

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

## **Recent Developments in Fiber Genomics of Tetraploid Cotton Species**

---

Mirzakamol S. Ayubov,  
Ibrokhim Y. Abdurakhmonov,  
Venkateswara R. Sripathi, Sukumar Saha,  
Tokhir M. Norov, Zabardast T. Buriev,  
Shukhrat E. Shermatov, Khurshida A. Ubaydullaeva,  
Jack C. McCarty, Dewayne D. Deng and  
Johnie N. Jenkins

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.72922>

---

### **Abstract**

Cotton (*Gossypium* spp.) produces naturally soft, single-celled trichomes as fiber on the seed coat supplying the main source of natural raw material for the textile industry. It is economically considered as one of the most leading cash crops in the world and evolutionarily very important as a model system for detailed scientific investigations. Cotton production is going through a big transition stage such as losing the market share in competition with the synthetic fibers, high popularity of *Bt* and herbicide resistance genes in cotton cultivars, and the recent shift of fiber demands to meet the standard fiber quality due to change of textile technologies to produce high superior quality of fibers in the global market. Recently, next-generation sequencing technologies through high-throughput sequencing at greatly reduced costs provided opportunities to sequence the diploid and tetraploid cotton genomes. With the availability of large volume of literatures on molecular mapping, new genomic resources, characterization of cotton genomes, discoveries of many novel genes, regulatory elements including small and microRNAs and new genetic tools such as gene silencing or gene editing technique for genome manipulation, this report attempted to provide the readers a comprehensive review on the recent advances of cotton fiber genomics research.

**Keywords:** fiber genomics, SSR markers, QTL analysis, genome-wide analysis

---

## 1. Introduction

Cotton (*Gossypium* spp.), a natural fiber source for the textile industry [1], has significant economic impact in about 80 countries including USA, China, India, Pakistan, Brazil, Egypt, and Uzbekistan [2]. Cotton seeds are a rich source of vegetable oils, medicinal compounds and byproducts. Cottonseed is also used in livestock feed [3]. The estimated world cotton area and production worldwide are 32–34 million hectares and ~26 million metric tons of fiber [4]. Biologically, cotton fibers are single-celled trichomes [5] that grow from the epidermal cell layer of the ovule in a boll. The allotetraploid cotton genomes contain two subgenomes,  $A_t$  and  $D_t$ , in its nucleus, which resulted from the ancestral allopolyploidization of progenitor A-genome and D-genome diploids. The presence of the  $A_t$  and  $D_t$  subgenomes further increases the complexity in understanding the evolution, function, and composition of the allotetraploid cotton. Most of the economically important traits in cotton are controlled by complex quantitative trait loci (QTL) composed of many genes that collectively express the phenotype. Not all fiber quality traits are positively related to lint yield, and many have a negative genetic correlation with important agronomic traits including lint yield [6]. The principle of cotton breeding is mainly to select desirable traits based on market demand, and breeders embrace this objective in their genetic improvement programs. Previously, cotton breeders' primarily emphasized yield and agronomic characteristics, but recent technological changes in the spinning industry necessitated genetic improvements in major cotton fiber quality traits such as length, strength, micronaire (fineness), uniformity, reflectance, elongation, short fiber content, etc. [7]. Accordingly, cotton producers demand superior lines that will produce not only high yield but also improved fiber quality.

Although significant progress has been achieved in developing genetically-engineered insect and herbicide resistant cultivars, which suggested usefulness and efficiency of an integration of molecular and conventional methods for the genetic improvement of cotton, there is currently a great challenge to efficiently and rapidly breed complex traits such as fiber quality. One of the great obstacles in this regard is the narrowness of genetic base of current cultivar germplasm, which is due to following reasons: (1) a "genetic bottleneck" occurring during domestication of cotton [8, 9]; and (2) repeatedly using the similar cotton genotypes for breeding programs as a major parents [10]. Many of these above prompted researchers to bring novel tools and change breeding approaches to have successful cotton improvement programs. With advances in modern genomic technologies, considerable progress has been made in utilizing innovative approaches to achieve progress in cotton genetic research and breeding program. This review emphasized the latest advances in molecular mapping, genome-wide analyses, genome sequence, characterization of fiber genes, and existing genomic approaches for the improvement of fiber traits in allotetraploid cotton species.

## 2. Molecular markers

DNA markers are "landmarks" in the genome that can be selected because of their proximity to a QTL of interest. The selection of DNA markers linked to a QTL increases the efficiency of breeding, and usually decreases costs and subjective phenotypic selection with minimal backcrossing. Molecular markers represent a site of detectable variation in the genomic DNA, and

they are broadly categorized as: (1) restriction enzyme-based DNA markers (such as RFLP and AFLP), (2) polymerase chain reaction (PCR)-based markers (mostly SSRs) and (3) single nucleotide polymorphism (SNP)-based markers.

The first RFLP map in an interspecific cotton population (*G. hirsutum* × *G. barbadense*) with 705 RFLP loci identified 41 linkage groups that covered 4675 cM locations mapped on 11 pairs of homoeologous chromosomes [11]. An RFLP map with 63 fiber QTLs linked to the A-subgenome (chromosome 3, 7, 9, 10, and 12) and 29 fiber QTLs associated with the D-subgenome (chromosome 14Lo, 20, and the long arm of chromosome 26) were reported in 2005 [12]. RFLP's were extensively used in identifying a wide-range of QTLs linked to fiber quality, length, strength, uniformity, wall thickness, micronaire, fineness, and maturity [13, 14]).

A second group of molecular markers, AFLPs, are powerful and efficient. They are being continuously used in cotton genomic investigations [15]. In the recent past, AFLP markers have been used in monitoring the differential expression of cotton fiber transcripts during elongation and secondary cell wall thickening in interspecific (*G. hirsutum* × *G. barbadense*) RI lines [16].

SSRs are the most informative, versatile and readily detectable DNA-based markers [17]. They have been used to determine agronomically and economically important genes, genetic linkage mapping, and linkage disequilibrium-based association studies in cotton [18, 19]. Although, traditional methods of microsatellite marker development are costly and time-consuming [20], large numbers of SSRs have been used to explore the molecular diversity, population structure and elite alleles of several Upland cotton cultivars [21–24]. As a result, many fiber quality traits associated marker loci and fiber development have been identified [21–23, 25].

### 3. Mapping for fiber-related traits in cotton

The genetic improvement in cotton with full utilization of conventional breeding has provided significant progress, but the complexity of the genome and limited understanding of economically important traits has deterred efficient breeding. Cotton molecular breeding has been a reliable tool for the characterization and mobilization of complex QTLs of interest. The advent of genomics-assisted breeding has become an effective method for selecting parents for agronomic, stress-responsive and fiber-related traits. In the past decade, focus of cotton genomic research has shifted from a few marker genotyping based QTL-mapping efforts to large-scale marker-based genome-wide association studies (GWAS), using high-throughput next-generation sequencing (NGS)-based genotyping methods.

An inter-specific high-density linkage map of allotetraploid cotton has been constructed using the F<sub>2</sub> population of *G. hirsutum* and *G. darwinii* with 2763 markers associated with 26 linkage groups covering 4176.7 cM genome and displaying a few differences between A<sub>t</sub> and D<sub>t</sub> subgenomes [26]. Among 601 distorted SSR loci reported [27], a lower number of segregation distortion loci were located in the A<sub>t</sub>-subgenome than the D<sub>t</sub>-subgenome. Recently, 185 cotton genotypes were evaluated for mapping of major fiber traits using 95 polymorphic SSRs [27]. This map also covered some other marker-trait associations, such as average boll weight, and gin turn out percentage together with fiber traits such as micronaire value, staple length, fiber bundle strength, and uniformity index. The results showed a clear association of MGHES-51,

MGHES-31 and MGHES-55 with all these traits, which is useful in future marker-assisted breeding and gene cloning studies in cotton [25, 27]. Similarly, MGHES-31 and MGHES-55 EST SSR markers were associated with lint percentage in a unique RIL populations derived from linted and lintless genotype crosses [21].

The SSRs have been used to examine molecular diversity, population structure, and to scan for polymorphisms. Genome-wide mapping in over 500 inbred cotton cultivars collected from China, United States, and the Soviet Union helped in identification of 494 fiber-quality-associated SSRs. Of the 216 markers related to fiber quality identified in this study [23], 13 were reported in other studies as fiber-trait-associated markers.

Gene-based markers were developed based on candidate genes and EST sequences to detect polymorphism in *Gossypium* species and for genetic and physical mapping. EST-derived microsatellites have been used to increase the number of microsatellites available for genetic map construction, and facilitated the use of functional genomics to elucidate fiber development process. EST-derived SSR-based high-density genetic maps for cotton fiber genes were reported in a number of studies [28–32]. The early EST-SSR studies in cotton [28, 29] focused on mapping loci in an interspecific (*G. hirsutum* × *G. barbadense*) RIL population while recent efforts aimed at mapping colored fiber loci (Lc 1 and Lc 2) [32].

With the advance on molecular marker development in cotton, three new marker types, indel (insertion-deletion), SNP, and retrotransposon-microsatellite amplified polymorphism (REMAP) were used to increase map density in allotetraploid cotton cultivars [33]. SNP markers in cotton can be used to associate genes with the respective fiber traits. Over one hundred genomic regions have been identified by tagging >2500 SSR and SNP markers using interspecific recombinant inbred lines (RILs) [34]. A large set of gene-associated SNPs were identified by comparative transcriptome profiling of four wild (*G. tomentosum*, *G. mustelinum*, *G. armourianum*, and *G. longicalyx*) and two cultivated cotton species (*G. barbadense* and *G. hirsutum*) [35]. By combining RNA-Seq and super bulked segregant analysis sequencing (sBSAseq) approaches, *Li<sub>2</sub>* mutant and its wild-type near-isogenic line (NIL) *G. hirsutum* cv. DP5690 were screened to identify the *Ligon-lintless 2* (*Li<sub>2</sub>*) gene sequence, and subgenome-specific SNPs were identified in the tetraploid cotton [36].

Expansin protein plays an important role in fiber length and quality. Using NGS approaches, SNPs linked to six expansin genes were identified [37]. An  $\alpha$ -type cyclin-dependent kinase (GhCDKA) protein has conserved cyclin-binding, ATP-binding, and catalytic domains. It plays a key role in fiber development. The CDKA gene expression was validated by northern blot and qRT-PCR analyses. Further, SNPs linked to CDKA gene locus was assigned to the chromosome 16 [38]. Using comparative and resequencing analyses, 24 million SNPs were identified between the A- and D-genomes in cotton. This analysis revealed that A-genome is relatively more variable (duplications and deletions) than the D-genome. In *G. hirsutum*, 1472 conversion events including 113 overlapping genetic events were identified between homologous chromosomes [39].

Two hundred twenty and 115 BAC contigs for two homoeologous chromosomes 12 and 26 of Upland cotton have been identified within 73.49 Mb and 34.23 Mb in physical length. Numerous fiber unigenes and non-unigene locations were found in both chromosomes [40]. New marker loci and linkage groups were applied in different cotton species using informative sequence-based markers and DNA sequence information. Some of the ESTs and BACs



were physically anchored and clustered into the high-density genetic map, and functionally annotated and classified into different *Gossypium* species [33].

A single dominant *Ligon lintless-2* ( $Li_2$ ) gene results in significantly shorter fibers than a wild type in cotton (*G. hirsutum*). Two near-isogenic lines of  $Li_2$  cotton (one mutant and one wild type) lines were painstakingly created by five backcrosses ( $BC_5$ ) generations. An additional cross was used to develop a linkage map of the  $Li_2$  locus located on chromosome 18. The SSR marker NAU3991 was successfully mapped and showed complete linkage with the  $Li_2$  locus [41]. Marker-assisted breeding and *in vitro* mutagenesis of cotton ovules can provide an additional insight into the regulatory aspects of the  $li_2$  mutation in cotton [42]. Similarly, using linkage mapping and analysis in *G. raimondii*, the  $Li_1$  gene on chromosome 22 was identified. Many genes were recognized to be altered in the  $Li_1$  mutant line for early termination of fiber elongation. Several additional studies have also identified factors that affect *Li*-associated genes at the downstream position [43].

