Evolution of pectin-synthesis-relevant galacturonosyltransferase gene family and its expression during cotton fiber development

FAN Senmiao (✉ fansenmiao@caas.cn)

CAAS

LIU Aiying

CAAS

ZOU Xianyan

CAAS

ZHANG Zhen

CAAS

GE Qun

CAAS

GONG Wankui

CAAS

LI Junwen

CAAS

GONG Juwu

CAAS

SHI Yuzhen

CAAS

DENG Xiaoying

CAAS

JIA Tingting

CAAS

YUAN Youlu

CAAS

SHANG Haihong

CAAS & Zhengzhou University

Research

Keywords: Cotton fiber development, Pectin, Galacturonosyltransferases, Evolution, Expression patterns

Posted Date: January 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-152342/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Cotton Research on August 2nd, 2021. See the published version at <https://doi.org/10.1186/s42397-021-00099-z>.

Abstract

Background: Pectin is a key substance involved in cell wall development, and the galacturonosyltransferases (GAUTs) gene family is a critical participant in the pectin synthesis pathway. Systematic and comprehensive research on GAUTs has not been performed in cotton. Analysis of the evolution and expression patterns of the GAUT gene family in different cotton species is needed to increase knowledge of the function of pectin in cotton fiber development.

Results: In this study, we identified 131 *GAUT* genes in the genomes of four *Gossypium* species (*G. raimondii*, *G. barbadense*, *G. hirsutum*, and *G. arboreum*), and classified them as GAUT-A, GAUT-B and GAUT-C. Among them, 15 GAUT genes encoded proteins (GAUT1 to GAUT15). All GAUT family genes except for the gene GAUT7 coding contained a conserved Glyco_transf_8 domain (H-DN-A-SVV-S-V-H-T-F). The conserved sequences of GAUT7 was a PLN (phospholamban) 02769 domain, categorized as a probable galacturonosyltransferase. According to cis-element analysis, GAUT genes expression may be regulated by hormones such as JA, GA, SA, ABA, Me-JA and IAA. The evolution and expression patterns of the GAUT gene family in different cotton species and the expression levels in upland cotton materials having different fiber strengths were analyzed. Peak expression of *GhGAUT* genes was observed before 15DPA; in the six materials with high fiber strength, the expression was concentrated from 10 to 15DPA; while the highest expression in low fiber strength materials was detected between 5 and 10 DPA. The results presented in this paper lays the foundation for future research on gene function during cotton fiber development.

Conclusions: The GAUT gene family may affect cotton fiber development, including fiber elongation and fiber thickening. In the low-strength-fiber lines, GAUTs mainly participate in fiber elongation, whereas their major effect on cotton with high-strength fiber is related to both elongation and thickening.

Background

Cotton, the world's most important economic crop, is tightly connected to the development of the national economies and human livelihoods. Cotton fiber formation and development is an important factor in the cotton breeding (Fan. 2013). The cell wall structure of cotton fiber is composed of pectin and cellulose. The synthesis and decomposition of pectin are therefore important factors affecting cotton fiber formation (Basra et al. 1984).

Pectin, which plays a critical role in plant growth and development, is mainly found in the primary and middle layers of plant cell walls during plant growth and development, tissue structure formation, and stress. Pectin biosynthesis is estimated to require at least 67 transferases including glycosyl-, methyl-, and acetyltransferases (Mohnen et al. 2008). Pectin is also the most complex cell wall polysaccharide (Scheller et al. 2007). The sugar nucleotide reverse transporter transports the sugar nucleotide to the Golgi apparatus under the action of a specific glycosyltransferase. The glycosyl group is then cleaved

and attached to the elongating polysaccharide chain to form pectin. The synthesis of pectin involves at least 53 different glycosyltransferases localized on the Golgi apparatus (Ridley et al. 2001).

Only two glycosyltransferase-encoding genes are recognized on the basis of previous research: homogalacturonan (HG) synthase and glucuronyltransferase involved in rhamnogalacturonanII (RG II) synthesis (Willats et al. 2006). Compared with currently cataloged HG pectin multimerism in plants, much research has been performed on galacturonyltransferase (HG: α -1,4-D-galacturonosyltransferase, HG:GalAT). The gene encoding HG:GalAT (EC 2.4.1.43) is named galacturonosyltransferase (GAUT) (Mohnen et al. 2008; Willats et al. 2006).

Galacturonosyltransferases (GAUTs), which are partly responsible for pectin biosynthesis, are glycosyltransferase (Harholt et al. 2010). According to evolutionary analysis, they constitute glycosyltransferase family 8 (GT8), GT8 consists of three separate protein classes, classes I and II contain mostly eukaryotic proteins, while class III consists almost entirely of bacterial proteins (Yin et al. 2010). Plant cell-wall-related proteins, including GAUT and GAUT-like (GATL) proteins, are all located in class I (Sterling et al. 2006). The GAUT gene family was first identified by Blast analysis of the Arabidopsis genome, on the basis of structural similarity, 15 GAUT and GATL family members have been classified as follows, GAUT1-GAUT7 in GAUT-A; GAUT8-GAUT11 in GAUT-B; and GAUT12-GAUT15 in GAUT-C, All GATL family members are clustered together and constitute a clade closely related to GAUT15. Experiments have shown that GAUT1 is equivalent to HG: GalAT and provides the necessary molecular machinery for HG and pectin synthesis (Sterling et al. 2006). GAUT1 also belongs to GT family 8, and the GAUT1-related superfamily also contains the 10 GAUT-like genes (Cantarel et al. 2008). The *gaut1* mutation can cause plant dwarfing, reduced cell adhesion, and a 25% reduction in GalA content of leaves (Orfila et al. 2005). The GAUT1-GAUT7 Core complex is held together by one or more covalent disulfide bonds and other noncovalent interaction. GAUT1 is dependent on GAUT7 and remains in the Golgi apparatus, where is considered to be a component of the HG:GalAT complex that participates in pectin synthesis (Atmodjo et al. 2011). The *gaut8* mutant can reduce the adhesion of epidermal cells of the young leaves and the marginal cells of the roots, thereby leading to plant dwarfing (Durand et al. 2009). The *gaut11* mutation can reduce the thickness of the seed mucus (Caffall et al. 2009). In the *gaut13 gaut14* double mutant, the distribution of pectin in the pollen tube wall is altered, which results in serious defects in pollen tube shape and growth (Wang et al. 2013).

