Phylogenetic and Functional Analysis of Cesa Genes in Cotton

Zhuikui Cui

Henan Agricultural University

Guiqin Sun (✉ 2441588775@qq.com)

Jiangxi Agricultural Engineering Vocational college

Quanzhi Zhao

Henan Agricultural University

Research article

Keywords: evolution, function, Cesa, cotton fiber, VIGS, Gossypium hirsutum

Posted Date: November 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-101509/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Cotton (*Gossypium hirsutum*) is widely distributed all over the world, and improving the quality of its fiber is one of the most important tasks in cotton breeding. Cotton fibers are primarily composed of cellulose, which is synthesized and regulated by cellulose synthase (CesAs). But the molecular mechanism of CesA genes in cotton was unclear.

Results: In this study, the phylogenetic history and purifying selection of CesA genes were investigated along with their functions. *CesA3* and *CesA6* are the two largest subgroups in *G. arboreum*, comprising 52.8% of the whole CesA family. These two CesA subgroups were then chosen for further research, and the results showed that they are highly differentiated in dicot groups. The two subgroups were also discriminated with the use of a Ka/Ks analysis. This indicated that they may play an important role in fiber development based on their unique phylogenetic status. Functional studies were subsequently conducted using the most purified genes (*Gohir.A08G144300.1* in CesA3 subgroup). The silencing of *Gohir.A08G144300.1* visibly inhibits the growth of cotton fiber, showing that it is critical for the growth of cotton fibers.

Conclusions: The results presented here a target gene *Gohir.A08G144300.1* based on the analysis of CesA gene members, and it is found that this gene was crucial to the growth of cotton fibers. This study provides more information for the understanding of the molecular mechanism of cotton fiber development.

Background

Cotton, an economic crop, is widely distributed all over the world. The cotton fiber is considered to be the most valuable part of this plant. The fibers coat seeds and primarily consist of cellulose [1]. Cotton is highly effective at synthesizing cellulose and fibers compared with other plants. Thus, the physiological process that underlies this mechanism merits exploration.

Cellulose and hemicellulose are two important components of the cell wall [2]. The CesA gene family has been reported to encode key subunits in the cellulose synthase complex (CSC), which is responsible for cellulose synthesis [3]. Its physiological functions have been studied in some other plant species before.

CesA, a large gene family in cotton plants, acts as a regulatory element during the process of fiber formation and other developmental pathways. In *Arabidopsis*, the CSC consists of *AtCesA1*, *AtCesA3* and *AtCesA6*, which are involved the synthesis of cellulose in primary cell walls (PCWs) [4]. Moreover, CSCs that consist of *AtCesA4*, *AtCesA7* and *AtCesA8* are involved in cellulose synthesis in secondary walls (SCWs) [5]. At least two CSCs have been identified in the xylem cell membrane in *Populus* [6]. Complex I is synthesized by *PtCesA4*, *PtCesA7*, *PtCesA8*, *PtCesA17* and *PtCesA18*, which have been reported to synthesize cellulose in the development of SCWs. Complex II is constituted by *PtCesA3*, *PtCesA10*, *PtCesA11*, *PtCesA13*, *PtCesA15* and *PtCesA16*, which is involved in the formation of both PCWs and SCWs.

Cotton is an important plant for fiber and oil production. Fiber cells primarily undergo two periods during their developmental process: cell elongation, which controls the length of fibers, and secondary cell wall thickening, which determines the strength of fibers. The two periods are arranged in a chronological order and do not overlap [7]. SCW thickening is closely related to the quality of the cotton fiber, which merits further study.

The development of cotton fibers is a complicated process which is regulated by the expression of a series of genes. The CesA gene family plays a central role in this process. A better understanding of its physiological functions and evolutionary history is of vital significance.