In addition, a comprehensive review of molecular markers, marker-assisted selection (MAS) and marker-assisted backcross (MAB) breeding have been presented recently [44]. Cleaved amplified polymorphic sequences (CAPS) and derived-CAPS (dCAPS) markers obtained from the genes of interest are becoming increasingly valuable markers in molecular breeding of crops along with SNP markers. Phytochrome gene-based CAPS, transcription factor (*HY5*), and specific dCAPS in cotton were developed using comparative sequence analysis of the *PHYA1*, *PHYB*, and *HY5* genes that showed close association of these genes with fiber quality and early flowering traits in cotton [45, 46].

One significant QTL can control multiple fiber quality traits, such as fiber length, micronaire, strength and uniformity, and helps to elucidate the molecular basis of fiber quality. One such fiber QTL was found and mapped between HAU2119 and SWU2302 markers in a *G. hirsutum* RIL population. Three candidate genes have been identified within this QTL by RNA-Seq and RT-PCR analysis [47].

Recently, 352 wild and domesticated cotton accessions were screened, and 93 domestication sweeps have been assigned to 74 Mb and 104 Mb of the A- and D-subgenomes in allotetraploid cotton, respectively. Further, genome-wide association study (GWAS) has identified 19 candidate loci for fiber-quality-related traits. This study suggested possible asymmetric sub-genome domestication for directional selection of long fibers. The effects of domestication on cis-regulatory divergence were shown by genome-wide screening for DNase I-hypersensitive sites and linking the variation to gene function [48].

Numerous additional genetic and genomic resources in cotton are available for the cotton research community through specific databases. In this context, CottonDB, established in 1995, is one of the earliest plant genome databases developed [49]. The International Cotton Genome Initiative (ICGI) was formed to coordinate the development of cotton genomics science including the creation of an integrated and saturated genetic and physical map in cotton [50]. The Cotton Microsatellite Database (CMD), an invaluable resource for cotton microsatellites, was developed to meet the goals of ICGI with the support received from Cotton Incorporated [51]. Later, the more comprehensive and robust database covering the genomic, genetic and breeding information collected from cotton was formed as CottonGen [52] database,

#	Author names	Publication year	Mapping population	Trait types mapped	Published journal	Reference
1	Iqbal and Rahman	2017	Inbred cultivars and segregating biparental population	Lint	Frontiers in Plant Science	[24]
2	Adhikari et al.	2017	F2 and F2:3	Fiber quality	Euphytica	[114]
3	Wang et al.	2017	BC3F2, BC3F2:3 and BC3F2:4	Fiber length	Theoretical and Applied Genetics	[44]
4	Su et al.	2016	Segregating biparental population	Fiber quality	Scientific Reports	[115]
5	Wang et al.	2016	RILs	Yield and fiber	PLoS ONE	[116]
6	Nie et al.	2016	Inbred cultivars	Fiber quality	BMC Genomics	[21]
7	Shang et al.	2016	RILs and BC	Fiber quality	G3:Genes Genomes Genetics	[117]
8	Liu et al.	2016	F2 and RIL118	Fiber quality	BMC Genomics	[43]
9	Badigannavar and Myers	2015	F2	Fiber	Journal of Cotton Science	[118]
10	Wang et al.	2015	RILs	Yield and Fiber	PLoS ONE	[119]

**Table 1.** Recent studies in mapping of fiber-related traits in cotton.

which has enhanced features such as data visualization, mining, sharing and retrieval [34]. Currently, there are 103 maps available in CottonGen [52]. A few recent mapping studies for fiber-related traits in cotton are presented (**Table 1**).

#### 4. Genome-sequencing efforts in cotton

The genome-sequencing endeavor in cotton has significantly advanced our existing genomic knowledge for the past decade. Available full sequence of cotton genomes should provide a better understanding of fiber and fiber-related traits. The initial efforts in cotton genome sequencing were started with the closest progenitor diploid species, *G. raimondii* (D5) and *G. arboreum* (A<sub>2</sub>), and later extended to other tetraploid species, *G. hirsutum* (AD)<sub>1</sub> and *G. barbadense* (AD)<sub>2</sub>. Currently, whole genome-sequencing efforts of the diploid progenitor species, *G. herbaceum* (A<sub>1</sub>), are progressing under collaboration between Alabama A&M University and USDA ARS [53]. Genome-sequencing information from the diploid and tetraploid species may aid in developing genetically engineered cotton lines, with superior agronomic traits. Recent reports in whole genome analysis in cotton are summarized in a tabular form (**Table 2**).

Authors	Year of publication	Species	Sequencing platform	Assembler used	Assembled genome size	Predicted genes	Journal published	Reference
Sripathi et al.	Unpublished	<i>Gossypium herbaceum</i>	Roche 454 and Illumina Hiseq2000	ABYSS, SSPACE and SEALER	~1.46 Gb	41,387	Unpublished	[55]
Li et al.	2015	<i>Gossypium hirsutum</i>	Illumina HiSeq 2000 platform	SOAPdenovo, SSPAC, and GapCloser	~2.17 Gb	76,943	<i>Nature Biotechnology</i>	[120]
Liu et al.	2015	<i>Gossypium barbadense</i>	Roche 454, Illumina Hiseq2000, PacBio SMRT	Newbler v2.3.	~2.20 Gb	77,526	Scientific Reports (Published online)	[57]
Li et al.	2014	<i>Gossypium arboreum</i>	Illumina HiSeq 2000 platform	SOAPdenovo	~1.94 Gb	41,330	<i>Nature Genetics</i>	[121]
Wang et al.	2012	<i>Gossypium raimondii</i>	Illumina HiSeq 2000 platform	SOAPdenovo and SSPACE	~0.58 Gb	40,976	<i>Nature Genetics</i>	[122]

**Table 2.** Genome-sequencing efforts in cotton.

Briefly, several key facts have been gleaned from the tetraploid cotton, *G. hirsutum* and *G. barbadense*, genomes-sequencing efforts. Using whole-genome shotgun reads, BAC-end sequences, and genotype-by-sequencing (GBS) genetic maps, the allotetraploid *G. hirsutum* TM-1 genome was sequenced [54]. It was subsequently assembled into ~2.56 Gb genome with 32,032 and 34,402 genes from  $A_t$  and  $D_t$  subgenomes, respectively. Comparative subgenome analysis revealed higher percentages of gene loss, disrupted genes, structural rearrangements, and sequence divergence in the  $A_t$  subgenome when compared to the  $D_t$  subgenome. This can be attributed to the evolutionary irregularities. However, no genome-wide expression dominances were found between the two subgenomes. It is important and should be noted that the  $A_t$  subgenome, with its positively selected genes (PSGs) for fiber improvement, and stress tolerance in the  $D_t$  subgenome, are tied to genomic signatures of selection and domestication [54]. Between the two allopolyploid cottons, *G. barbadense* is the preferred species for producing superior, extra-long fibers. The WGS analysis of *G. barbadense* suggested that the uneven genome-wide duplication was 20 million years ago (MYA) and pseudogenization 11–20 MYA might be a probable cause of genomic divergence. Further, based on sequenced genomic information, the role of a fiber-specific cell elongation regulator, *PRE1* (with  $A_t$  subgenome origin), conditioning extra-long fibers was revealed [55].

Since the completion of whole genome sequencing (WGS) of the nuclear genome of cotton, the focus has been shifted to chloroplast and mitochondrial genomes [56]. The complete cotton chloroplast genome is 160,301 bp in length with a total of 131 genes, of which 112 are unique genes and 19 are duplicated genes. The cotton chloroplast genome lacks *rpl22* and *infA*, and contains a number of dispersed direct and inverted repeats. The phylogenetic analysis of cotton



and 25 other completely sequenced angiosperm chloroplast genomes revealed strong relationships among *Malvales*, *Brassicales* and *Myrtales* within the rosids clades [57].

The complete mitochondrial (mt) DNA sequence of *G. raimondii* was assembled into a circular genome of (676,078 bp), and then compared and analyzed with other plants. The analysis showed 39 protein-coding, 6 rRNA, and 25 tRNA genes. Interestingly, almost all of the *G. raimondii* mitochondrial (mt) genome has been transferred to Chr1 within the nucleus. The phylogenetic analysis with the other related mt genomes of rosids showed that *G. raimondii* is closely related to *G. barbadense*. Similar to the plastid genome analysis, the phylogenetic analysis of mt genomes revealed *Brassicales* were closely related to *Malvales* in the rosids clades [1]. Sequencing and characterization of both nuclear and cytoplasmic genomes of the *Gossypium* species will enrich the knowledge used to identify fiber-related genes for the improvement of cotton fiber quality trait using modern genetic engineering tools [58, 59].

The whole genome sequences of *G. hirsutum* and *G. barbadense* have revolutionized molecular biology investigations in cotton. Knowledge gained from decoding the cotton genome has helped to improve our understanding of gene function to ultimately benefit growers with improved yield and fiber quality. It has provided an unprecedented opportunity to bridge the gap between functional and structural genomic research by using the reference sequences of the tetraploid cotton genome. Scientists are using new advanced technologies “to mine” for useful genes and understand the molecular processes of fiber development for cotton germplasm enhancement. For example, Paterson [60] reported that among 48 genes for which expression is upregulated in domesticated *G. hirsutum* fibers at 10 dpa, 20 genes show 10-fold enrichment relative to random genes. They are within QTL hotspot affecting length, uniformity, and short fiber content. Thirteen genes show 15-fold enrichment and they are in the homologous hotspot affecting fiber elongation and fineness. The reference tetraploid genome sequence revealed that non-random patterns of diverse data sets that are concentrated in the specific small regions of the At and Dt genome [60]. Having such enriched genomic data in hand, scientists are much closer to identifying the causal gene(s). For example, expression patterns of genes in *G. hirsutum* wild type and its near isogenic fuzzless/lintless mutant at the stage of fiber initiation were analyzed using the RNA-Seq data [61]. Recently, Chen et al. [62] reported that differential gene regulation causes the difference in the quality of fiber between *G. barbadense* and *G. hirsutum* based on integrated genome-wide expression profiling markers [62].

## 5. Fiber-related transcriptome and gene expression studies in cotton

The functions of cotton transcriptome were studied using multiple techniques such as comparative genomics, BLAST, Gene Ontology (GO) analysis, and pathway enrichment by Kyoto Encyclopedia of Genes and Genomes (KEGG). Some prominent fiber-related findings were found. In this section, we briefly summarize some of key findings on this regard.

To study key fiber development genes, fuzzless/lintless (*fl*) cotton mutants were considered to elucidate molecular mechanisms relevant to fiber length development [63]. Furthermore, different fiber developmental stages have been studied to understand comprehensive mechanisms of overall fiber development. For example, *G. hirsutum* wild type (WT) and its near

isogenic fuzzless/lintless (*fl*) mutant were used in comparative transcriptome analysis and microarray studies of developing ovules [64, 65]. In *fl* mutants, at the fiber initiation stage of fiber development, calcium and phytohormone mediated signal transduction pathways, biosynthesis of auxin and ethylene, and stress-responsive transcription factors (TFs) were downregulated, whereas researchers observed a strong downregulation of genes related to carbohydrate and lipid metabolisms, mitochondrial electron transport system (mETS) and cell wall loosening and elongation at the fiber elongation stage of development. A number of genes including *CELLULOSE SYNTHASES* and *SUCROSE SYNTHASE C* were down regulated in *fl* mutants at fiber initiation and secondary cell wall biosynthesis stages compared to the WT [64]. Interestingly, it was reported that some of genes related to phytohormone signaling and stress response, upregulated in the WT genotypes in the early period of fiber initiation, started their expression in the later period of 15 day of post anthesis (dpa) in *fl* mutants [64].