The GAUT family is large, and more than 67 members have been found in tomato (*Solanum lycopersicum*). The GAUT family member with the highest expression level in tomato is *gaut4* (Godoy et al. 2013). In the *gaut4* mutant, pectin structure is changed significantly, and other fruit traits, such as starch content, fruit yield, and single fruit quality, are also altered, consistent with an observed increase in firmness. In addition, harvest index is significantly decreased because of a reduction in fruit weight and number (Godoy et al. 2013). Hyodo et al. have demonstrated that the *gaut1* gene is expressed continuously during fruit development, especially in the fruit epidermis, while *gaut1* is first expression in the mesocarp, endocarp, septum, locular tissue, and nucleus after the fruit enters maturity (Hyodo et al. 2013).

To improve the quality of high-fiber cotton materials, our research team has been focused on cotton fiber development. Zhang et al. used Recombinant Inbred Lines (RILs) to identify 16 stable quantitative trait loci (QTLs) related to fiber strength (Zhang et al. 2017). In addition, Zou et al. identified 3364 candidate genes related to cotton fiber strength, and analyzed differences in fiber development (5 to 30 DPA) using two extreme materials in a RIL population. A total of 363 differentially expressed genes (DEGs) comprising 4, 75, 39, 62, and 183 genes at 10, 15, 20, 25, and 30 DPA were detected. Among them, 228 (62.6%) genes were upregulated in high strength materials, while *Gh_A07G1907* (GAUT6) was downregulated at 15DPA and upregulated at 25DPA (Zou et al. 2019).

As discussed above, the GAUT gene family has been studied in plants such as *Arabidopsis* and tomato, et al., but no systematic investigation has been carried out in cotton. In the current study, we systematically and comprehensively analyzed the chromosomal location, structures, and phylogenetic relationships of the GAUT gene family in different *Gossypium* species. We also focused on the expression of these genes during the fiber development.

Results

Identification of cotton *GAUT* genes

Detailed phylogenetic analyses have divided GT8 proteins into two distantly related clades: 1) the GAUT1 (galacturonosyltransferase1)-related family, including the GAUT and GATL proteins, known as galacturonosyltransferase proteins, 2) a group including plant glycogen protein-like starch starters (PGSIPs) and galactitol synthases (GolSs) (Yin et al. 2018). According to the Pfam database and a bioinformatics analysis, all inferred proteins have a Glyco_transf_8-like domain (PF01501), which indicates that the corresponding genes belong to the GAUT gene family (Kikuchi et al. 2003). To identify the GAUT gene in *Gossypium* species, we identified 187 GAUT genes from eight *Gossypium* species (Fig S1), including 131 genes from the following cotton species: *G. hirsutum* (41 genes), *G. barbadense* (42 genes), *G. arboreum* (25 genes) and *G. raimondii* (23 genes) (Additional file 3). The length range of coding regions in the *GAUT* gene family ranged from 1,098 to 4,899 bp, and the encoded proteins comprised of 365 to 1,632 amino acids. The length of 32 *GAUT* genes was less than 3,000 bp; 88 had lengths of 3,000 to 7,000 bp, while the remaining 11 were longer than 7,000 bp (Fig S2: Table S1).

Phylogenetic analysis and classification of the cotton GAUT gene family in cotton

Using published genome sequencing data for eight species, we determined the phylogenetic relationships of GAUT gene family members from multiple species. In a previous study (Sterling et al. 2006), three types of gene sequences, GAUT-A, GAUT-B, and GAUT-C, were found by multiple sequence alignment of the 187 GAUT genes of these eight species to *Arabidopsis* homologs (Fig 1a). In the present study, we analyzed the GAUT genes of four cotton species, thereby classifying 62 genes into GAUT-A, 38 into GAUT-B, and 31 into GAUT-C (Fig 1b). In addition, GAUT6 had the largest number of homologs, 20, in *Gossypium* species, and *GhA07G1907*, identified in a previous transcriptome analysis (Zou et al. 2019), was assigned to GAUT6. We found 16 GAUT13 homologs, 14 homologs each of GAUT7, GAUT9, and

GAUT11, and 11 homologs each of GAUT2 and GAUT12. The remaining genes had fewer than 10 homologs. No genes homologous to GAUT14 were detected in any of the four cotton species, and no GAUT5 homologs were identified in *G. hirsutum*. Only one homolog each of GAUT5 and GAUT10 were detected, namely, *GrGAUT18* and *GhGAUT03*, respectively. Four homologs each of GAUT2, 4, 6, 7, 9, 12, and 13, three homologs each of GAUT8 and GAUT 11, and two homologs each of GAUT-A genes GAUT1 and GAUT3 were discovered in *G. hirsutum*.