With its rapid development, the large-scale application of next generation sequencing (NGS) technology has helped to decipher a substantial amount of genetic and transcriptomic information [8]. NGS technology has been a powerful tool in the research of many plants, including *Arabidopsis* [9], rice [10], soybean [11] and sesame [12]. This bioinformatic method has provided much information about CesA genes at the genomic level. However, the evolution and function of CesA genes in cotton has not been systematically studied. In this study, the evolutionary history of CesA genes in cotton were investigated among a series of selected plant species. In addition to this, the function of representative CesA gene *Gohir.A08G144300.1* was studied, providing more references to expand our understanding of cotton CesA genes at the evolutionary and functional levels.

Results

Evaluation of transcriptome sequencing results

In our previous transcriptome data, six different fiber samples, including three 5DPA and three 15DPA, were obtained using Illumina sequencing technology (San Diego, CA, USA), and the statistical results of transcriptomes are shown in Table 1. The ratio of clean reads and the total mapped rate were 90.0% and 93.0%, respectively. A total of 36G data were obtained after quality control, and the Q30 base ratio was above 92.0%. In addition, the uniquely mapped rate was higher than 94.0%. All these indicators demonstrated that the transcriptomic data were highly accurate for subsequent analyses. Six differentially expressed CesA genes were identified from the transcriptome data (**Table S1**). KEGG pathway analysis showed that the CesA genes were enriched in signaling pathways, such as the ATP binding pathway (Fig. 1).

Table 1
Overview of high-quality of the transcriptome sequencing data

Sample	Clean Reads No.	Clean Reads %	N (%)	Q30 (%)	Total Mapped (%)	Uniquely Mapped (%)
15DPA1	39292978	92.90	0.001869	93.39	37546270 (95.55)	35759208 (95.24)
15DPA2	41001628	93.08	0.001863	93.27	39217817(95.65)	37417619 (95.41)
15DPA1	39213268	92.90	0.001862	93.33	36545369 (93.20)	34587599 (94.64)
5DPA1	38202518	90.82	0.001859	92.96	36437115 (95.38)	34694598 (95.22)
5DPA2	41301746	92.06	0.001853	93.53	39214123 (94.95)	37089080 (94.58)
5DPA3	40624562	93.07	0.001434	92.99	38573488 (94.95)	36695222 (95.13)
Note: Clean Reads No. represents the number of high-quality sequencing reads. Clean Reads (%) represents the ratio of high-quality sequencing reads among all the sequenced reads. N (%) represents the percentage of fuzzy bases. Q30 (%) represents the percentage of bases whose base recognition accuracy is above 99.9%. Total_Mapped (%) represents the total number of clean reads mapped on the reference genome. Uniquely_Mapped (%) represents the total number of clean reads which uniquely mapped on the reference genome.						

Classification And Phylogeny Of The Cesa Gene Family

In order to investigate the evolution of all CesA gene members, the BLASTp method was used to retrieve and identify the CesA genes in *G. arboreum* [13], *G. raimondii* [14] and *G. hirsutum* [15], and 70 CesA genes were finally identified based on the conserved structure of CesA (**Table S2-S4**). MEGA7 software was used to construct the evolutionary tree of CESA. The systematic structure showed that CesA was divided into seven subfamilies (Fig. 2), respectively (*CesA1*, *CesA3*, *CesA4*, *CesA6*, *CesA7*, *CesA8* and *CesA9*). Among of them, 21 *CesA6* genes (30.0%) and 16 *CesA3* genes (22.8%) were discovered with a larger number of gene members, indicating that the subgroup *CesA3* and *CesA6* may be significant in cellulose-related physiological processes in *Gossypium*.

CesA3 and CesA6 genes are highly differentiated in dicot groups

To study the evolutionary history of cotton CesA3 and CesA6 genes and the evolutionary relationship of other model species in more detail, we constructed a phylogenetic tree using the CDS sequences of CesA genes in *Selaginella* [16], grape [17], *Arabidopsis* [18, 19] and cotton with the neighbor joining method.

