124: 295-300.

Wilkins TA, Jernstedt JA (1999). Molecular genetics of developing cotton fibers., Hawthorne Press, New York.

Wilkins TA, Rajasekaran K, Anderson DM (2000). Cotton Biotechnology. Critical Rev. Plant Sci. 15: 511-550.

Wu Y, Llewellyn DJ, White R, Ruggiero K, Al-Ghazi Y, Dennis ES (2007). Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of the rapidly expanding fibreinitial cells on the surface of cotton ovules. Planta 226: 1475-1490.

Wu Y, Machado A, White RG, Llewellyn DJ, Dennis ES (2006). Identification of early genes expressed during cotton fiber initiation using cDNA microarrays. Plant Cell Physiol. 47: 107-127.

Wu Y, Rozenfeld S, Defferrard A, Ruggiero K, Udall JA, Kim HR, Llewellyn DJ, Dennis ES (2005) Cycloheximide treatment of cotton ovules alters the abundance of specific classes of mRNAs and generates novel ESTs for microarray expression profiling. Mol. Genet. Genomics 274: 477-493.

Xu WL, Wang XL, Wang H, Li XB (2007). Molecular characterization and expression analysis of nine cotton *GhEF1a* genes encoding translation elongation factor 1A. Gene. 389: 27-35.

Xu Z, Zou F, Vision T (2005). Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. Genetics. 170: 401-408.

Yang SS, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD, Chen ZJ (2006). Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fibrecell development in allotetraploid cotton. Plant J. 47: 761-775.

Yao Y, Yang YW, Liu JY (2006). An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. Electrophoresis. 27: 4559-4569.

Zhang S, Klessig DF (2001). MAPK cascades in plant defense signaling. Trends Plant Sci. 6: 520-527.

Zuo K, Wang J, Wu W, Chai Y, Sun X, Tang K (2005). Identification and characterization of differentially expressed ESTs of Gossypium barbadense infected by *Verticillium dahliae* with suppression subtractive hybridization. Mol. Biol. 39: 191-199.