Comparative transcriptome analysis showed that only a few genes were differentially expressed in zero dpa ovules, and three dpa fibers. The importance of auxin signaling and sugar signaling pathways in modulation of different fiber developmental stages was studied using pathway studio analysis [65].

Another transcriptome analysis of *G. hirsutum* WT and its mutant *fl* ovules in fiber initiation and elongation stages has been implemented using high-throughput tag-sequencing (Tag-seq). Differentially expressed gene (DEG) analyses results revealed substantial changes in gene type and abundance between the wild type and mutant libraries. Most of the DEG in WT1/M1 and WT2/M2 libraries developed for the study of the fiber cell development included cellulose synthase, phosphatase, and dehydrogenase genes [66].

In order to identify important genes of Ligon Lintless-1 (*Li-1*) mutants during the secondary cell wall synthesis stage, high-throughput microarray technology and real-time PCR were employed. There were at least 2-fold differences in at least 100 expressed transcripts found during secondary wall biogenesis using transcriptome analysis. Expansin, sucrose synthase, and tubulin expression gene families were identified in *li-1* mutant. This signifies *Li-1* gene activities during later fiber developmental stages [67].

A comparative microarray analysis was used to study fiber elongation in two short fiber mutants, and their near isogenic WT to identify key genes or metabolic pathways. Energy production, increasing mitochondrial electron transport activity, and response-to-reactive oxygen-related genes showed higher expression in short fiber mutants than in wild type. At least 88 fiber elongation genes were identified that were not affected by growth condition [63].

Improving the defects in the fiber secondary cell wall (SCW) resulted in non-fluffy fibers, low dry weight, and fineness fiber in the *immature* fiber mutant (*im* mutant) of *G. hirsutum*. Lower cellulose content and thinner cell walls were found in *im* mutant than its near-isogenic WT line (NIL) TM-1 at the same fiber developmental stage. Sucrose content, an important carbon source for cellulose synthesis, was also significantly lower in *im* mutant than TM-1 during 25~35 DPA. Comparative analysis of fiber transcriptional profiling indicated that SCW biosynthesis was 3 days later in *im* mutant than TM-1. Cellulose synthesis, secondary cell wall biogenesis, cell wall thickening, and sucrose metabolism were associated with genes significantly upregulated in TM-1. Quantitative reverse transcription PCR (qRT-PCR) analysis

validated that carbohydrate metabolism had 12 related genes that were differentially expressed at the earlier transition. qRT-PCR also showed differences in the SCW biosynthesis stages of fiber development between TM-1 and *im* mutant, and the importance of the *im* regulatory gene functions in fiber SCW biosynthesis [68].

Human selection altered the phenotypic evolution of fiber development over long periods of selective breeding. This has been shown by comparative transcriptome profiling of developing fiber using RNA-Seq analysis. Over 5000 differentially expressed genes were found with a regulatory or participatory role in primary and secondary cell wall synthesis between wild and domesticated cottons, which arose from artificial selection [69]. Some 210,965 unigenes of more than 100 bp were obtained from 47.2 million paired end reads of the anthers of TM-1 through transcriptome sequencing. Many enriched genes were found in the processes of transcription, translation, and posttranslation as well as hormone signal transduction. In addition, the participation of transcription factor families and cell wall-related genes active during cell expansion and carbohydrate metabolism were analyzed [70]. To determine fiber elongation and cell wall differentiation, combined transcriptome and metabolome analyses were studied in *G. barbadense* and *G. hirsutum* cultivars. Results showed that 10–28 dpa *G. barbadense* fibers showed primary cell wall synthesis to support elongation, transitional cell wall remodeling, and secondary wall cellulose synthesis for continued fiber elongation. Deep sequencing of transcriptomes and non-targeted metabolomes of single-celled cotton fiber showed that the discrete developmental stage of transitional cell wall remodeling occurs before secondary wall cellulose synthesis begins in both genotypes. Among all 40,000 transcripts, expressed in the fiber and all of the cell wall-related transcripts, expression was similar between genotypes. However, cellulose synthase gene expression patterns were more complex than expected. Oxidative stress in fiber tissues was lower in *G. barbadense* when compared to *G. hirsutum*. Using deep-sequencing transcriptomes and non-targeted metabolomes, a transcriptional repression of lignification during cell wall synthesis was

#	Authors	Year of publication	Methods	Tissue used	Journal published	Reference
1	MacMillan et al.	2017	RNA-Seq	Fiber quality	BMC Genomics	[123]
2	Miao et al.	2017	small RNA-Seq	Fiber quality	PLoS ONE	[124]
3	Li et al.	2017	RNA-Seq	Fiber development	BMC Genomics	[125]
4	Hu et al.	2017	small RNA-Seq	Fiber yield	Plant Biotechnology Journal	[126]
5	Thyssen et al.	2017	RNA-Seq	Lint	Plant Journal	[127]
6	Naoumkina et al.	2017	RNA-Seq	Lint	Genomics	[128]
7	Ma et al.	2016	RNA-Seq	Fiber development	PLoS ONE	[129]
8	Naoumkina et al.	2016	RNA-Seq	Fiber development	BMC Genomics	[89]
9	Hinchliffe et al.	2016	RNA-Seq	Fiber color	Journal of Experimental Botany	[130]
10	Zou et al.	2016	RNA-Seq	Fiber development	Science, China	[131]

**Table 3.** Recent transcriptome analyses studies in cotton.

identified. The results implicated a positive contribution of the ascorbate-glutathione cycle in improving fiber length by the enhanced capacity of reactive oxygen species (ROS) [71].

Crossed and backcrossed inbred lines of *G. hirsutum* and *G. barbadense* have been developed to investigate fiber yield per acre. Using these experimental populations, a number of yield and yield component QTL co-localized differentially expressed genes (DEGs) and DEG-based SSCP-SNP markers have been identified [72]. Numerous (1486) DEGs were found from (BIL) population using a microarray-based comparative transcriptome analysis in 10 dpa fibers. In 24 yield QTL regions and 11 yield QTL hotspots, the 212 DEGs were mapped. Within the 7 lint-yield QTL, identified in the BIL population, additional 81 DEGs were mapped [72].

In another study, using the cotton EST sequences, available at NCBI, 29,547 and 19,578 unigenes were assigned to the D<sub>1</sub> and A<sub>1</sub> subgenomes of tetraploid cotton, respectively. Among these, the majority of the abundantly expressed genes played intricate roles in cotton fiber development [73]. Selected publication for recent fiber development-related investigations are summarized in **Table 3**.

## 6. Characterization of specific genes for cotton fiber

A cascade of genes is expressed during the stages of fiber development. Recently, several novel genes were identified.  $\alpha$ -expansins are one such gene family, which are cell wall proteins that disrupt non-covalent bonds between wall components to facilitate cell wall extension. Six  $\alpha$ -expansin mRNA encoding genes were isolated using a genomic library screen and PCR-based strategies. Four genes in the expansin gene family (*GhExp3-GhExp6*) are expressed in multiple tissues of cotton, and only two genes (*GhExp1* and *GhExp2*) showed expression in developing cotton fibers. *GhExp1* transcripts are highly expressed, while *GhExp2* transcripts were detected at low levels in the fiber. Of these two, *GhExp1* is relatively of greater importance to cell wall extension during fiber development [74].

Actin-depolymerizing factor (ADF) is also one of the important genes that modulates the polymeration and depolymeration of the actin filaments, and is also important for fiber development. *GhADF2*, *GhADF3*, *GhADF4*, and *GhADF5* genes, encoding ADF proteins, have been isolated from cotton (*G. hirsutum*) cDNAs. Bioinformatic analyses have shown the molecular evolutionary relationships of these genes, including their highly conserved status. Interestingly, *GhADF2* was predominantly expressed in the fibers and not in other tissues [75]. Downregulation of *GhADF1* in the transgenic cotton plants showed increased fiber length and strength as compared to the wild-type fiber. Transgenic fibers also contained more abundant F-actin filaments in the cortical region of the cells than control fibers. In transgenic fibers, the secondary cell wall appeared thicker and the cellulose content was higher than that of the control. This showed the critical role of *GhADF1* in the processes of elongation and secondary cell wall formation during fiber development [76].

Plants have a signaling system that mediates responses to environmental stimuli. Co-expression and preferential interaction between two calcineurin B-like (CBL) proteins and CBL-interacting protein kinase (*CIPK*) genes in the elongating fiber cells of *G. hirsutum* were determined [77]. Very specific characterization of a receptor-like kinase gene (*GhRKL1*), especially during secondary cell wall synthesis in the cotton fiber cells, has been studied. In cotton,



the location of *GhRLK1* products is considered to be in the plasma membrane. Additionally, the *GhRLK1* gene's function is thought to be in the signal transduction pathway, i.e., the induction and maintenance of active secondary cell wall formation during fiber development [78].

The *DET3* gene, which encodes subunit C of the vacuolar ATPase (V-ATPase), participates in Brassinosteroid-induced cell elongation. Seven candidates' expressed sequence tags (ESTs) were screened to analyze the function of *GhDET3* on the elongation of cotton fibers, and yielded detail data about this gene. Results showed that the amino acid sequence of *GhDET3* had high homology with *DET3* from *Arabidopsis*, rice and maize. Ubiquitous expression of this gene in all the tissues and organs has been detected by qRT-PCR analysis. The highest accumulation of *GhDET3* mRNA peaked during the fiber elongation stage (12 dpa), compared with the lowest level at the fiber initiation stage (0 dpa ovules) underscoring the vital role *GhDET3* plays in cotton fiber elongation [79].

One of the transcription factor genes in the *MADS*-box family has wide-ranging roles in many diverse aspects of plant development. The gene transcripts have been detected in developing cotton fiber cells and in other plant tissues. Alternative splicing results showed that transcripts may have altered cellular roles. This was due to encoded proteins with altered K-domains and/or C-terminal regions, and their subsequently variant proteins [80].

Reactive oxygen species (ROS) plays a prominent role in signal transduction and cellular homeostasis, as well as in plant cell development [81, 82]. However, growing and maturing cells encounter oxidative stress when resource imbalances occur [81]. In order to study oxidative stress-related genes and their expression levels in a single plant cell, microarray analyses have been conducted [81, 82]. Antioxidant genes were substantially upregulated in domesticated diploid and polyploid cotton (*Gossypium*) in contrast to WT. In contrast, no significant influence was shown on regulation of ROS-related genes with genomic merger and ancient allopolyploid formation in three wild allopolyploid species. The ROS-related processes were regulated by different sets of antioxidant genes. [81]. Reduced expression of ROS gene has also been observed in *im* mature fiber mutants [83].

Ten cotton class III peroxidase coding (*GhPOX*) genes were isolated from *G. hirsutum*. Class III peroxidase, an heme-containing enzyme, is encoded by a large, multigene family. These genes participate in the release or utilization of ROS. Among them, *GhPOX1* showed the most predominant expression in fast-elongating cotton fiber cells, and transcription level of *GhPOX1* was over 400-fold higher in growing fiber cells than in ovules, flowers, roots, stems and leaves. Results suggested that *GhPOX1* plays an important role during fiber cell elongation, possibly by mediating production of ROS [82].

*GhPFN1* is a gene isoform that was found to be preferentially expressed in cotton fibers. *GhPFN1* was also found to be tightly associated with the fast elongation of cotton fibers. Overexpression and quantitative analyses also confirmed that *GhPFN1* may play a critical role in the rapid elongation of cotton fibers by promoting actin polymerization [84]. Brassinosteroids (BRs) promote fiber elongation. The *BIN2* gene is a member of the shaggy-like protein kinase family, and functions as a negative regulator of BR signaling in *Arabidopsis*. Cotton *BIN2* genes have been characterized for investigation of BR-mediated responses in the development of cotton fibers. To further elucidate their role, cotton *BIN2* gene was



transformed into the *Arabidopsis* genome. Resulting transgenic *BIN2*-mutant *Arabidopsis* plants exhibited reduced growth, confirming the role of *BIN2* genes. The *BIN2* gene thus encodes functional bin2 isoforms that inhibit growth by negative regulation of BR signaling [85].