Analysis of conserved motif and GAUT gene structures

The following motif is conserved in 15 Arabidopsis GAUT genes and their orthologues in cotton: H-DN-A-SVV-S-V-H-T-F (H-x(2)-[ILV]-x-[ST]-D-N-[IV]-[IL]-A-[ASTV]-S-V-V-[AIV]-x-S-x-[AIV]-x(2)-[AS]-x(2)-[PS]-x(3)-V-[FL]-H-[ILV]-[ILV]-T-[DN]-x (2)-[NST]-x (2)-[AGP]-[IM]-x (3)-F) (Sterling et al. 2006). The GAUT gene family encodes a protein with a molecular mass between 61 and 78 kDa (Sterling et al. 2006; Godoy et al. 2013). Consistent with topological predictions, most GAUT genes can encode a type II membrane protein containing a putative transmembrane domain in its hypervariable N-terminal region (Reithmeier et al. 1992). Among the GAUT proteins analyzed in our study, three GAUT proteins (GAUT 3, 4, and 5) which belonging to GAUT-A contained an N-terminal signal peptide rather than a transmembrane domain. The only GAUT gene family members predicted to have no N-terminal transmembrane domain or signal peptide in cotton were GAUT2 genes (Fig 2b). We also found some GAUT1, GAUT3 and GAUT11 genes with the above characteristics. Among 14 GAUT7 genes belonging to group GAUT-A in eight *Gossypium* species (Fig S2: Table S1), 10 contained the same conserved as PLN02769 domain (Fig 2b), which was assigned to the the category of “Probable Galacturonosyltransferase” (<https://www.ncbi.nlm.nih.gov/proteinclusters/?term=PLN02769>). The remaining GAUT family members contained a conserved Glyco_transf_8 domain. The prediction motifs of each member are shown in Fig 2C, and specific structural information is given in Fig S3.

Analysis of collinearity and repeating elements in the GAUT gene

According to the results of MCScan analysis, there were no tandem repeat elements were present in the GAUT gene family in *Gossypium*. In *G. hirsutum*, *GhGAUT01* and *GhGAUT19*, *GhGAUT15* and *GhGAUT34*, *GhGAUT16* and *GhGAUT36*, *GhGAUT17* and *GhGAUT38*, and *GhGAUT18* and *GhGAUT39* were homologous to the GAUT 3, GAUT 11, GAUT 12, and GAUT 13 respectively, and were also segment repeats. Genes in diploid *Gossypium* species corresponding to the above repeated genes are shown in Fig 3. According to their relative order in *Gossypium*, GAUT genes were categorized into five groups (1 to 5). With the exception of group 4, which belonged to GAUT-A, all groups were members of GAUT-C.

Analysis of GAUT expression patterns

As determined by collinearity and repetitive element analyses of the above homologous genes in combination with transcriptome data from different fiber developmental stages of diploid species, *GaGAUT02* had the highest expression at 15DPA but almost no expression during other periods. The other members of group 1 (Fig 3) were also barely expressed in *G. raimondii* and tetraploid cotton, with

fragments per kilobase of transcript per million fragments mapped (FPKM) values of less than 2.0. In group 2, *GhGAUT15* and *GhGAUT34* had the highest expressions during the late stage (20 to 25DPA) of fiber development (Fig 4). Group 3 members *GhGAUT16*, *GhGAUT36*, *GrGAUT14*, and *GaGAUT22* were not expressed at all during the fiber developmental period. Among group 4 genes, *GhGAUT17* and *GhGAUT38* had their highest expression during fiber developmental from 5 to 10 DPA, and *GrGAUT22* and *GaGAUT24* had the expression peaks at 0 and 15 DPA, respectively. All members of group 5 except for *GaGAUT25* were expressed at 15 DPA, and the remaining genes had almost no expression during fiber development. In four cotton species, six genes had peak FPKM values greater than 40, namely *GaGAUT08*, *GaGAUT12*, *GaGAUT13*, *GrGAUT03*, *GrGAUT18*, and *GhGAUT25*. The expression peaks of these six genes all occurred before 15 DPA, which suggests that the GAUT gene family plays a role in early cotton fiber development.

We selected a RIL population containing high-strength fiber and low-strength fiber lines for quantitative real time polymerase chain reaction (qRT-PCR) analysis (Wang et al. 2014). For this analysis, we selected *GhGAUT08* (*Gh_A07G1907*) (Zou et al. 2019) and *GhGAUT25*, both belonging to GAUT-A and *GhGAUT10*, *GhGAUT11*, and *GhGAUT29* belonging to GAUT-B and having FPKM values greater than 10 (Fig S4: Table S2). The qRT-PCR analysis revealed that the GAUT gene family has an important influence on fiber development before 15 DPA. From 5 to 30 DPA, the overall expressions of *GhGAUT08* and *GhGAUT10* were higher in high-strength materials than in low-strength materials. At 5 DPA, the expressions of *GhGAUT11* and *GhGAUT29* were higher in low-strength materials than in high-strength materials, with the opposite true from 10 to 30 DPA. *GhGAUT25* expression was higher in low-strength materials than high-strength materials from 5 to 10 DPA, with the reverse pattern observed after 15DPA. In six high-strength materials, peak GAUT expression was from 10 to 15 DPA (Fig 5), whereas the period of highest expression in six low-strength materials was 5 to 10 DPA (Fig 5).

The GAUT gene family affects the synthesis of pectin, we measured the pectin content of different materials. The results showed that the peak pectin content of high-strength fiber materials appeared at 15DPA, while the low-strength materials appeared at 10DPA (Fig 6). This result was similar to the description of gene expressions.

