Journal of Cotton Research

RESEARCH Open Access

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

FAN Senmiao[1], LIU Aiying[1], ZOU Xianyan[1], ZHANG Zhen[1], GE Qun[1], GONG Wankui[1], LI Junwen[1], GONG Juwu[1], SHI Yuzhen[1], DENG Xiaoying[1], JIA Tingting[1], YUAN Youlu[1,2*†] and SHANG Haihong[1,2*†]

## 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 have 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*, which coding probable galacturonosyltransferases. Among them, the GAUT genes encode proteins GAUT1 to GAUT15. All GAUT proteins except for GAUT7 contain a conserved glycosyl transferase family 8 domain (H-DN-A-SVV-S-V-H-T-F). The conserved sequence of GAUT7 is PLN (phospholamban) 02769 domain. According to *cis*-elemet analysis, GAUT genes transcript levels may be regulated by hormones such as JA, GA, SA, ABA, Me-JA, and IAA. The evolution and transcription patterns of the *GAUT* gene family in different cotton species and the transcript levels in upland cotton lines with different fiber strength were analyzed. Peak transcript level of *GhGAUT* genes have been observed before 15 DPA. In the six materials with high fiber strength, the transcription of *GhGAUT* genes were concentrated from 10 to 15 DPA; while the highest transcript levels in low fiber strength materials were detected between 5 and 10 DPA. These results 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.

**Keywords:** Cotton fiber development, Pectin, Galacturonosyltransferases, Evolution, Transcription patterns

* Correspondence: yuanyoulu@caas.cn; shanghaihong@caas.cn
†Yuan YL and Shang HH contributed equally to this work.
[1]State Key Laboratory of Cotton Biology; Key Laboratory of Biological and Genetic Breeding of Cotton, Ministry of Agriculture and Rural Affairs, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China
Full list of author information is available at the end of the article

## Background

Cotton, the world's most important economic fiber crop, is tightly connected to the ecomonic development and human livelihoods. Cotton fiber formation and development is an important factor in 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, and participates in the formation of plant tissue structure, biotic and abiotic stress processes. 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-coding genes are recognized on the basis of previous research in *Arabidopsis*: 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 researches have 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 (GT8s). GT8 consists of three separate protein classes, classes I and II contain mostly eukaryotic proteins, while almost the entire class III consists of bacterial proteins (Yin et al. 2010a, b). 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, GalAT and the GAUT1-related gene family provides the genetic and biochemical tools required to study the function of these genes in pectin synthesis (Sterling et al. 2006). GAUT1 also belongs to GT8, and the GAUT1-related superfamily also contains 10 GAUT-like genes (Cantarel et al. 2008). The *gaut1* mutation can cause plant dwarfing, reduce 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 it is considered to be a component of the HG:GalAT complex that participates in pectin synthesis (Atmodjo et al. 2011). The *gaut8* mutation 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 expressed 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. (2013) demonstrated that the *gaut1* gene had higher transcription level during fruit development, especially in the fruit epidermis, while *gaut1* showed higher expression level in the mesocarp, endocarp, septum, locular tissue, and nucleus after the fruit enters maturity.

To improve the quality of high fiber strength cotton materials, we have focused on cotton fiber development. Zhang et al. (2017) used recombinant inbred lines (RILs) to identify 16 stable quantitative trait loci (QTLs) related to fiber strength. In addition, Zou et al. (2019) identified 3 364 candidate genes related to cotton fiber strength, and analyzed the differences in fiber development (5 to 30 DPA) using two extremely different 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, respectively, were detected. Among them, 228 (62.6%) genes were upregulated in high-strength materials, while *Gh_A07G1907* (homologous to *GAUT6*) was downregulated at 15 DPA but upregulated at 25 DPA.

As discussed aboved, the GAUT gene family has been studied in plants such as *Arabidopsis* and tomato etc., 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 transcription 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-related family, including the GAUT and GATL proteins, known as galacturonosyltransferase proteins, and 2) a group including plant glycogen protein-like starch starters (PGSIPs) and galactitol synthases (GolSs) (Yin et al. 2010a, b). 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 species (Fig. S1), including 131 genes from the following species: *G. hirsutum* (41 genes), *G. barbadense* (42 genes), *G. arboreum* (25 genes) and *G. raimondii* (23 genes). The length 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 residues. The length of 32 GAUT genes was less than 3 000 bp; 88 genes had lengths of 3 000 to 7 000 bp, while the remaining 11 were longer than 7 000 bp (Table 1).

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

Using published genome sequence data from eight species, we determined the phylogenetic relationships of GAUT gene family members among multiple species. In a previous study (Sterling et al. 2006), three types of protein sequences, GAUT-A, GAUT-B, and GAUT-C, were found by multiple sequence alignment of the 187 GAUT proteins 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 proteins into GAUT-A, 38 into GAUT-B, and 31 into GAUT-C (Fig. 1b). In addition, *GhA07G1907* was a differentially expressed gene in the QTL which was related to fiber strength. It was a homologous gene to *AtGAUT6*. And genes homologous to *AtGAUT6* had the most number in *Gossypium* species (Zou et al. 2019). We found 16 *AtGAUT13* homologs, 14 genes were homologous to each of *AtGAUT7*, *AtGAUT9*, and *AtGAUT11*, and 11 genes were homologous to each of *AtGAUT2* and *AtGAUT12*. The remaining genes had fewer than 10 homologs. No gene homologous to *AtGAUT14* were detected in any of the four cotton species, and no *AtGAUT5* homolog was identified in *G. hirsutum*. Only one homolog of each of *AtGAUT5* and *AtGAUT10* were detected, namely, *GrGAUT18* and *GhGAUT03*, respectively. Four homologs each of *AtGAUT2*, *4, 6, 7, 9, 12*, and *13*, three homologs each of *AtGAUT8* and *AtGAUT11*, and two homologous genes to each of *AtGAUT1*, and *AtGAUT3* were discovered in *G. hirsutum*.