One approach used to detect novel genes during fiber development is by identification of fiber-associated gene-rich islands on cotton chromosomes. For this purpose, 10 gene-rich islands have been found in different stages of *G. hirsutum* fiber development on chromosomes 5, 10, 14, and 15 [86]. Distributions of a large number of fiber genes across the  $A_t$  and  $D_t$  subgenomes of AD tetraploid cotton have been studied extensively. In an attempt to develop an integrated genetic and physical map of fiber development, 103 fiber transcription factors, 259 fiber development genes, and 173 expressed sequence tag-short sequence repeats (EST-SSRs) have been mapped [87]. According to this study, major transcription factor genes and more fiber QTLs were mapped to the  $D_t$  subgenome than the  $A_t$  subgenome, whereas more fiber development genes were mapped to the  $A_t$  subgenome than the  $D_t$  subgenome. The  $D_t$  subgenome may provide the transcription factor genes that potentially regulate the expression of the fiber genes in the  $A_t$  subgenome [87].

Differential gene expression of candidate genes between wild type and mutant during fiber elongation stage has been shown by RNA-sequencing and qPCR analysis [88]. Twelve candidate genes of the Ligon lintless-1 mutant (*li-1*) were found in  $F_2$  mapping populations derived from the cross of Li 1 and H7124 genotypes. In *li-1* mutant genotype genes encoding ribosomal proteins, actin protein, ATP synthase, and beta-tubulin 5 were found as a putative candidates impacting fiber development process [88].

Heat shock transcription factors (Hsfs) have an important role in both plant stress and development. Due to global warming, cotton is also experiencing increased exposure to elevated temperature. Consequently, the development of the yield and the quality of lint are affected. Forty *GhHsf* genes were selected by Wang et al. [84] that were characterized into three groups: A, B, and C. In cotton, these *GhHsfs* were observed in the majority of the plant tissues, particularly around the developing ovules. Exposure to high temperature in cotton plants showed that *GhHsf39* demonstrated the most rapid response to the heat shock. It has been suggested that a differential expression of *Hsfs* may thus play a role in the fiber development that requires further study [89].

UDP-glucuronosyltransferase gene is a cytosolic glycosyltransferase that catalyzes the transfer of the glucuronic acid component of UDP-glucuronic acid to a small hydrophobic molecule. UGT makes up one of the largest and most important multigene family in plants. The cotton UGT, *GhGlcAT1*, gene promoter contains specific transcription regulatory elements, and provides clues about the roles of *GhGlcAT1* in cotton fiber development, especially during fiber elongation [90]. A phylogenetic study of the UGT proteins of cotton was studied in selected cultivars and wild *Gossypium* species. The study identified, analyzed, and compared 142 UGTs in *G. raimondii*, 146 in *G. arboreum*, and 196 in *G. hirsutum*. The conserved consensus sequence had 44 amino acids. It additionally showed a possibility of regrouping the *GrUGTs* and *GaUGTs* into 16 phylogenetic groups (A-P) and *GhUGTs* into 15 groups. Additionally, RNA-Seq data was used to study the expression patterns of the UGT genes in *G. hirsutum* wild type and its isogenic fuzzless/lintless mutant during fiber initiation [61].

WRKY gene products aid in managing stress responses in multiple plant species but they have not been extensively studied at various stages of fiber development. Ding et al. [87]

identified the relation between *WRKY* transcriptome factors and fiber development of *G. raimondii* and *G. arboreum* by studying their genome and transcriptome of 112 *G. raimondii* and 109 *G. arboreum* *WRKY* genes. The transcriptome analysis identified several *WRKY* genes active during fiber initiation, elongation, and maturation with different expression patterns between species. The association of *WRKY* allelic gene expression ( $D_t$  and  $A_t$ ) in *G. hirsutum* and alternative splicing events were likewise seen in both diploid and tetraploid cotton during the developmental stage of the fiber. In summary, this study provided new results for the evolution and role of *WRKY* gene family in cotton species [91].

The role of the *MYB* family transcription factors was evaluated during the developmental stage of cotton fiber. Within 1986 *MYB* and *MYB*-related putative proteins, 524 non-redundant cotton *MYB* genes were identified and regrouped into four subgroups (1*R*-*MYB*, 2*R*-*MYB*, 3*R*-*MYB*, and 4*R*-*MYB*). In addition, *MYB* transcription factors were classified into 16 subgroups according to the phylogenetic tree analysis. After analysis, 69.1% of all *GhMYBs* genes were identified as 2*R*-*MYB* subfamily. Conclusively, this study highlights important aspects regarding the functions of *MYB* transcriptome factors in cotton fiber development. Furthermore, it contributes to the understanding of the regulatory network of *MYB* in affecting other functions of cotton fiber development [92].

*CrRLK1L*, one of the receptor-like kinase (RLK) gene family subgroups, has previously been demonstrated to be important in the development pattern and spatial regulation in cotton. *CrRLK1L* family is believed to act as sensors for the integrity of the cell wall and regulators of polar elongation. This study focuses on *CrRLK1L* in cotton fiber development. A total of 44 *CrRLK1L* genes were isolated from *G. raimondii*, 40 from *G. arboreum*, and 79 from *G. hirsutum*. Among these, six genes played an important role in fiber development [93].

To visualize PME expression levels, 80 PME genes (*GaPME01*-*GaPME80*) were isolated from *G. arboreum*, 78 (*GrPME01*-*GrPME78*) from *G. raimondii*, and 135 (*GhPME001*-*GhPME135*) from *G. hirsutum*. The differences in the PMEs expression levels at the developmental stage of fiber was observed using qRT-PCR. Predominant expression in fiber was during the secondary cell wall thickening stage suggesting tissue-specific expression patterns in cotton fiber [94].

*LPAAT* is an enzyme from the Kennedy pathway in higher plants encoded by a multigene family. Recently, the role of modified-*LPAAT* gene ( $A_t$ -*Gh13LPAAT5*) in increasing the cottonseed oil content and fiber quality has been proposed by combining the genome-wide and transcriptome analyses [95].

## 7. Small RNA-mediated gene regulation studies in cotton

Small RNAs (including microRNAs, tasiRNAs, and piRNAs) are mainly 17–24 nucleotide long sequences that are scattered across the plant genomes and play an important role in regulating target gene expression via posttranscriptional and translational repression at different stages of plant development. The discovery of novel miRNA genes will help in understanding the key mechanisms associated with miRNA genesis and regulation of fiber development in cotton. Although the regulatory mechanisms of microRNAs and small non-coding RNAs were determined in overall plant growth and development [96], their specific role in the regulation

of fiber cell elongation and developmental processes were more widely elucidated primarily after 2008. The roles of small interfering and microRNAs in the development of the cotton ovule and fiber elongation were annotated by Abdurakhmonov and his colleagues [97], as a first attempt in fiber genomics. They identified three plant microRNAs (miR172, miR390, and ath-miR853-like) and demonstrated dpa-specific small RNA expression profiles during ovule development. These results suggested the complex dpa-specific small RNA regulation in ovule development covering 0–10 dpa fiber development stages [97].

Multiple approaches have since been developed to identify the role of small RNAs in fiber initiation and elongation. For example, a deep-sequencing approach was used to investigate global expression and complexity of small RNAs in wild type and fuzzless/lintless cotton ovules. Over 20 conserved candidate miRNA families, including their 111 members, were identified during fiber initiation and elongation. More than 100 unique target genes were predicted for most of the conserved miRNAs; two cell-type-specific novel miRNA candidates were also determined in cotton ovules [98]. More than 4 million small RNA sequences have been analyzed from fiber and non-fiber cotton tissues. Thirty one miRNA families, including 27 conserved and 4 novel miRNAs, have been identified from these tissues. In addition, 19 unique miRNA families were also identified representing 32 miRNA precursors. Seven families had been previously reported, and 25 new miRNA precursors have also been found.

The enrichment of siRNAs in ovules and fibers in small RNA metabolism and chromatin modification becomes active during fiber development [99]. A recent study identified 46 novel and 96 known miRNAs in elongating cotton fibers. They also found 64 differentially expressed miRNAs, and of those, 16 were predicted as novel miRNAs [100]. Several novel fiber miRNAs have been identified using high-throughput sequencing technologies during the secondary cell wall thickening stage in cotton [101]. Small RNA libraries were constructed from developing fiber cells of the short fiber mutants *Li-1*, *Li-2*, and their near-isogenic wild-type lines. Among 24 conserved and 147 novel identified families, four miRNAs revealed significant negative correlations with fiber lengths [96].

Earlier miRNA-specific DNA markers were developed and mapped in cotton to study the genetic variation of miRNAs and their putative target genes [98]. A number of pre-miRNA and putative target gene primers have been examined and polymorphic loci were mapped on the total tetraploid cotton chromosomes. MiRNA-based sequence-related amplified polymorphism (SRAP) markers were used in order to map more miRNA loci. RT-PCR analysis revealed unique expression patterns across different fiber development stages between the parents in pre-miRNAs and putative target genes [102].

In *G. hirsutum*, ~300 miRNAs have been identified targeting over 3000 genes that possibly regulate stress responses, metabolism, hormone signal transduction and fiber development [103]. Among 79 and 46 miRNA families identified in *G. hirsutum* and *G. raimondii*, respectively, eight miRNAs were specifically related to fiber elongation and associated pathways such as calcium and auxin signal transduction, fatty acid metabolism and anthocyanin biosynthesis, and xylem tissue differentiation. In addition, one tasiRNA was identified and its target, *ARF4*, was experimentally validated *in vivo* [104].

Approximately, 10 million non-coding RNAs (ncRNAs) from fiber tissue of the allotetraploid cotton (*G. hirsutum*) were sequenced 7 days after flowering (DAF), to identify 24 nt ncRNA

as the dominant species, followed by 21 nt ncRNA, and 23 nt ncRNA. This study further screened ~560 miRNA gene loci and suggested the role of miRNAs in elongation and secondary cell wall synthesis stages of cotton fiber development [105].