Prediction of *cis*-elements in GAUT genes promoter regions

To investigate the potential reasons for different expression patterns among GAUT genes, we analyzed the promoter elements of the GAUT gene family in upland cotton. This analysis was performed because *cis*-elements (Fig 7) can affect gene expression regulation (Higo et al. 1999). We analyzed 32 *cis*-elements in upland cotton genes, namely, elements responsive to anaerobic conditions, different hormones (Me-JA, GA, SA, ABA and IAA), and abiotic stress conditions, including drought, low temperature and other stress conditions. CGTCA and TGACG motifs are associated with responsiveness to methyl jasmonate (Me-JA) (Basyuni et al. 2018), while the GARE-motif (TCTGTTG) and P-box (CCTTTTG) are related to gibberellin (GA) (Porto et al. 2014). The TCA-element (CCATCTTTT) is a salicylic acid (SA) responsive element (Herrera et al. 2015). AuxRR-core (GGTCCAT) and the TGA-element

(AACGAC) are related to the responsiveness to auxin (Herrera et al. 2015), and the ABRE-motif (ACGTG) is associated with the abscisic acid (ABA) response (Mishra et al. 2014). The ARE-motif (AAACCA) is related to the anaerobic environment, while the LTR-motif (CCGAAA) participates in response to low temperature (Chen et al. 2018). Finally, the TC-rich repeats (ATTCTCTAAC) is the response element associated with stress (Wei et al. 2009). In future work, we plan to further verify the regulation of the above hormones on the GAUT gene family in upland cotton.

Expression analysis of prominent fiber-expressed genes under abiotic stress and phytohormone treatments

For a more in-depth study of *GhGAUTs* expression levels induced by abiotic stress, the expression patterns of four *GhGAUT* genes after NaCl, PEG, abscisic acid (ABA), naphthylacetic acid (NAA), salicylic acid (SA), and methyl jasmonate (MeJA) treatment were analyzed by qRT-PCR (Fig 8). We examined the effects of various hormones on the expression of the five *GhGAUT* genes. We observed that within 1 hour after all treatments, the relative expression levels of these five genes were rapidly increased and decreased after 24 hours. The peak of the up-regulation response were between 3 hours and 12 hours, except for *GhGAUT11*. *GhGAUT11* did not respond to the treatment of the three hormones for ABA, SA, and MeJA. *GhGAUT29* responds to all stress and hormone treatments, and the detected expression level was higher; *GhGAUT08* responded to ABA and SA treatment with the higher expression levels. *GhGAUT25* also responded to two stresses and four hormone treatments, but the expression levels were the highest under the treatment of PEG and ABA, and the response peaks were at 6h and 12h, respectively. *GhGAUT10* also responded to the treatment of ABA, SA, and MeJA, with the peak expression level at 6h, 12h, and 12h; the response level to NaCl, PEG and NAA treatment was low, with slight response at 3h and 6h, respectively. The level was low. These results indicated that after different hormone treatments, different genes had different response times and response patterns, which were closely related to their hormone response elements and expression patterns.

Discussion

Phylogenetic analysis of the GAUT gene family

Research on the GAUT gene family is currently focused on Arabidopsis (Sterling et al. 2006; Cantarel et al. 2008), tomato (Godoy et al. 2013; Vasco et al. 2011), ramie (Chen et al. 2014), sweet cherry (Campoy et al. 2015), and the other species. Studies have also been carried out in cotton, but a systematically conducting evolutionary analysis has not yet been conducted. Among the 131 GAUT members in cotton, 20 are homologous to GAUT6; no genes in cotton are homologous to GAUT14, and no GAUT5 homologs are present in *G. hirsutum*.

The GAUT gene family encodes galacturonosyltransferases involved in the synthesis of pectin, a compound related to cell wall formation. The conserved motif of the GAUT gene family is H-DN-A-SVV-S-V-H-T-F. In tomato, GAUT1 had been predicted to be a membrane protein with a single N-terminal transmembrane helix and a major globular domain in the Golgi cavity, but no pectin biosynthetic enzyme

has been found. Analysis of the GAUT gene family has shown that all members share a conserved glycosyltransferase family 8 domain of (Godoy et al. 2013; Vasco et al. 2011). In addition to possessing the conserved Glyco_transf_8 domain, however, GAUT genes in the four species of cotton have a PLN02769 domain; 14 genes are homologous to GAUT7, two of which are present in *G.raimondii*, with the other three species having an average of 4. We speculated that changes have occurred in the genetic domain during the evolution of the different species.

The possible role of the GAUT gene family in fiber development

The formation of bast fiber is an important determinant ramie fiber quality, with the development of ramie fiber affecting the rate of hemp formation and ultimately the value of ramie directly (Chen et al. 2014). Analysis the cDNA sequence of ramie GalAT, a key pectin biosynthetic homologous to GAUT4, has known that most GAUT4 accumulates in roots, followed by leaves, phloem, and xylem (Liu et al. 2009). Similar to the situation in hemp, the development of cotton fiber cells in the main factor affecting cotton fiber quality (Haigler et al. 2012). *GbGAUT1*, a high galacturonic acid (HG) GAUT protein containing a conserved GAUT gene family domain, falls into group GAUT-A in the phylogenetic tree and is preferentially expressed during fiber secondary cell wall thickening, especially at 35 DPA. These results indicate that the *GbGAUT1* gene may play an important role in fiber development (Chi et al. 2009).

In our study, the peak expression of *GhGAUT* genes was concentrated before 15 DPA. In the six high-strength-fiber materials, the highest expression was from 10 to 15DPA, when the periods of fiber secondary wall thickening and fiber elongation overlap (Ji. 2011). In the low-strength lines, the expression of *GhGAUT* genes was concentrated between 5 to 10 DPA, which corresponds to the fiber elongation period (Fan. 2013).