The phylogenetic tree of four species indicated that all the genes can be categorized into several groups, and almost all AtCesA genes located at the most basal branches (**Fig**

. **3a**). MapChart software was used to draw the diagram of the position of cotton CESA on each chromosome (**Table S5**). The results showed four CesA3 genes and five CesA6 genes were identified among seven chromosomes (**Fig. 3b**), with the gene length varying from 3.0 kb to 4.6 kb (**Table S6**). The CesA genes from other plants, such as *Solanum lycopersicum* and *Arabidopsis thaliana*, were so complicated, since it appeared that they are not clustered based on the organisms or gene subfamilies. However, the CesA3 genes (blue) and CesA6 genes (red) from *G. hirsutum* appeared to be most differentiated, since their position in phylogeny tree is the most advanced. Three CesA3 genes were clustered, indicating their relatively close relationship. However, *Gohir.D13G163700.1*, a member of CesA3, did not appear in concert with the other three CesA3 genes. So we can speculate that the genes from CesA3 and CesA6 are highly differentiated in cotton.

Gene structure and selection pressure of CesA3 and CesA6 during the evolutionary process

The conserved motif structures were predicted to clarify the conserved structure of CesA3 and CesA6 genes that encode proteins. All the CesA3 and CesA6 genes from the three species of *Gossypium* (*G. arboreum*, *G. raimondii* and *G. hirsutum*) were queried to identify motifs in CesA3 and CesA6 to improve the quality of motif prediction. The three most likely motifs were regarded as the final results (**Fig. 4**). The six predicted motifs are highly conserved, with E-values that range from $3.5E^{-268}$ to $2.9E^{-284}$. The length of genes clearly vary (3.0 kb to 4.4 kb in CesA3, 3.8 kb to 4.5 kb in CesA6; **Table S6**), and all genes have the three motifs which are similarly distributed on the transcripts. The order of three motifs also remained unchanged among the CesA3 and CesA6 genes. To understand the evolutionary history of *Gossypium arboreum* CesA3 and CesA6 genes in more detail, the selective pressure (Ka, Ks, Ka/Ks ratio) of the CesA gene family was estimated in the evolutionary process using the *kekeThecc1EG002092t1* [20] gene as a reference (**Table S7**). We statistically analyzed the amount of selective pressure (**Table S5**). The selection pressure of CesA genes during the evolutionary process is less than 1, i.e., indicating that they are subject to purifying selection.

Analysis of the expression patterns of CesA genes in *Gossypium hirsutum*

In order to research the function of CesA genes, expression pattern analysis was conducted. We selected two representative genes to study the CesA gene in more detail. The expression levels of the *Gohir.A08G144300.1* and *Gohir.D05G245300.1* genes in different tissues and fiber development stages were detected at 5, 10, 15, 20 and 25 dpa using fluorescence quantification method and the results were shown in **Fig. 5**. The results indicated that expression level of two genes in the fiber was significantly higher than other tissues. Besides, expression levels of two genes at different fiber development stages were analyzed. The *Gohir.D05G245300.1* gene reached a peak at 25 dpa in fiber development. However,

the *Gohir.A08G144300.1* gene reached a peak at 10 and 15 dpa in fiber development and then decreased significantly. This gene was highly expressed throughout the entire fiber period.

Gohir.A08G144300.1 gene may play important roles in cotton fiber development

To study the possible biological functions of CesA genes during the process of plant growth, we selected *Gohir.A08G144300.1* to construct the CesA plant interference vector [21] and quantified the results of cotton phenotype (Fig. 6). The number and size of cotton bolls was reduced (Fig. 6a, b), with the average length decreasing by 20% from 35 mm (wild type, wt) to 28 mm. The growth of *Gohir.A08G144300.1* fibers was inhibited significantly compared with that of the control group (CK) (Fig. 6c-f). The fiber length was shortened (Fig. 6e and f), with the average length decreasing by 13% from 25 mm to 18 mm. The variations of cotton fibers may be caused by the changes in the cell volume and number.