## 8. Functional genomics and genome-editing technologies in regulation of fiber genes

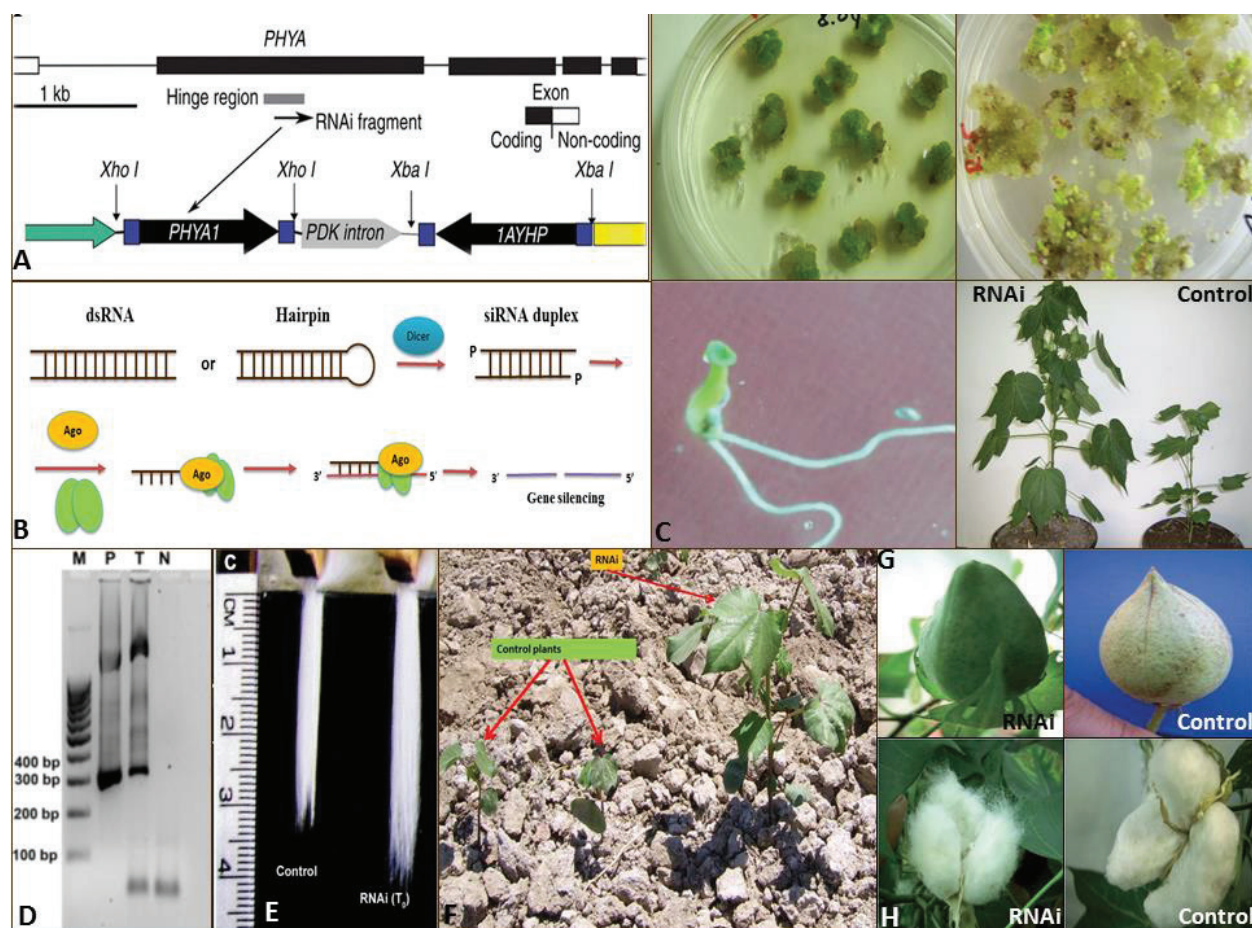
Genome modification (GM) and genome-editing technologies (GETs) are invaluable in the discovery of genes of interest, and support functional genomics of many organisms, including cotton. Various GM and GETs have been developed to investigate regulation mechanisms of genes. One of the most commonly used GM approaches is RNA interference (*RNAi*) that drew its historic support from antisense technology in the discovery of gene structures and the functions of organisms. *RNAi* is a new emerging technique based on homology-dependent post transcriptional gene silencing, induced by double-stranded RNA (dsRNA). Associations of many important genes with fiber development were detected using *RNAi* [3]. Recently, considerable work has been done highlighting the suitability of this method in cotton improvement [106–108]. For example, modifying or regulating flowering time is arduous in a plant improvement program, but it is sometimes a critical step to produce novel varieties with high yield that are better adapted to a specific environment.

In a collaborative project, scientists from Uzbekistan and USA developed cotton plants with early flowering, higher yield, and improved fiber qualities when *RNAi* technology was applied to regulate phytochrome gene [108] (**Figure 1**). Considering the potential of this research, scientists have patented this technology. A number of novel *RNAi* cultivars were successfully fielded trialed in over 60,000 hectares in Uzbekistan [3]. This is the first time a major crop is developed through the new *RNAi* technology and has been planted in such a large area.

A majority of *RNAi* published studies have focused on the functional aspects of cotton fiber-related genes [3, 109]. Later, Wang et al. [110] characterized dihydroflavanol 4-reductase (*DFR*) enzyme that mediates the biosynthesis of two polyphenols (anthocyanins and proanthocyanidins (PAs)) in Upland cotton. In order to silence *GhDFR1* in cotton, *DFR* gene was cloned from developing fibers, and used for virus-induced gene silencing. The results show a significant decrease in accumulation of anthocyanins and PAs when *GhDFR1* is silenced. In addition, a high decrease of two PA monomers, (-)-epicatchin and (-)-epigallocatechin, occurred in *GhDFR1*-silenced plant fibers while two new monomers, (-)-catchin and (-)-gallocatechin, were present compared to control plants infected with the empty vector. *GhDFR1* contribution in the biosynthesis of anthocyanins and PAs in cotton has thus been confirmed [110].

Overexpression of an important gene in an alien/host genome can be useful to detect gene functions and structure. The cellular functions of the class I of TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factor *GhTCP14*, from Upland, cotton were characterized by Wang et al. [111]. According to their work, the main expression of *GhTCP14* gene was detected in fiber cells at the initiation and elongation stages of development, and its expression increased in response to exogenous auxin. Overexpression of *GhTCP14* in *Arabidopsis thaliana* enhanced initiation and elongation of trichomes and root hairs. Moreover, it affected root





**Figure 1.** Different stages on the application of RNAi technology silencing *Phytochrome A* gene in *Gossypium hirsutum* grown at Uzbekistan (unpublished information collected from the reference 108). (A) Vector designed to introduce RNAi *PHYA1*; (B) RNAi mechanism in the cell; (C) *Agrobacterium tumefaciens* mediated cotton transformation with pHellsgate-8-*PHYA1* RNAi vector and plant regeneration using somatic embryogenesis through tissue culture system; (D) PCR-based marker confirmation of *PHYA1* RNAi regenerated cotton plant; (E) RNAiPhyA cotton line produces longer improved fiber compared to the control Cocker 312 plant; (F) RNAi plant showed rapid growth compared to the control at the young stage; (G) the cotton bolls showed different shape in control versus RNAiPhyA plant in the field; (H) matured bolls with fiber in RNAi Cocker 312 and control Cocker 312 cotton plants.

gravitropism like a mutant of the auxin efflux carrier PIN-FORMED2 (*PIN2*) gene. Expression of the auxin uptake carrier *AUXIN1* (*AUX1*) showed its upregulation, while *PIN2* was down-regulated in the *GhTCP14*-expressing plants. *GhTCP14* showed transcription activity by binding to the promoters of the *PIN2*, *IAA3*, and *AUX1* genes; these are auxin response genes that use electrophoretic mobility shift assays. All results demonstrated the potential regulation of *GhTCP14* gene in auxin-mediated differentiation and elongation of cotton fiber cells [111]. Overexpression of the actin-bundling protein *GhFIM2* was functionally characterized in cotton. The abundance of actin bundles is accompanied with accelerated fiber growth at the fast-elongating stage, and it increased when the *GhFIM2* gene was overexpressed. Secondary cell wall biogenesis also showed activation when the *GhFIM2* gene was overexpressed. These results indicated the importance of *GhFIM2* gene in the development of cotton fiber cells [112].

Recently, the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (*Cas*) 9 protein system have emerged as a simple and efficient tool for genome editing in eukaryotic cells. Most of the commercially grown cotton is tetraploid, and it is much more



difficult to target both sets of homologous alleles. In an initial effort to standardize CRISPR/Cas9 in the tetraploid cotton, a single copy gene, green fluorescent protein (GFP), has been utilized to determine the efficacy of the system in generating targeted mutations (indels) using three independent sgRNAs [113]. Literature analysis showed that application of novel generation GETs for cotton in general and fiber trait improvements in specific are in their very early stages and require more future attentions, coordinated efforts, and continuous investments.

## 9. Conclusion

For all crops—cotton in particular—where limited genetic diversity exists among the agriculturally elite types, genetic improvements will depend on innovative exploitation of genetic resources, and efficient strategies that effectively utilize both conventional and advanced molecular technologies. Comprehensive information is needed from independent and diverse research to understand molecular and genetic mechanisms associated with fiber development and additional agronomic traits. A cotton fiber provides a single cell crucible to understand the mechanisms of primary and secondary cell wall synthesis. It facilitates not only the study of cotton fibers, but also helps for further understanding of how all plant cell walls grow in relation to cell division and cell elongation. The elongating fibers also exemplify a scheme for production, and utilization of complex biochemical substances are first synthesized and then transported beyond the metabolic confines of cell membrane in the so-called “outer-space.” A cotton fiber cell can be studied in detail relatively away from the noise of metabolic activity within the “inner-space” where thousands (and soon to be tens of thousands) of functionally characterized genes participate in a spatial and temporal interplay of plant growth. These genes first promote the regulated cell growth in organ-specific manner, and later march the specialized fiber cells toward senescence and apoptosis. To study the complex but very interesting process of fiber development, the cotton research community has extensively applied the genetic tools from mapping to genome modification technologies through characterization of key genes and development of molecular markers. This enabled researchers to tag complex fiber QTLs, clone and characterize genes and breed novel superior fiber cultivars using MAS and GE technologies. However, functional genomics of fiber is still behind other crops in the utilization of new generation native GETs due to complexity and multi-allelic nature of fiber-related genes, which require well-planned cooperative research activities and larger investments.

## Acknowledgements

We thank Academy of Sciences of Uzbekistan, and Committee for Coordination Science and Technology Development of Uzbekistan for basic science (FA-F5-T030), and several applied (FA-A6-T081, FA-A6-T085 and YF5-FA-T-005) research grants. We greatly acknowledge the Office of International Research Programs (OIRP) of the United States Department of Agriculture (USDA) – Agricultural Research Service (ARS) and U.S. Civilian Research and Development Foundation (CRDF) for international cooperative grants P121, P121B, and UZB-TA-2992, which were devoted to cotton microRNA research. We also thank Dr. Sharma for his efforts to review this manuscript.

## Author details

Mirzakamol S. Ayubov<sup>1</sup>, Ibrokhim Y. Abdurakhmonov<sup>1</sup>, Venkateswara R. Sripathi<sup>2</sup>, Sukumar Saha<sup>3\*</sup>, Tokhir M. Norov<sup>1</sup>, Zabardast T. Buriev<sup>1</sup>, Shukhrat E. Shermatov<sup>1</sup>, Khurshida A. Ubaydullaeva<sup>1</sup>, Jack C. McCarty<sup>3</sup>, Dewayne D. Deng<sup>3</sup> and Johnie N. Jenkins<sup>3</sup>

\*Address all correspondence to: [sukumar.saha@ars.usda.gov](mailto:sukumar.saha@ars.usda.gov)

1 Center of Genomics and Bioinformatics, Academy Sciences of Uzbekistan, Tashkent, Uzbekistan

2 Center for Molecular Biology, Department of Biological and Environmental Sciences, Alabama A and M University, Alabama, USA

3 Genetics and Sustainable Agriculture Research, USDA-ARS, Mississippi, USA

## References

- [1] Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. *BioMed Research International*. 2016;**2016**:5040598
- [2] Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proceedings of the National Academy of Sciences USA*. 2006;**103**:18054-18059. DOI: 10.1073/pnas.0605389103
- [3] Abdurakhmonov IY, Ayubov MS, Ubaydullaeva KA, Buriev ZT, Shermatov SE, Ruziboev HS, Shapulatov UM, et al. RNA interference for functional genomics and improvement of cotton (*Gossypium* sp.). *Frontiers in Plant Science*. 2016;**7**:202. DOI: 10.3389/fpls.2016.00202
- [4] Stable Global Stocks in 2017/18 [Internet]. 2017. Available from: <https://www.icac.org/Press-Release/2017/PR-30-17-Stable-Global-Stocks-in-2017-18>
- [5] Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA. Functional genomics of cell elongation in developing cotton fibers. *Plant Molecular Biology*. 2004;**54**:911-929. DOI: 10.1007/s11103-004-0392-y
- [6] Desalegn ZN, Ratanadilok N, Kaveeta R. Correlation and heritability for yield and fiber quality parameters of Ethiopian cotton (*Gossypium hirsutum* L.) estimated from 15 (diallel) crosses. *Kasetsart Journal (Natural Science)*. 2009;**43**:1-11
- [7] Chee PW, Shen X, Lubbers EL, Paterson AH. Fine mapping for fiber length on chromosome 1 in cotton [Internet]. 2006. Available from: <http://www.ugacotton.com/vault/rer/2006/p73.pdf>
- [8] Wendel JF. New world cottons contain old world cytoplasm. *Proceedings of the National Academy of Sciences*. 1989;**86**:4132-4136. DOI: 10.1073/pnas.86.11.4132