Relationship between pectin substances and cotton fiber quality

As ia well known, that cotton fiber cells develop from a single cell, and their main components are cellulosic materials and pectin substances (Wang. 2012). Pectin synthesis and decomposition affect cotton fiber strength (Fan. 2013). Pectin methylesterase (PME), a common enzyme in plants that is related to cell wall structure, participates in pectin decomposition catalyzes pectin deesterification to produce pectate and methanol (Fan. 2013, Li et al. 2016). According to previous studies, PME levels increase during fiber development, with the lowest enzyme activity found in high-strength-fiber strains (Fan. 2013, Li. 2016). By comparing multiple transcriptome datasets, Guo et al. identified two genes related to pectin esterase, which is involved in the hydrolysis of pectin into gelatinic acid, and the expression of these genes were sharply upregulated starting at 12 DPA (Gou et al. 2007). In addition, two enzymes involved in pectin synthesis, UDP-glucose 6-dehydrogenase and UDP-D-glucuronic acid 4-epimerase, were down-regulated during secondary wall synthesis. The authors found that the amount and molecular weight of pectin was decreased in cells at the late stage of cotton fiber development (Gou et al. 2007). Research based on immunohistochemistry has revealed that unesterified homogalacturonan is sparse in epidermal cells, which do not develop into fibers, whereas this compound is abundant in elongated cotton fiber cells (Zhao et al. 2012). The above observations suggest that pectin synthesis

affects early fiber development and the hydrolysis of pectin is related to the formation of fibers during the later stage.

Conclusions

In this study, we characterized the GAUT galacturonosyltransferase gene family associated with pectin synthesis by analyzing their phylogenetic relationships, conserved motifs, gene structures, promoter sequences, and expression in cotton lines having different fiber strengths. Comprehensive expression and bioinformatics analysis indicated that the peak expression of *GhGAUT* genes was concentrated before 15DPA. Gene expression in the six materials with high-strength fiber were concentrated between 10 and 15DPA, the beginning of the fiber secondary wall thickening period and also part of the fiber elongation and thickening phase. In contrast, GAUT gene expression in the six materials with low-strength fiber was highest from 5 to 10 DPA, which was the fiber elongation period. The result should lay the foundation for future research associated with pectin synthesis during cotton fiber development.

Methods

Identification of cotton *GAUT* family members

To identify all homologous *GAUT* gene family in Arabidopsis (Sterling et al. 2006), we used a Hidden Markov Model approach as implemental in HMMER 3.0 (Finn et al. 2011) to search for the Glyco_transf_8 domain (PF01501) in the following species: *Carica papaya* L. (Tang et al. 2008; <http://www.phytozome.net>), *Theobroma cacao* (Argout et al. 2011), *G. raimondii* (Wang et al. 2012; <http://www.phytozome.net>), *Oryza sativa* (Du et al. 2017; <http://www.mbkbase.org/R498/>), *A. thaliana* (Cao et al. 2011; <http://www.arabidopsis.org>), *G. barbadense* (Liu et al. 2015b; <http://database.chgc.sh.cn/cotton/index.html>), *G. hirsutum* (Li et al. 2015; Zhang et al. 2015; <http://mascotton.njau.edu.cn>), and *G. arboreum* (Du et al. 2018; <ftp://bioinfo.ayit.edu.cn/downloads/>). The sequences were verified using BLASTp with *AtGAUT* sequences as the query sequences and an e-value threshold of 1×10^{-5} . After multiple sequence alignment in ClustalW program (Thompson et al. 2002), incomplete sequences was manually deleted.

Phylogenetic tree construction, analysis of gene structure and localization

A neighbor-joining phylogenetic tree was constructed from the aligned sequences using MEGA v6.06 with 1,000 bootstrap repeats. To confirm GAUT gene structure in *Gossypium* species, information about GAUT gene exons and introns was retrieved from the GFF3 file, and the exon/intron structure was visualized using the Gene Structure Display Server 2.0 (Hu et al. 2014). Conserved domains were predicted by using a conserved domain database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler et al. 2011). Motifs were explored using the MEME program (<http://meme.nbcr.net/meme>) (Bailey et al. 2006), with the maximum number of motifs set to 20. Analysis and visualization of the chromosomal distribution of the GAUT genes was carried out using MapChart 2.2 (Voorrips et al. 2002).

Promoter region and collinearity analysis of four cotton species

A 2,000-bp sequence upstream of the start codon in the upland cotton genomic sequence was extracted. Analysis of cis-acting elements was carried out using the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>; Lescot et al. 2002). Repetitive elements in the GAUT gene family were identified by collinear analysis using the entire BLAST array (e-value=1e-5) in the MCScan (Tang et al. 2008).

Plant materials

Plant were grown using standard field management practices in Anyang, China. Plant materials consisted of six high-fiber-strength lines and six low-fiber-strength lines from a RILs populations (Sun et al. 2012). The date of flowering was recorded as 0 days post-anthesis (0 DPA). Plant were sampled every 5 days from 0 to 30 DPA. After collection, fibers were separated from the cotton bolls with a sterile knife, immediately frozed in liquid nitrogen, and stored at -80 ° C, and fiber samples were dried at 45°C for determination of pectin content, RNA was extracted from cotton fiber tissue, and reversed transcribed into cDNA was formed by reverse-transcription PCR (RT-PCR). The resulting cDNA was stored at -20 ° C for subsequent use in qRT-PCR experiments with three biological replicates (Tuttle et al. 2015).