Fiber cell glass method was used to study the cause of reduced fibers. Significant differences in cell volume and number and fiber length were observed between silencing plants in which target gene *Gohir.A08G144300.1* was silenced and normal plants. Fiber cell glass method was used to study the cause of reduced fibers of the TRV-CaCesA knockout and the film was observed under a 200X optical microscope (Fig. 6). The results suggested that this gene play an important role in fiber growth.

Discussion

Evolutionary analysis of the CesA genes

Cotton is a highly valuable resource plant species. Improving the yield and quality of its fibers is one of the essential issues in the field of cotton breeding. In recent years, the gradual improvements in technology have resulted in substantial progress in the study of developmental mechanisms for cotton fibers [22–24]. This has been made possible by such new technologies as RNA-Seq and Chip-Seq, which enable studies of the regulation of transcriptomes, proteomes, metabolomes and transcription and translation. The mechanism of cotton fiber development has been analyzed from many different perspectives, and a large number of fiber-related genes have been excavated. The CesA gene family is an essential component of cellulose synthase and is simultaneously directly responsible for the process of production of cotton fibers. In this study, we analyzed the cotton fiber transcriptome and found that the CesA genes are widely involved in the process of cotton fiber development. We focused on the phylogeny and function of the CesA genes. We also investigated the phylogenetic history of CesA genes in cotton (*G. hirsutum*) and examined *A08G144300.1* at the functional level.

It can easily be deduced that the CesA3 and A6 genes from *G. hirsutum* are located on the most differentiated branches on the dicot phylogenetic tree (Fig. 2a), which suggests that there are strong links

between the CesA genes and specific characteristics in cotton. We used the evolutionary node plants *S. lycopersicum*, *V. vinifera* and Arabidopsis to examine the evolutionary history of CesA genes in more detail. We did not compare CesA3 *Gohir.A08G144300.1* and CesA6 *Gohir.D05G245300* with the genes in other species. Notably, we also found that these results are not caused solely by the chronological order of the species along the overall physiological history. Therefore, we concluded that *Gohir.A08G144300.1* and *CesA6 Gohir.D05G245300* are very import for cotton and retention it.

Functional Analysis Of The Cesa Genes

There are many members of the CesA transcription factor family that are largely involved in the regulation of plant growth and development, response to biological stress and other processes [25–27]. The *AtCesA3* and *AtCesA6* genes in *A. thaliana* are involved in the composition of primary cell wall and cellulose synthesis. With the advent of the post-genomic era, increasing numbers of functions for CesA gene have been verified [28–31]. However, CesA is rarely reported in genes that regulate the development of fiber. In this study, a series of biochemical and functional identification experiments, such as the determination of gene structure and Ka/Ks and the quantification of fluorescence, were conducted on CesA genes The structure of CesA gene family is highly conservative, and these genes appear to have been the subject of selection during the process of evolution. To study the potential function of the Cest genes in more detail, we constructed the A08G144300.1 gene interference vectors and found that the length of fibers of the TRV-CmANT-1 and TRV-CesA.A08G144300.1 knockout strains was significantly reduced compared with those of the TRV control group. The fiber length appears visually as significantly shorter. Similar results have been observed in other species, For example, when the CESA gene is mutated in Arabidopsis [32, 33], the synthesis of cellulose is hindered. This leads to a decrease in the thickness of the cell wall, which causes a series of changes in cell morphology, that result in a new perspective on the study of signal pathways of *A08G144300.1*. In addition, the *Gohir.D05G245300.1* gene was found to be highly expressed in the fiber. This implies that this gene plays an important role in fiber growth. We will present an advanced analysis of their detailed functions in a future report.