- [9] Iqbal MJ, Reddy OU, El-Zik KM, Pepper AE. A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theoretical and Applied Genetics*. 2001;**103**:547-554
- [10] Van Esbroeck GA, Bowman DT, May OL, Calhoun DS. Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. *Crop Science*. 1999;**39**:323-328
- [11] Reinisch AJ, Dong JM, Brubaker CL, Stelly DM, Wendel JF, Paterson AHA. Detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. *Genetics*. 1994;**138**:829-847
- [12] Ulloa M, Saha S, Jenkins JN, Meredith WR Jr, McCarty JC Jr, Stelly DM. Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) Joinmap. *Journal of Heredity*. 2005;**96**:132-144. DOI: 10.1093/jhered/esi020
- [13] Lacape JM, Llewellyn D, Jacobs J, Arioli T, Becker D, Calhoun S, Viot C. Meta-analysis of cotton fiber quality QTLs across diverse environments in a *Gossypium hirsutum* × *G. barbadense* RIL population. *BMC Plant Biology*. 2010;**10**:132. DOI: 10.1186/1471-2229-10-132
- [14] Zhang S-W, Zhu X-F, Feng L-C, Gao X, Yang B, Zhang T-Z, Zhoua B-L. Mapping of fiber quality QTLs reveals useful variation and footprints of cotton domestication using introgression lines. *Scientific Reports*. 2016;**6**:31954. DOI: 10.1038/srep31954
- [15] Zhang JF, Yuan Y, Niu C, Hinchliffe DJ, Lu Y, Shuxun Y, Percy RG, Cantrell RG. AFLP-RGA markers in comparison with RGA and AFLP in cultivated tetraploid cotton. *Crop Science*. 2007;**47**:180-187. DOI: 10.2135/cropsci2006.04.0249
- [16] Claverie M, Souquet M, Jean J, Forestier-Chiron N, Lepitre V, Pré M, Jacobs J, Llewellyn D, Lacape JM. cDNA-AFLP-based genetical genomics in cotton fibers. *Theoretical and Applied Genetics*. 2012;**124**:665-683. DOI: 10.1007/s00122-011-1738-x
- [17] Saeed AF, Wang R, Wang S. Microsatellites in pursuit of microbial genome evolution. *Frontiers in Microbiology*. 2015;**6**:1462. DOI: 10.3389/fmicb.2015.01462
- [18] McCough SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biology*. 1997;**35**:89-99
- [19] Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*. 1996;**1**:215-222. DOI: 10.1016/S1360-1385(96)86898-0
- [20] Lopez L, Barreiro R, Fischer M, Koch MA. Mining microsatellite markers from public expressed sequence tags databases for the study of threatened plants. *BMC Genomics*. 2015;**16**:781. DOI: 10.1186/s12864-015-2031-1
- [21] Abdurakhmonov IY, Buriev ZT, Saha S, Pepper AE, Musaev JA, Almatov A, Shermatov SE, Kushanov FN, Mavlonov GT, Reddy UK, JZ Y, Jenkins JN, Kohel RJ, Abdukarimov A. Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*. *Euphytica*. 2007;**156**:141-156. DOI: 10.1007/s10681-007-9361-2

- [22] Abdurakhmonov IY, Kohel RJ, JZ Y, Pepper AE, Abdullaev AA, Kushanov FN, et al. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics*. 2008;**92**:478-487. DOI: 10.1016/j.ygeno.2008.07.013
- [23] Abdurakhmonov IY, Saha S, Jenkins JN, Buriev ZT, Shermatov SE, Scheffler BE, et al. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica*. 2009;**136**:401-417. DOI: 10.1007/s10709-008-9337-8
- [24] Nie X, Huang C, You C, Li W, Zhao W, Shen C, Zhang B, Wang H, Yan Z, Dai B, Wang M, Zhang X, Lin Z. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *BMC Genomics*. 2016;**17**:352. DOI: 10.1186/s12864-016-2662-x
- [25] Hinze LL, Gazave E, Gore MA, Fang DD, Scheffler BE, JZ Y, Jones DC, Frelichowski J, Percy RG. Genetic diversity of the two commercial tetraploid cotton species in the *Gossypium* diversity reference set. *Journal of Heredity*. 2016;**107**:274-286. DOI: 10.1093/jhered/esw004
- [26] Chen H, Khan MK, Zhou Z, Wang X, Cai X, Ilyas MK, Wang C, Wang Y, Li Y, Liu F, Wang K. A high-density SSR genetic map constructed from a F2 population of *Gossypium hirsutum* and *Gossypium darwinii*. *Genetics*. 2015;**574**:273-286. DOI: 10.1016/j.gene.2015.08.022
- [27] Iqbal MA, Rahman MU. Identification of marker-trait associations for lint traits in cotton. *Frontiers in Plant Science*. 2017;**8**:86. DOI: 10.3389/fpls.2017.00086
- [28] Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, el-Shihy OM, Cantrell RG. Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Molecular Genetics and Genomics*. 2005;**274**:428-441. DOI: 10.1007/s00438-005-0037-0
- [29] Han ZG, Guo WZ, Song XL, Zhang TZ. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. *Molecular Genetics and Genomics*. 2004;**272**:308-327. DOI: 10.1007/s00438-004-1059-8
- [30] Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T. A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics*. 2007;**176**:527-541. DOI: 10.1534/genetics.107.070375
- [31] Buyyarapu R, Kantety RV, JZ Y, Saha S, Sharma GC. Development of new candidate gene and EST-based molecular markers for *Gossypium* species. *International Journal of Plant Genomics*. 2011;**2011**:894598. DOI: 10.1155/2011/894598
- [32] Wang L, Liu H, Li X, Xiao X, Ai X, Luo C, Zhu L, Li X. Genetic mapping of fiber color genes on two brown cotton cultivars in Xinjiang. *Spring*. 2014;**3**:480. DOI: 10.1186/2193-1801-3-480
- [33] Zhao L, Yuanda L, Caiping C, Xiangchao T, Xiangdong C, Wei Z, Hao D, Xiuhua G, Wangzhen G. Toward allotetraploid cotton genome assembly: Integration of a high-density molecular genetic linkage map with DNA sequence information. *BMC Genomics*. 2012;**13**:539. DOI: 10.1186/1471-2164-13-539



- [34] JZ Y, Ulloa M, Hoffman SM, Kohel RJ, Pepper AE, Fang DD, Percy RG, Burke JJ. Mapping genomic loci for cotton plant architecture, yield components, and fiber properties in an interspecific (*G. hirsutum* L. × *G. barbadense* L.) RIL population. *Molecular Genetics and Genomics*. 2014;**289**:1347-1367. DOI: 10.1007/s00438-014-0930-5
- [35] Hulse-Kemp AM, Ashrafi H, Zheng X, Wang F, Hoegenauer KA, Maeda AB, Yang SS, Stoffel K, Matvienko M, Clemons K, Udall JA, Van Deynze A, Jones DC, Stelly DM. Development and bin mapping of gene-associated interspecific SNPs for cotton (*Gossypium hirsutum* L.) introgression breeding efforts. *BMC Genomics*. 2014;**15**:945. DOI: 10.1186/1471-2164-15-945
- [36] Thyssen GN, Fang DD, Turley RB, Florane C, Li P, Naoumkina M. Next generation genetic mapping of the ligo-lintless-2 (Li2) locus in upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*. 2014;**127**:2183-2192. DOI: 10.1007/s00122-014-2372-1
- [37] An C, Saha S, Jenkins JN, Scheffler BE, Wilkins TA, Stelly DM. Transcriptome profiling, sequence characterization, and SNP-based chromosomal assignment of the EXPANSIN genes in cotton. *Molecular Genetics and Genomics*. 2007;**278**:539-553. DOI: 10.1007/s00438-007-0270-9
- [38] Gao W, Saha S, Ma DP, Guo Y, Jenkins JN, Stelly DMA. Cotton-fiber-associated cyclin-dependent kinase a gene: Characterization and chromosomal location. *International Journal of Plant Genomics*. 2012;**2012**:613812. DOI: 10.1155/2012/613812
- [39] Page JT, Huynh MD, Liechty ZS, Grupp K, Stelly D, Hulse AM, Ashrafi H, Van Deynze A, Wendel JF, Udall JA. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3 (Bethesda)*. 2013;**3**:1809-1818. DOI: 10.1534/g3.113.007229
- [40] Xu Z, Kohel RJ, Song G, Cho J, Yu J, Yu S, Tomkins J, Yu JZ. An integrated genetic and physical map of homoeologous chromosomes 12 and 26 in upland cotton (*G. hirsutum* L.). *BMC Genomics*. 2008;**9**:108. DOI: 10.1186/1471-2164-9-108
- [41] Hinchliffe DJ, Turley RB, Naoumkina M, Kim HJ, Tang Y, Yeater KM, Li P, Fang DDA. Combined functional and structural genomics approach identified an EST-SSR marker with complete linkage to the Ligon lintless-2 genetic locus in cotton (*Gossypium hirsutum* L.). *BMC Genomics*. 2011;**12**:445. DOI: 10.1186/1471-2164-12-445
- [42] Boopathi NM, Thiyagu K, Urbi B, Santhoshkumar M, Gopikrishnan A, Aravind S, Swapnashri G, Ravikesavan R. Marker-assisted breeding as next-generation strategy for genetic improvement of productivity and quality: Can it be realized in cotton? *International Journal of Plant Genomics*. 2011;**2011**:670104. DOI: 10.1155/2011/670104
- [43] Gilbert MK, Turley RB, Kim HJ, Li P, Thyssen G, Tang Y, Delhom CD, Naoumkina M, Fang DD. Transcript profiling by microarray and marker analysis of the short cotton (*Gossypium hirsutum* L.) fiber mutant Ligon lintless-1 (Li1). *BMC Genomics*. 2013;**14**:403. DOI: 10.1186/1471-2164-14-403
- [44] Bolek Y, Hayat K, Bardak A, Azhar M. Molecular breeding of cotton. In: Abdurakhmonov IY, editor. *Cotton Research*. Croatia: InTech; 2016. DOI: 10.5772/64593