Transcriptome analyses and qRT-PCR

The transcriptome data were downloaded from the Sequence Read Archive (SRA) of the NCBI database (<https://www.ncbi.nlm.nih.gov/>) (Paterson et al. 2012). The SRA data for *G. raimondii* (PRJNA79005; Wang et al. 2012), *G. barbadense* (PRJNA251673; Liu et al. 2015b), *G. hirsutum* (PRJNA248163; Zhang et al. 2015), and *G. arboreum* (PRJNA179447; Du et al. 2018) were converted to FASTQ format data with the SRA Toolkit, and then analyzed and filtered with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). TopHat2 (Kim et al. 2013) was used to map the clean data to the index genomes constructed by Bowtie2 (Langmead et al. 2012) with the library-type and fr-unstranded parameters. The Cufflinks program was used to calculate the expression levels of the genes in the reference genome (Trapnell et al. 2010). Visualized gene expression levels (Fig S4: Table S2) were output using a homogenization method based on $\log_2(\text{FPKM} + 1)$ in the pheatmap (<https://CRAN.R-project.org/package=pheatmap>). qRT-PCR assays of selected genes were performed using specific designed primers (Fig S5: Table S3) on a Roche 480 II PCR system. Gene expression levels were calculated using the $2^{-\Delta\text{Ct}}$ method, and the experimental design included three biological replicates and three technical replicates (Livak et al. 2001; Pfaffl et al. 2001).

List Of Abbreviations

GAUT: galacturonosyltransferase

GAUTs: galacturonosyltransferases

PLN: phospholamban

DPA: days post-anthesis

HG: homogalacturonan

RG II: rhamnogalacturonan II

HG: GalAT: HG: α -1,4-D-Galacturonosyltransferase

GT8: glycosyltransferase 8

RIL: recombinant inbred lines

QTLs: quantitative trait loci

DEGs: differentially expressed genes

PGSIPs: protein-like starch starters

GolSs: galactitol synthase

FPKM: fragments per kilobase of transcript per million fragments mapped

Q-PCR: quantitative real time polymerase chain reaction

RT-PCR: reverse-transcription PCR

qRT-PCR: quantitative real-time PCR

SRA: Sequence Read Archive

ABA: abscisic acid

NAA: naphthylacetic acid

SA: salicylic acid

Me-JA: methyl jasmonate

GA: gibberellin

JA: jasmonate

IAA: indole acetic acid

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by the Major Research Plan of National Natural Science Foundation of China (NO.31690093); the National Agricultural Science and Technology Innovation project for CAAS (CAAS-ASTIP-2016-ICR), and the Central Level of the Scientific Research Institutes for Basic R & D Special Fund Business (Y2017PT51).The funder has no role in the design of the study and collection, analysis, or interpretation of data and writing of the manuscript.

Authors' contributions

Yuan YL and Shang HH conceived and managed the project. Fan SM coordinated the overall project and wrote the manuscript. Liu AY prepared plant materials and extracted RNA, and Zou XY completed the qRT-PCR assay. Zhang Z contributed to the multiple sequence alignments and the phylogenetic analysis. Ge Q helped analyze the data. Gong WK analyzed domains and predicted the promoters. Li JW and Gong JW wrote scripts. Shi YZ contributed to the chromosomal localization and gene structural analysis. Deng XY and Jia TT discussed the results and commented on the manuscript. All authors read, edited, and approved the current version of the manuscript.

Acknowledgments

Not applicable.

References

1. Argout X, Salse J, Aury JM, et al. The genome of *Theobroma cacao*. *Nature Genetics*, 2011, 43(2): 101.

2. Atmodjo MA, Sakuragi Y, Zhu X, et al. Galacturonosyltransferase (GAUT) 1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan: galacturonosyltransferase complex. *Proceedings of the National Academy of Sciences*, 2011, 108(50): 20225-20230.
3. Bailey TL, Williams N, Misleh C, Li WW: MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 2006, 34(suppl_2): W369-W373.
4. Basra AS, Malik CP. Development of the cotton fiber. *International Review of Cytology*, 1984, 89: 65-113.
5. Basyuni M, Wati R, Sulistiyono N, et al. Isolation and analysis of a multifunctional triterpene synthase KcMS promoter region from mangrove plant *kandelia candel*. *Conference Series: Earth and Environmental Science*, 2018, 130(1): 012013.
6. Caffall KH, Pattathil S, Phillips SE, et al. *Arabidopsis thaliana* T-DNA mutants implicate GAUT genes in the biosynthesis of pectin and xylan in cell walls and seed testa. *Molecular Plant*, 2009, 2(5): 1000-1014.
7. Campoy JA, Le Dantec L, Barreneche T, et al. New insights into fruit firmness and weight control in sweet cherry. *Plant Molecular Biology Reporter*, 2015, 33(4): 783-796.
8. Cantarel BL, Coutinho PM, Rancurel C, et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research*, 2008, 37(suppl_1): D233-D238.
9. Cao J, Schneeberger K, Ossowski S, et al: Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics*, 2011, 43:956.
10. Chen J, Pei Z, Dai L, et al. Transcriptome profiling using pyrosequencing shows genes associated with bast fiber development in ramie (*Boehmeria nivea* L.). *BMC Genomics*, 2014, 15(1): 919.
11. Chen J, Piao Y, Liu Y, et al. Genome-wide identification and expression analysis of chitinase gene family in *Brassica rapa* reveals its role in clubroot resistance. *Plant Science*, 2018, 270: 257-267.
12. Chi JN, Han GY, Wang XF, et al. Isolation and molecular characterization of a novel homogalacturonan galacturonosyl-transferase gene (*GbGAUT1*) from *Gossypium barbadense*. *African Journal of Biotechnology*, 2009, 8(19).
13. de Godoy F, Bermúdez L, Lira BS, et al. Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. *Journal of Experimental Botany*, 2013, 64(8): 2449-2466.
14. Durand C, Vicié-Gibouin M, Follet-Gueye M L, et al. The organization pattern of root border-like cells of *Arabidopsis* is dependent on cell wall homogalacturonan. *Plant Physiology*, 2009, 150(3): 1411-1421.
15. Du HL, Yu Y, Ma YF, et al. Sequencing and de novo assembly of a near complete indica rice genome. *Nature Communications*, 2017, 8: 15324.
16. Du XM, Huang G, He SP, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nature Genetics*, 2018, 50(6): 796.