Conclusions

In this study, the phylogenetic history of CesA gene families, and the function of a representative gene *Gohir.A08G144300.1* were investigated. *CesA3* and *CesA6* were two are the two largest subgroups in CesA gene family, and target gene *Gohir.A08G144300.1* play an vital role in cotton fiber development by functional analysis, providing more insights into the research of molecular research of CesA genes in cotton. Although we obtained primary functional information of the genes, more experimental and computational evidence is needed to fully elucidate the function of the genes and process of cotton fiber growth.

Methods

Materials

A standard system for genetics is provided by the National Cotton Breeding Center of the Cotton Research Institute of the Chinese Academy of Agricultural Sciences (Zhengzhou, China). New roots, stems, young leaves, flowers (during the flowering period) and different stages of fibers during development were frozen in liquid nitrogen and stored at -80°C until their RNA was extracted for an analysis of tissue expression.

Identification of the cotton CESA gene family

All of the gene sequence and annotation information of the cotton CESA family was downloaded from the CFGD (<http://www.cottonfgd.org/>) database [34] under the accession number iPF03552 (**Table S2–S4**). Six cotton CESA protein domain sequences were used as probes to examine other families using the probe search method [35], and the results were compared with the Phytozome botanical genome database (<https://phytozome.jgi.doe.gov/pz/portal.html#>) (**Additional file 8**). The e-value was set to 10^{-5} . The transcriptome data was completed by Shanghai Personalbio (Shanghai, China).

An Illumina HiSeq platform was used to sequence the genes. The raw data obtained from the computer was processed by quality control, filtering, comparing, and other analyses. Clustal X1.8 software was used to perform multiple sequence alignment of protein sequences, and the alignment results were analyzed using MEGA7 [36]. MapChart software was used to map the distribution of genes on chromosomes. The online software MEME [37] was simultaneously used to pre-analyze the conserved motifs of proteins using the D cotton CESA sequence, and KaKs_Calculator2.0 software [38] was used to calculate the selection pressure between collinearity genes.

Quantitative real-time PCR

The cotton plant was split into several parts, including the root, shoot, leaves, flowers and fibers. Fibers were collected at different stages (5, 10, 15, 20 and 25 d) after flowering. The RNA was extracted separately. A quantitative real-time PCR experiment was conducted using TIANGEN RealUniversal Color PreMix (SYBR Green) (QKD-201, Tiangen Biotech, Beijing, China) following the manufacturer's instructions. GaHistone3 was used as the reference gene.

Paraffin sectioning

Three-day-old cotton fibers were fixed in a fixing solution that contained 4% FAA (formaldehyde-glacial acetic acid-absolute ethanol) for 24 hours, dried under vacuum and then incubated overnight at 4°C. The samples were dehydrated using a series of gradient concentrations of ethanol (50%, 70%, 85%, 95% and 100%) with each gradient consisting of 30 min. The soaked tissues were embedded in liquid paraffin and then cooled at -20°C. The samples were cut into 4 µm-thick sections with a paraffin section base. The sections were suspended in a 40°C water bath to flatten them before their placement on a glass slide.

They were dried overnight at 37°C. The sections were then stained with safranin and Fast Green, and photographed with a digital camera under a microscope.

Virus-induced gene silencing (VIGS) experiments

Full-length *Gohir.A08G144300.1* was amplified from *G. arboreum* cDNA (**Table S6**). The *EcoRI* and *KpnI* sites of pTRV:RNA2 were used as cloning sites, and the target sequences were inserted. The recombinant vectors were transformed into *Agrobacterium* GV3101 competent cells following the manufacturer's instructions. The *Agrobacterium* transformant was cultivated in liquid Luria-Bertani (LB) medium containing 25 µg/mL⁻¹ rifampicin to an OD₆₀₀ from 1.8 to 2.2. The OD₆₀₀ of the medium was adjusted to 1.5 using buffer that contained 0.5 mol/L⁻¹ MES, 200 mmol/L⁻¹ acetosyringone and 1 mol/L⁻¹ MgCl₂ for transfection. Liquid medium containing pTRV:RNA2-Gohir.A08G144300.1 and pTRV:RNA2 were mixed with pTRV at an equimolar ratio. They were then injected into cotton cotyledons at the three-leaf stage. Each group had 10 replications. The cotton plants were exposed to negative pressure and subsequently grown in the dark for 48 h. The plants were then moved to a greenhouse and grown under a 16-hour photoperiod for 30 d. Finally, they were grown under an 8-hour photoperiod to trigger the differentiation of flower buds.