- [45] Kushanov FN, Pepper AE, JZ Y, Buriev ZT, Shermatov SE, Saha S, Ulloa M, Jenkins JN, Abdurakimov A, Abdurakhmonov IY. Development, genetic mapping and QTL association of cotton PHYA, PHYB, and HY5-specific CAPS and dCAPS markers. *BMC Genetics*. 2016;**17**:141. DOI: 10.1186/s12863-016-0448-4
- [46] Kushanov FN, Buriev ZT, Shermatov SE, Turaev OS, Norov TM, Pepper AE, Saha S, Ulloa M, Yu JZ, Jenkins JN, Abdurakimov A, Abdurakhmonov IY. QTL mapping for flowering-time and photoperiod insensitivity of cotton *Gossypium darwinii* Watt. *PLoS One*. 2017;**12**:e0186240. DOI: 10.1371/journal.pone.0186240
- [47] Liu D, Zhang J, Liu X, Wang W, Liu D, Teng Z, Fang X, Tan Z, Tang S, Yang J, Zhong J, Zhang Z. Fine mapping and RNA-Seq unravels candidate genes for a major QTL controlling multiple fiber quality traits at the T1 region in upland cotton. *BMC Genomics*. 2016;**17**:295. DOI: 10.1186/s12864-016-2605-6
- [48] Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, Zhou X, Nie X, Li Z, Guo K, Ma Y, Huang C, Jin S, Zhu L, Yang X, Min L, Yuan D, Zhang Q, Lindsey K, Zhang X. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics*. 2017;**49**:579-587. DOI: 10.1038/ng.3807
- [49] Yu J, Jung S, Cheng C-H, et al. CottonGen: A genomics, genetics and breeding database for cotton research. *Nucleic Acids Research*. 2014;**42**(Database issue):D1229-D1236. DOI: 10.1093/nar/gkt1064
- [50] Brubaker CL, Cantrell RG, Giband M, Lyon BR, Wilkins TA. Letter to Journal of Cotton Science Community: Formation of the International Cotton Genome Initiative, ICGI. *Journal of Cotton Science*. 2000;**4**:149
- [51] Blenda A, Yellambalase P, Palmer M, Cantrell R, Main D. Cotton Microsatellite Database for comparative characterization of SSRs in *Gossypium*. In: *Proceeding of International Cotton Genome Initiative (ICGI) Research Conference; 18-20 September 2006; Brasilia, Brazil*. pp. 25-26
- [52] CottonGen [Internet]. 2017. Available from: <http://www.cottongen.org>
- [53] Sripathi VR. Towards understanding the genome of diploid cotton, *Gossypium herbaceum* using deep sequencing. In: *Plant and Animal Genome XXIII Conference; 10-14 January 2015; San Diego, CA, USA*
- [54] Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*. 2015;**33**:531-537. DOI: 10.1038/nbt.3207
- [55] Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L, et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports*. 2015;**5**:14139. DOI: 10.1038/srep14139

- [56] Yan R, Liang C, Meng Z, Malik W, Zhu T, Zong X, Guo S, Zhang R. Progress in genome sequencing will accelerate molecular breeding in cotton (*Gossypium* spp.). 3. Biotech. 2016;**6**:217. DOI: 10.1007/s13205-016-0534-3
- [57] Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H. The complete chloroplast genome sequence of *Gossypium hirsutum*: Organization and phylogenetic relationships to other angiosperms. BMC Genomics. 2006;**7**:61
- [58] Abdurakhmonov IY. Genomics era for plants and crop species—Advances made and needed tasks ahead. In: Abdurakhmonov IY, editor. Plant Genomics. Croatia: InTech; 2016. DOI: 10.5772/62083
- [59] Abdurakhmonov IY. Cotton research highlights. In: Abdurakhmonov IY, editor. Cotton Research. Croatia: InTech; 2016. pp. 3-15. DOI: 10.5772/65456
- [60] Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature. 2012;**492**:423-427. DOI: 10.1038/nature11798
- [61] Huang J, Pang C, Fan S, Song M, Yu J, Wei H, Ma Q, Li L, Zhang C, Yu S. Genome-wide analysis of the family 1 glycosyltransferases in cotton. Molecular Genetics and Genomics. 2015;**290**:1805-1818. DOI: 10.1007/s00438-015-1040-8
- [62] Chen X, Guo W, Liu B, Zhang Y, Song X, Cheng Y, Zhang L, Zhang T. Molecular mechanisms of fiber differential development between *G. barbadense* and *G. hirsutum* revealed by genetical genomics. PLoS One. 2012;**7**:e30056. DOI: 10.1371/journal.pone.0030056
- [63] Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD. Comparative transcriptome analysis of short fiber mutants Ligon-lintless 1 and 2 reveals common mechanisms pertinent to fiber elongation in cotton (*Gossypium hirsutum* L.). PLoS One. 2014;**9**:e95554. DOI: 10.1371/journal.pone.0095554
- [64] Padmalatha KV, Patil DP, Kumar K, Dhandapani G, Kanakachari M, et al. Functional genomics of fuzzless-lintless mutant of *Gossypium hirsutum* L. cv. MCU5 reveal key genes and pathways involved in cotton fibre initiation and elongation. BMC Genomics. 2012;**13**:624. DOI: 10.1186/1471-2164-13-624
- [65] Liu K, Sun J, Yao L, Yuan Y. Transcriptome analysis reveals critical genes and key pathways for early cotton fiber elongation in Ligon lintless-1 mutant. Genomics. 2012;**100**:42-50. DOI: 10.1016/j.ygeno.2012.04.007
- [66] Wang QQ, Liu F, Chen XS, Ma XJ, Zeng HQ, Yang ZM. Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. Genomics. 2010;**96**:369-376. DOI: 10.1016/j.ygeno.2010.08.009
- [67] Bolton JJ, Soliman KM, Wilkins TA, Jenkins JN. Aberrant expression of critical genes during secondary cell wall biogenesis in a cotton mutant, Ligon lintless-1 (Li-1). Comparative and Functional Genomics. 2009;**2009**:659301. DOI: 10.1155/2009/659301

- [68] Wang C, Lv Y, Xu W, Zhang T, Guo W. Aberrant phenotype and transcriptome expression during fiber cell wall thickening caused by the mutation of the Im gene in immature fiber (im) mutant in *Gossypium hirsutum* L. BMC Genomics. 2014;**15**:94. DOI: 10.1186/1471-2164-15-94
- [69] Yoo MJ, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. PLOS Genetics. 2014;**10**:e1004073. DOI: 10.1371/journal.pgen.1004073
- [70] Zhang X, Ye Z, Wang T, Xiong H, Yuan X, Zhang Z, Yuan Y, Liu Z. Characterization of the global transcriptome for cotton (*Gossypium hirsutum* L.) anther and development of SSR marker. Gene. 2014;**551**:206-213. DOI: 10.1016/j.gene.2014.08.058
- [71] Tuttle JR, Nah G, Duke MV, Alexander DC, Guan X, Song Q, Chen ZJ, Scheffler BE, Haigler CH. Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. BMC Genomics. 2015;**16**:477. DOI: 10.1186/s12864-015-1708-9
- [72] Man W, Zhang L, Li X, Xie X, Pei W, Yu J, Yu S, Zhang J. A comparative transcriptome analysis of two sets of backcross inbred lines differing in lint-yield derived from a *Gossypium hirsutum* × *Gossypium barbadense* population. Molecular Genetics and Genomics. 2016;**291**:1749-1767. DOI: 10.1007/s00438-016-1216-x
- [73] Jin X, Li Q, Xiao G, Zhu YX. Using genome-referenced expressed sequence tag assembly to analyze the origin and expression patterns of *Gossypium hirsutum* transcripts. Journal of Integrative Plant Biology. 2013;**55**:576-585. DOI: 10.1111/jipb.12066
- [74] Harmer SE, Orford SJ, Timmis JN. Characterisation of six alpha-expansin genes in *Gossypium hirsutum* (upland cotton). Molecular Genetics and Genomics. 2002;**268**:1-9
- [75] Zhang C, Guo L, Wang X, Zhang H, Shi H, Xu W, Li X. Molecular characterization of four ADF genes differentially expressed in cotton. Journal of Genetics and Genomics. 2007;**34**:347-354. DOI: 10.1016/S1673-8527(07)60037-X
- [76] Wang HY, Wang J, Gao P, Jiao GL, Zhao PM, Li Y, Wang GL, Xia GX. Down-regulation of GhADF1 gene expression affects cotton fibre properties. Plant Biotechnology Journal. 2009;**7**:13-23. DOI: 10.1111/j.1467-7652.2008.00367.x
- [77] Gao P, Zhao PM, Wang J, Wang HY, XM D, Wang GL, Xia GX. Co-expression and preferential interaction between two calcineurin B-like proteins and a CBL-interacting protein kinase from cotton. Plant Physiology and Biochemistry. 2008;**46**:935-940. DOI: 10.1016/j.plaphy.2008.05.001
- [78] Li YL, Sun J, Xia GX. Cloning and characterization of a gene for an LRR receptor-like protein kinase associated with cotton fiber development. Molecular Genetics and Genomics. 2005;**273**:217-224. DOI: 10.1007/s00438-005-1115-z
- [79] Xiao Z, Tan K, Hu M, Liao P, Chen K, Luo M. Cloning and expression analysis of GhDET3, a vacuolar H<sup>+</sup>-ATPase subunit C gene, from cotton. Journal of Genetics and Genomics. 2008;**35**:307-312. DOI: 10.1016/S1673-8527(08)60044-2

- [80] Lightfoot DJ, Malone KM, Timmis JN, Orford SJ. Evidence for alternative splicing of MADS-box transcripts in developing cotton fibre cells. *Molecular Genetics and Genomics*. 2008;**279**:75-85. DOI: 10.1007/s00438-007-0297-y
- [81] Chaudhary B, Hovav R, Flagel L, Mittler R, Wendel JF. Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*). *BMC Genomics*. 2009;**10**:378. DOI: 10.1186/1471-2164-10-378
- [82] Mei W, Qin Y, Song W, Li J, Zhu Y. Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. *Journal of Genetics and Genomics*. 2009;**36**:141-150. DOI: 10.1016/S1673-8527(08)60101-0
- [83] Kim HJ, Tang Y, Moon HS, Delhom CD, Fang DD. Functional analyses of cotton (*Gossypium hirsutum* L.) immature fiber (im) mutant infer that fiber cell wall development is associated with stress responses. *BMC Genomics*. 2013;**14**:889. DOI: 10.1186/1471-2164-14-889
- [84] Wang HY, Yu Y, Chen ZL, Xia GX. Functional characterization of *Gossypium hirsutum* profiling 1 gene (GhPFN1) in tobacco suspension cells. Characterization of in vivo functions of a cotton profilin gene. *Planta*. 2005;**222**:594-603. DOI: 10.1007/s00425-005-0005-2
- [85] Sun Y, Allen RD. Functional analysis of the bin 2 genes of cotton. *Molecular Genetics and Genomics*. 2005;**274**:51-59
- [86] Xu Z, Kohel RJ, Song G, Cho J, Alabady M, Yu J, Koo P, Chu J, Yu S, Wilkins TA, Zhu Y, Yu JZ. Gene-rich islands for fiber development in the cotton genome. *Genomics*. 2008;**92**:173-183. DOI: 10.1016/j.ygeno.2008.05.010
- [87] Xu Z, JZ Y, Cho J, Yu J, Kohel RJ, Percy RG. Polyploidization altered gene functions in cotton (*Gossypium* spp.). *PLoS One*. 2010;**5**:e14351. DOI: 10.1371/journal.pone.0014351
- [88] Jiang Y, Ding M, Cao Y, Yang F, Zhang H, He S, Dai H, Hao H, Rong J. Genetic fine mapping and candidate gene analysis of the *Gossypium hirsutum* Ligon lintless-1 (li-1) mutant on chromosome 22(D). *Molecular Genetics and Genomics*. 2015;**290**:2199-2211. DOI: 10.1007/s00438-015-1070-2
- [89] Wang J, Sun N, Deng T, Zhang L, Zuo K. Genome-wide cloning, identification, classification and functional analysis of cotton heat shock transcription factors in cotton (*Gossypium hirsutum*). *BMC Genomics*. 2014;**15**:961. DOI: 10.1186/1471-2164-15-961
- [90] AM W, Lv SY, Liu JY. Functional analysis of a cotton glucuronosyltransferase promoter in transgenic tobaccos. *Cell Research*. 2007;**17**:174-183. DOI: 10.1038/sj.cr.7310119
- [91] Ding M, Chen J, Jiang Y, Lin L, Cao Y, Wang M, Zhang Y, Rong J, Ye W. Genome-wide investigation and transcriptome analysis of the WRKY gene family in *Gossypium*. *Molecular Genetics and Genomics*. 2015;**290**:151-171. DOI: 10.1007/s00438-014-0904-7
- [92] Salih H, Gong W, He S, Sun G, Sun J, Genome-wide DX. Characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. *BMC Genetics*. 2016;**17**:129. DOI: 10.1186/s12863-016-0436-8