17. Fan SM. The relationship between the cellulose synthesis and cotton fiber quality. Henan University, Kaifeng, 2013.
18. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 2011, 39(suppl_2): W29-W37.
19. Gou JY, Wang LJ, Chen SP, et al. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Research*, 2007, 17(5): 422.
20. Haigler CH, Betancur L, Stiff MR, et al. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Frontiers in Plant Science*, 2012, 3: 104.
21. Harholt J, Suttangkakul A, Scheller HV. Biosynthesis of pectin. *Plant Physiology*, 2010, 153(2): 384-395.
22. Herrera-Vásquez A, Salinas P, Holuigue L. Salicylic acid and reactive oxygen species interplay in the transcriptional control of defense genes expression. *Frontiers in Plant Science*, 2015, 6: 171.
23. Higo K, Ugawa Y, Iwamoto M, et al. Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Research*, 1999, 27(1): 297-300.
24. Hu B, Jin J, Guo AY, et al. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 2014, 31(8): 1296-1297.
25. Hyodo H, Terao A, Furukawa J, et al. Tissue specific localization of pectin–Ca²⁺ cross-linkages and pectin methyl-esterification during fruit ripening in tomato (*Solanum lycopersicum*). *PLoS One*, 2013, 8(11): e78949.
26. Ji ZB. Gene Expression Profile Analyses of Upland Cotton (*Gossypium hirsutum*) Fiber during Secondary Cell Wall Deposition. Huazhong Agricultural University University, Wuhan, 2011.
27. Kikuchi N, Kwon YD, Gotoh M, et al. Comparison of glycosyltransferase families using the profile hidden Markov model. *Biochemical and Biophysical Research Communications*, 2003, 310(2): 574-579.
28. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 2013, 14(4): R36.
29. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012, 9(4): 357.
30. Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 2002, 30(1): 325-327.
31. Li FG, Fan GY, Lu C, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology*, 2015, 33(5): 524.
32. Liu JX, Yu CM, Tang SW, et al. Cloning and expression of key enzyme gene GalAT in ramie pectin biosynthesis. *Agricultural Sciences in China*, 2009, 8(6): 664-670.
33. Li WJ, Shang HH, Ge Q, et al. Genome-wide identification, phylogeny, and expression analysis of pectin methylesterases reveal their major role in cotton fiber development. *BMC Genomics*, 2016,

- 17(1): 1000.
34. Li WJ. Phylogeny analysis of PME's gene family and function of *GhJAZ* in cotton. Henan University, Kaifeng, 2016.
  35. Liu X, Zhao B, Chen XY, et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports*, 2015, 5: 14139.
  36. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Methods, 2001, 25(4):402-408.
  37. Marchler-Bauer A, Lu S, Anderson J B, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, 2010, 39(suppl_1): D225-D229.
  38. Mishra S, Shukla A, Upadhyay S, et al. Identification, occurrence, and validation of DRE and ABRE Cis-regulatory motifs in the promoter regions of genes of *Arabidopsis thaliana*. *Journal of Integrative Plant Biology*, 2014, 56(4): 388-399.
  39. Mohnen D. Pectin structure and biosynthesis. *Current Opinion in Plant Biology*, 2008, 11(3): 266-277.
  40. Muñoz P, Norambuena L, Orellana A. Evidence for a UDP-glucose transporter in Golgi apparatus-derived vesicles from pea and its possible role in polysaccharide biosynthesis. *Plant Physiology*, 1996, 112(4): 1585-1594.
  41. Neckelmann G, Orellana A. Metabolism of uridine 5'-diphosphate-glucose in Golgi vesicles from pea stems. *Plant Physiology*, 1998, 117(3): 1007-1014.
  42. Orfila C, Sørensen SO, Harholt J, et al. QUASIMODO1 is expressed in vascular tissue of *Arabidopsis thaliana* inflorescence stems, and affects homogalacturonan and xylan biosynthesis. *Planta*, 2005, 222(4): 613-622.
  43. Paterson AH, Wendel JF, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, 2012, 492(7429): 423.
  44. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 2001, 29(9): e45-e45.
  45. Porto MS, Pinheiro MP, Batista VG, et al. Plant promoters: an approach of structure and function. *Molecular Biotechnology*, 2014, 56(1): 38-49.
  46. Reithmeier RA, Deber CM. Intrinsic membrane protein structure: principle and prediction. *The Structure of Biological Membranes*, 1992: 337-393.
  47. Ridley BL, O'Neill MA, Mohnen D. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, 2001, 57(6): 929-967.
  48. Scheller HV, Jensen JK, Sørensen SO, et al. Biosynthesis of pectin. *Physiologia Plantarum*, 2007, 129(2): 283-295.
  49. Sterling JD, Atmodjo MA, Inwood SE, et al. Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proceedings of the National Academy of Sciences*, 2006, 103(13): 5236-5241.