Declarations

Author Contributions: Conceived and designed the experiments: Z.C., G.S. and Q.Z. Performed the experiments: Z.C. Analyzed the data and wrote the paper: Z.C.

Funding: This project was supported by the Special fund for Modern Agriculture of Jiangxi Province (JXARS-22) and Science and Technology Research Project of Education Department of Jiangxi Province (181366).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Krakhmalev VA, Zakirov TA. In vivo observations on the structure and growth of cotton ovule hairs. Russ J Plant Physiol +. 2000;47(2):249–55.

2. Zhang N, Li S, Xiong LM, Hong Y, Chen YP. Cellulose-hemicellulose interaction in wood secondary cell-wall. *Model Simul Mater Sc* **2015**, 23, (8).
3. Kumar M, Atanassov I, Turner S. Functional analysis of cellulose synthase (CESA) protein class specificity. *Plant physiology* 2017, 173, (2): 970–983.
4. Wang J, Elliott JE, Williamson RE. Features of the primary wall CESA complex in wild type and cellulose-deficient mutants of *Arabidopsis thaliana*. *J Exp Bot*. 2008;59(10):2627–37.
5. Atanassov II, Pittman JK, Turner SR. Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of *Arabidopsis* secondary cell walls. *The Journal of biological chemistry* 2009, 284, (6): 3833–41.
6. Xi W, Song D, Sun J, Shen J, Li L. Formation of wood secondary cell wall may involve two type cellulose synthase complexes in *Populus*. *Plant Mol Biol* 2017, 93, (4–5): 419–429.
7. Gou JY, Wang LJ, Chen SP, Hu WL, Chen XY. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell research* 2007, 17, (5): 422 – 34.
8. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333–51.
9. Loraine AE, McCormick S, Estrada A, Patel K, Qin P. RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant physiology* 2013, 162, (2): 1092 – 109.
10. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 2013, 6.
11. Chaudhary J, Patil GB, Sonah H, Deshmukh RK, Vuong TD. Expanding Omics Resources for Improvement of Soybean Seed Composition Traits. *Front Plant Sci* 2015, 6.
12. Wang LH, Yu S, Tong CB, Zhao YZ, Liu Y: Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome biology* 2014, 15, (2).
13. Li FG, Fan GY, Wang KB, Sun FM, Yuan YL. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature genetics* 2014, 46, (6): 567–572.
14. Wang KB, Wang ZW, Li FG, Ye WW, Wang JY. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature genetics* 2012, 44, (10): 1098–+.
15. Li FG, Fan GY, Lu CR, Xiao GH, Zou CS. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature biotechnology* 2015, 33, (5): 524–U242.
16. Weng JK, Tanurdzic M, Chapple C. Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. *BMC genomics* 2005, 6.
17. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007;449(7161):463–U5.
18. Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 2012, 485, (7400): 635–641.

19. Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000, 408, (6814): 796–815.
20. Liu X, Zhao B, Zheng HJ, Hu Y, Lu G. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific reports* 2015, 5.
21. Wu CJ, Jia LL, Goggin F. The reliability of virus-induced gene silencing experiments using tobacco rattle virus in tomato is influenced by the size of the vector control. *Mol Plant Pathol*. 2011;12(3):299–305.
22. Kim HJ, Hinchliffe DJ, Triplett BA, Chen ZJ, Stelly DM. Phytohormonal Networks Promote Differentiation of Fiber Initials on Pre-Anthesis Cotton Ovules Grown In Vitro and In Planta. *Plos One* 2015, 10, (4).
23. Zou C, Wang Q, Lu C, Yang W, Zhang Y. Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). *Science China Life sciences*. 2016;59(2):164–71.
24. Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nature communications* 2014, 5.
25. Suzuki S, Li L, Sun YH, Chiang VL. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant physiology* 2006, 142, (3): 1233–45.
26. Nobles DR, Romanovicz DK, Brown RM Jr. Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? *Plant physiology* 2001, 127, (2): 529 – 42.
27. Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP. A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant physiology* 2000, 123, (4): 1313–24.
28. Desprez T, Vernhettes S, Fagard M, Refregier G, Desnos T. Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform CESA6. *Plant physiology*. 2002;128(2):482–90.
29. Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR. The irregular xylem3 locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell*. 1999;11(5):769–80.
30. Ellis C, Turner JG. The *Arabidopsis* mutant *cev1* has constitutively active jasmonate and ethylene signal pathways and enhanced resistance to pathogens. *The Plant cell* 2001, 13, (5): 1025–33.
31. Yoshikawa T, Eiguchi M, Hibara KI, Ito JI, Nagato Y. Rice SLENDER LEAF 1 gene encodes cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell proliferation. *J Exp Bot*. 2013;64(7):2049–61.
32. Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z. Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. 2007;104(39):15572–7.

33. Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. *Science*. 2006;311(5769):1940–2.
34. Zhu T, Liang C, Meng Z, Sun G, Guo S. CottonFGD: an integrated functional genomics database for cotton. *BMC plant biology* 2017, 17, (1), 101.
35. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*. 2012;40(Database issue):D1178-86.
36. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* 2018, 35, (6), 1547–1549.
37. Bailey TL, Boden M, Buske FA, Frith M, Grant CE. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* 2009, 37, W202-W208.
38. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, 25, (11): 1451–2.

Figures

Figure 1

GO enrichment of differentially expressed genes in *Gossypium hirsutum* L. Red arrow indicated that the ATP binding pathway in which CesA genes were enriched.

Figure 4

Motif prediction and distribution of CesA3 and CesA6 genes in *G. arboreum*. Motif analysis in *G. arboreum* CesA3 (a) and CesA6 (b). Three motifs were recognized for each gene family. Identified motifs are completely different between CesA3 and CesA6 family. Motif distribution resembles within gene families. a) Motif analysis in *G. arboreum* CesA3. b) Motif analysis in *G. arboreum* CesA6.

Figure 5

Relative expression analysis of Gohir.A08G144300.1 and Gohir.D05G245300.1

Figure 6

Functional analysis of Gohir.A08G144300.1 (CesA3) a) The results of Paraffin section of normal cells. b) The results of Paraffin section of gene-silencing (Gohir.A08G144300.1) cells. c) Morphology of normal cotton bolls. d) Morphology of gene-silencing (Gohir.A08G144300.1) cotton bolls. e) The length of normal cotton bolls. f) The length of gene-silencing (Gohir.A08G144300.1) cotton bolls.

cotton fiber. f) The cotton fiber length of gene-silencing (Gohir.A08G144300.1) plants was obviously shorter than the normal. g) Gene expression analysis in gene silencing plants and normal plants. CK1, CK2 and CK3 represent three normal plants, and S1, S2 and S3 represent three gene silencing plants.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [renamed8fccc.xlsx](#)
- [renamed8bde9.xlsx](#)
- [renamed3198d.xlsx](#)
- [renamed4b502.xlsx](#)
- [renamedc91b9.xlsx](#)
- [renamed2ce87.xlsx](#)
- [renamed1f249.docx](#)
- [Additionalfile8.txt](#)