- [93] Niu E, Cai C, Zheng Y, Shang X, Fang L, Guo W. Genome-wide analysis of CrRLK1L gene family in *Gossypium* and identification of candidate CrRLK1L genes related to fiber development. *Molecular Genetics and Genomics*. 2016;**291**:1137-1154. DOI: 10.1007/s00438-016-1169-0
- [94] Li W, Shang H, Ge Q, Zou C, Cai J, Wang D, Fan S, Zhang Z, Deng X, Tan Y, Song W, Li P, Koffi PK, Jamshed M, Lu Q, Gong W, Li J, Shi Y, Chen T, Gong J, Liu A, Yuan Y. Genome-wide identification, phylogeny, and expression analysis of pectin methylesterases reveal their major role in cotton fiber development. *BMC Genomics*. 2016;**17**:1000. DOI: 10.1186/s12864-016-3365-z
- [95] Wang N, Ma J, Pei W, Wu M, Li H, Li X, Yu S, Zhang J, Yu JA. Genome-wide analysis of the lysophosphatidate acyltransferase (LPAAT) gene family in cotton: Organization, expression, sequence variation, and association with seed oil content and fiber quality. *BMC Genomics*. 2017;**18**:218. DOI: 10.1186/s12864-017-3594-9
- [96] Naoumkina M, Thyssen GN, Fang DD, Hinchliffe DJ, Florane CB, Jenkins JN, Small RNA. Sequencing and degradome analysis of developing fibers of short fiber mutants Ligon-lintles-1 (li 1) and -2 (li 2) revealed a role for miRNAs and their targets in cotton fiber elongation. *BMC Genomics*. 2016;**17**:360. DOI: 10.1186/s12864-016-2715-1
- [97] Abdurakhmonov IY, Devor EJ, Buriev ZT, Huang L, Makamov A, Shermatov SE, Bozorov T, Kushanov FN, Mavlonov GT, Abdukarimov A. Small RNA regulation of ovule development in the cotton plant, *G. hirsutum* L. *BMC Plant Biology*. 2008;**8**:93. DOI: 10.1186/1471-2229-8-93
- [98] Kwak PB, Wang QQ, Chen XS, Qiu CX, Yang ZM. Enrichment of a set of microRNAs during the cotton fiber development. *BMC Genomics*. 2009;**10**:457. DOI: 10.1186/1471-2164-10-457
- [99] Pang M, Woodward AW, Agarwal V, Guan X, Ha M, Ramachandran V, Chen X, Triplett BA, Stelly DM, Chen ZJ. Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biology*. 2009;**10**:122. DOI: 10.1186/gb-2009-10-11-r122
- [100] Wang Y, Ding Y, Liu JY. Identification and Profiling of microRNAs expressed in elongating cotton fibers using small RNA deep sequencing. *Frontiers in Plant Sciences*. 2016;**7**:1722. DOI: 10.3389/fpls.2016.01722
- [101] Yu D, Wang Y, Xue W, Fan S, Yu S, Liu JY. Identification and profiling of known and novel fiber microRNAs during the secondary wall thickening stage in cotton (*Gossypium hirsutum*) via high-throughput sequencing. *Journal of Genetics and Genomics*. 2014;**41**:553-556. DOI: 10.1016/j.jgg.2014.08.002
- [102] Chen X, Gao W, Zhang J, Zhang X, Lin Z. Linkage mapping and expression analysis of miRNAs and their target genes during fiber development in cotton. *BMC Genomics*. 2013;**14**:706. DOI: 10.1186/1471-2164-14-706



- [103] Xie F, Sun G, Stiller JW, Zhang B. Genome-wide functional analysis of the cotton transcriptome by creating an integrated EST database. *PLoS One*. 2011;**6**:e26980. DOI: 10.1371/journal.pone.0026980
- [104] Xue W, Wang Z, Du M, Liu Y, Liu JY. Genome-wide analysis of small RNAs reveals eight fiber elongation-related and 257 novel microRNAs in elongating cotton fiber cells. *BMC Genomics*. 2013;**14**:629. DOI: 10.1186/1471-2164-14-629
- [105] Li Q, Jin X, Zhu YX. Identification and analyses of miRNA genes in allotetraploid *Gossypium hirsutum* fiber cells based on the sequenced diploid *G. raimondii* genome. *Journal of Genetics and Genomics*. 2012;**39**:351-360. DOI: 10.1016/j.jgg.2012.04.008
- [106] Hao J, Tu L, Hu H, Tan J, Deng F, Tang W, Nie Y, Zhang X. GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *Journal of Experimental Botany*. 2012;**63**:6267-6281. DOI: 10.1093/jxb/ers278
- [107] Tang W, Tu L, Yang X, Tan J, Deng F, Hao J, Guo K, Lindsey K, Zhang X. The calcium sensor GhCaM7 promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. *New Phytologist*. 2014;**202**:509-520. DOI: 10.1111/nph.12676
- [108] Abdurakhmonov IY, Buriev ZT, Saha S, Jenkins JN, Abdurakarimov A, Pepper AE. Phytochrome RNAi enhances major fibre quality and agronomic traits of the cotton *Gossypium hirsutum* L. *Nature Communications*. 2014;**5**:3062. DOI: 10.1038/ncomms4062
- [109] Abdurakhmonov IYRNA. Interference. Croatia: InTech; 2016. Available from: [www.intechopen.com/welcome/9edcfa43c752e926f9e51ecb610e34db/](http://www.intechopen.com/welcome/9edcfa43c752e926f9e51ecb610e34db/)
- [110] Wang L, Zhu Y, Wang P, Fan Q, Wu Y, Peng QZ, Xia GX, Wu JH. Functional characterization of a dihydroflavanol 4-reductase from the fiber of upland cotton (*Gossypium hirsutum*). *Molecules*. 2016;**21**:32. DOI: 10.3390/molecules21020032
- [111] Wang MY, Zhao PM, Cheng HQ, Han LB, XM W, Gao P, Wang HY, Yang CL, Zhong NQ, Zuo JR, Xia GX. The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. *Plant Physiology*. 2013;**162**:1669-1680. DOI: 10.1104/pp.113.215673
- [112] Zhang M, Han LB, Wang WY, SJ W, Jiao GL, Su L, Xia GX, Wang HY. Overexpression of GhFIM2 propels cotton fiber development by enhancing actin bundle formation. *Journal of Integrative Plant Biology*. 2002;**59**:531-534. DOI: 10.1111/jipb.12552
- [113] Janga MR, Campbell LM, Rathore KS. CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (*Gossypium hirsutum* L.). *Plant Molecular Biology*. 2017;**94**:349-360. DOI: 10.1007/s11103-017-0599-3
- [114] Adhikari J, Das S, Wang Z, Khanal S, Chandnani R, Patel JD, Goff V, Auckland S, Rainville LK, Jones D, Paterson AH. Targeted identification of association between cotton fiber quality traits and microsatellite markers. *Euphytica*. 2017;**213**:65. DOI: 10.1007/s10681-017-1853-0

- [115] Su J, Li L, Pang C, Wei H, Wang C, Song M, Yu S. Two genomic regions associated with fiber quality traits in Chinese upland cotton under apparent breeding selection. *Scientific Reports*. 2016;**6**:38496. DOI: 10.1038/srep38496
- [116] Wang H, Huang C, Zhao W, Dai B, Shen C, Zhang B, Lin Z. Identification of QTL for fiber quality and yield traits using two immortalized backcross populations in upland cotton. *PLoS One*. 2016;**11**:e0166970. DOI: 10.1371/journal.pone.0166970
- [117] Shang L, Wang Y, Wang X, Liu F, Abduweli A, Cai S, Li Y, Ma L, Wang K, Hua J. Genetic analysis and QTL detection on fiber traits using two recombinant inbred lines and their backcross populations in upland cotton. *Genes|Genomes|Genetics*. 2016;**6**:2717-2724. DOI: 10.1534/g3.116.031302
- [118] Badigannavar A, Myers GO. Construction of genetic linkage map and QTL analysis for fiber traits in diploid cotton (*Gossypium arboreum* × *Gossypium herbaceum*). *Journal of Cotton Science*. 2015;**19**:15-26
- [119] Wang H, Huang C, Guo H, Li X, Zhao W, Dai B, Lin Z. QTL mapping for fiber and yield traits in upland cotton under multiple environments. *PLoS One*. 2015;**10**:e0130742. DOI: 10.1371/journal.pone.0130742
- [120] Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J, Liang X, Huang G, Percy RG, Liu K, Yang W, Chen W, Du X, Shi C, Yuan Y, Ye W, Liu X, Zhang X, Liu W, Wei H, Wei S, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology*. 2015;**33**:524-530. DOI: 10.1038/nbt.3208
- [121] Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*. 2014;**46**:567-572. DOI: 10.1038/ng.2987
- [122] Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics*. 2012;**44**:1098-1103. DOI: 10.1038/ng.2371
- [123] MacMillan CP, Birke H, Chuah A, Brill E, Tsuji Y, Ralph J, Dennis ES, Llewellyn D, Pettolino FA. Tissue and cell-specific transcriptomes in cotton reveal the subtleties of gene regulation underlying the diversity of plant secondary cell walls. *BMC Genomics*. 2017;**18**:539. DOI: 10.1186/s12864-017-3902-4
- [124] Miao Q, Deng P, Saha S, Jenkins JN, Hsu C-Y, Abdurakhmonov IY, et al. Genome-wide identification and characterization of microRNAs differentially expressed in fibers in a cotton phytochrome A1 RNAi line. *PLoS One*. 2017;**12**:e0179381. DOI: 10.1371/journal.pone.0179381
- [125] Li X, Wu M, Liu G, Pei W, Zhai H, Yu Z, Zhang J, Yu S. Identification of candidate genes for fiber length quantitative trait loci through RNA-Seq and linkage and physical mapping in cotton. *BMC Genomics*. 2017;**18**:427. DOI: 10.1186/s12864-017-3812-5

- [126] Hu H, Wang M, Ding Y, Zhu S, Zhao G, Tu L, Zhang X. Transcriptomic repertoires depict the initiation of lint and fuzz fibers in cotton (*Gossypium hirsutum* L.). *Plant Biotechnology Journal*. 2017;1467-7652. DOI: 10.1111/pbi.12844
- [127] Thyssen GN, Fang DD, Turley RB, Florane CB, Li P, Mattison CP, Naoumkina M. A Gly65Val substitution in an actin, GhACT\_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. *Plant Journal*. 2017;**90**:111-121. DOI: 10.1111/tpj.13477
- [128] Naoumkina M, Bechere E, Fang DD, Thyssen GN, Florane CB. Genome-wide analysis of gene expression of EMS-induced short fiber mutant Ligon lintless-y (li y) in cotton (*Gossypium hirsutum* L.). *Genomics*. 2017;**109**:320-329. DOI: 10.1016/j.ygeno.2017.05.007
- [129] Ma Q, Wu M, Pei W, Wang X, Zhai H, Wang W, et al. RNA-Seq-mediated transcriptome analysis of a fiberless mutant cotton and its possible origin based on SNP markers. *PLoS One*. 2016;**11**:e0151994. DOI: 10.1371/journal.pone.0151994
- [130] Hinchliffe DJ, Condon BD, Thyssen G, et al. The GhTT2\_A07 gene is linked to the brown colour and natural flame retardancy phenotypes of Lc1 cotton (*Gossypium hirsutum* L.) fibres. *Journal of Experimental Botany*. 2016;**67**:5461-5471. DOI: 10.1093/jxb/erw312
- [131] Zou C, Wang Q, Lu C, Yang W, Zhang Y, Cheng H, Feng X, Prosper MA, Song G. Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). *Science China Life Sciences*. 2016;**59**:164-171. DOI: 10.1007/s11427-016-5000-2