50. Sun FD, Zhang JH, Wang SF, et al. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. *Molecular Breeding*, 2012, 30(1):569-582.
51. Tang H, Bowers JE, Wang X, et al. Synteny and collinearity in plant genomes. *Science*, 2008, 320(5875): 486-488.
52. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using Clustal W and Clustal X. *Current Protocols in Bioinformatics*, 2003(1): 2-3.
53. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 2010, 28(5): 511.
54. Tuttle JR, Nah G, Duke MV, et al. Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. *BMC Genomics*, 2015, 16(1): 477.
55. Vasco M, De Stefano R, Di Matteo A, et al. Candidate genes controlling fruit quality in a tomato introgression line tolerant to water deficit. *Proceedings of the Joint Meeting AGI-SIBV-SIGA*. 2011: 19-22.
56. Voorrips RE: MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity*, 2002, 93(1):77-78.
57. Wang KB, Wang ZW, Li FG, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics*, 2012, 44(10): 1098.
58. Wang L, Wang W, Wang YQ, et al. Arabidopsis Galacturonosyltransferase (GAUT) 13 and GAUT14 have redundant functions in pollen tube growth. *Molecular Plant*, 2013, 6(4): 1131-1148.
59. Wang SG. Construction of Forward Subtractive Library and Analysis of Differentially Expressed Genes During Secondary Cell Wall Thickening Stage of Upland Cotton (*Gossypium hirsutum*) Fiber. Huazhong Agricultural University, Wuhan, 2012.
60. Wang YB, Li CH, Shang HH, et al. Study of Cotton Matured Fibre Quality and the Super-Molecular Structure in Upland Cotton RILs. *Fibres & Textiles in Eastern Europe*, 2014.
61. Wei LY, Zhang DF, Xiang F, et al. Differentially expressed miRNAs potentially involved in the regulation of defense mechanism to drought stress in maize seedlings. *International Journal of Plant Sciences*, 2009, 170(8): 979-989.
62. Willats WG, Knox JP, Mikkelsen JD. Pectin: new insights into an old polymer are starting to gel. *Trends in Food Science & Technology*, 2006, 17(3): 97-104.
63. Yin Y, Chen H, Hahn MG, et al. Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. *Plant Physiology*, 2010, 153(4): 1729-1746.
64. Yin Y, Mohnen D, Xu Y, et al. Glycosyltransferases of the GT8 family. *Annual Plant Reviews online*, 2018: 167-211.
65. Zhang TZ, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*, 2015, 33(5): 531.

66. Zhang Z, Ge Q, Liu AY, et al. Construction of a high-density genetic map and its application to QTL identification for fiber strength in Upland cotton. *Crop Science*, 2017, 57(2): 774-788.

67. Zhao L,Ding Q, Zeng J. et al. Immunohistochemical study of cell wall pectin during cotton fiber development. *Journal of Chinese Electron Microscopy Society*, 2012, 31(2): 179-183.

68. Zou XY, Liu AY, Zhang Z, et al. Co-Expression Network Analysis and Hub Gene Selection for High-Quality Fiber in Upland Cotton (*Gossypium hirsutum*) Using RNA Sequencing Analysis. *Genes*, 2019, 10(2): 119.

Figures

Figure 1

Phylogenetic tree of the Galacturonosyltransferase (GAUT) gene family. a. Neighbor-joining phylogenetic tree of Galacturonosyltransferase (GAUT) genes from eight species. b. Neighbor-joining phylogenetic tree of Galacturonosyltransferase (GAUT) family form the four cotton species.

Figure 2

The stucture analysis of Galacturonosyltransferase (GAUT) gene family. a: Phylogenetic tree of the Galacturonosyltransferase (GAUT) gene family. b: Analysis of the genomic sequences of the GAUT gene family domains. c: Analysis of the genomic sequences of the GAUT gene family motifs.

Figure 3

Collinearity of Galacturonosyltransferase genes in *Gossypium arboreum*, *Gossypium hirsutum* and *Gossypium raimondii*.

Figure 4

Heatmap presenting the Galacturonosyltransferase gene expression levels based on a transcriptome analysis. a: Galacturonosyltransferase gene expression levels during fiber development in *Gossypium arboreum* (0 to 15 DPA). b: Galacturonosyltransferase gene expression levels during fiber development in *Gossypium raimondii* (0 to 15 DPA). c: Galacturonosyltransferase gene expression levels during fiber development in *Gossypium arboreum* (10 to 28 DPA). d: Galacturonosyltransferase gene expression levels in the root, stem, leaf, and fiber tissues (5 to 30 DPA) from *Gossypium hirsutum*.

Figure 5

Analysis of GhGAUT expression levels in 6 high-strength fiber materials and 6 low-strength fiber materials. (a:0-153, b:69262, c:69272, d:69328, e:69327, f:69307, g: sGK9708, h:69312, i:69404, j:69305, k:69412, l:69362)

Figure 6

The pectin content of different samples. a: 5 high-strength fiber materials, b: 5 low-strength fiber materials

Figure 7

Analysis of cis-elements in GhGAUT genes

Figure 8

Analysis of GhGAUTs expression levels with abiotic stressand and phytohormone treatments according to a qRT-PCR assay. a:200 μ M NaCl treatment; b:20% PEG 6000 treatment; c: 100 μ M NAA treatment; d: 100 μ M ABA treatment; e: 100 μ M SA treatment; f: 100 μ M MeJA treatment.

Supplementary Files

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

- [FigS1.tif](#)

- [FigS2.png](#)
- [FigS3.tif](#)
- [FigS4.png](#)
- [FigS5.png](#)