### Analysis of conserved motif and GAUT protein structures

The following motif is conserved in 15 *Arabidopsis* GAUT protein 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 encode proteins with a molecular mass between 61 and 78 kDa (Sterling et al. 2006; Godoy et al. 2013). Consistent with topological predictions, most GAUT proteins can encode a type II membrane protein containing a putative transmemberance domain in its hypervariable N-terminal region. Among the GAUT proteins analyzed in our study, three GAUT proteins (GAUT 3, 4, and 5) which belongeding to GAUT-A contained an N-terminal signal peptide rather than a transmemberance domain. The only GAUT gene family members predicted to have no N-terminal transmemberance domain or signal peptide in cotton were GAUT2 proteins (Fig. 2b). We also found some GAUT1, GAUT3 and GAUT11 proteins with the above-mentioned characteristics. Among 14 GAUT7 proteins belonging to GAUT-A group in four *Gossypium* species (Table 1), the proteins encoded by 10 genes homologous to GAUT7 contained the conserved PLN02769 domain (Fig. 2b), which was assigned to the category of "Probable Galacturonosyltransferase" (https://www.ncbi.nlm.nih.gov/proteinclusters/?term=PLN02769). The remaining GAUT family members contained a conserved glycosyl transferase family 8 domain. The predicted motifs of each member are shown in Fig. 2C, and the specific structural information is given in Fig. S2.

### Analysis of collinearity and repeating elements in the GAUT gene

According to the results of MCScan analysis, there was no tandem repeat element 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

**Table 1** Chromosomal locus ID and length of the encoded galacturonosyltransferase proteins in cotton

| cotton species | Gene model | ID | Chromosomes | Start | End | CDS length / bp | Protein length / aa | Arabidopsis ortholog | Arabidopsis protein description | classification |
|---|---|---|---|---|---|---|---|---|---|---|
| *G. arboreum* | *GaGAUT01* | *Ga01G0140* | Chr01 | 912 602 | 919 551 | 1 974 | 657 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GaGAUT02* | *Ga01G0796* | Chr01 | 11 759 252 | 11 761 952 | 1 605 | 534 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GaGAUT03* | *Ga03G0997* | Chr03 | 28 017 614 | 28 023 124 | 1 827 | 608 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GaGAUT04* | *Ga04G0446* | Chr04 | 6 075 632 | 6 079 138 | 1 602 | 533 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GaGAUT05* | *Ga04G1634* | Chr04 | 90 270 116 | 90 274 871 | 1 602 | 533 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GaGAUT06* | *Ga04G1718* | Chr04 | 92 087 105 | 92 095 131 | 1 719 | 572 | *AT3G58790.1* | GAUT15 | GAUT-C |
| | *GaGAUT07* | *Ga05G1886* | Chr05 | 17 017 554 | 17 030 587 | 2 328 | 775 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GaGAUT08* | *Ga06G1924* | Chr06 | 120 827 230 | 120 831 246 | 1 863 | 620 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GaGAUT09* | *Ga07G2432* | Chr07 | 93 674 605 | 93 678 413 | 1 800 | 599 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GaGAUT10* | *Ga07G1617* | Chr07 | 33 502 551 | 33 506 162 | 2 046 | 681 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GaGAUT11* | *Ga07G2025* | Chr07 | 79 624 753 | 79 627 031 | 1 674 | 557 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GaGAUT12* | *Ga07G1793* | Chr07 | 47 136 518 | 47 138 700 | 1 695 | 564 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GaGAUT13* | *Ga07G1005* | Chr07 | 13 869 804 | 13 874 208 | 1 695 | 564 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GaGAUT14* | *Ga08G2216* | Chr08 | 120 833 143 | 120 836 579 | 1 758 | 585 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GaGAUT15* | *Ga08G0724* | Chr08 | 11 007 501 | 11 030 719 | 1 803 | 600 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GaGAUT16* | *Ga08G1537* | Chr08 | 103 324 886 | 103 331 544 | 2 376 | 791 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GaGAUT17* | *Ga09G2452* | Chr09 | 81 931 389 | 81 936 357 | 1 830 | 609 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GaGAUT18* | *Ga09G0159* | Chr09 | 3 533 841 | 3 539 737 | 1 476 | 491 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GaGAUT19* | *Ga11G2709* | Chr11 | 109 497 004 | 109 501 318 | 1 797 | 598 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GaGAUT20* | *Ga11G3730* | Chr11 | 121 180 096 | 121 183 320 | 1 602 | 533 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GaGAUT21* | *Ga11G3192* | Chr11 | 116 488 843 | 116 492 318 | 2 031 | 676 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GaGAUT22* | *Ga11G3319* | Chr11 | 117 670 007 | 117 673 612 | 1 602 | 533 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GaGAUT23* | *Ga12G2518* | Chr12 | 93 835 048 | 93 837 447 | 1 683 | 560 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GaGAUT24* | *Ga13G0584* | Chr13 | 8 732 738 | 8 736 526 | 2 046 | 681 | *AT3G61130.1* | GAUT1 | GAUT-A |
| | *GaGAUT25* | *Ga13G1828* | Chr13 | 108 148 486 | 108 151 333 | 1 623 | 540 | *AT5G54690.1* | GAUT12 | GAUT-C |
| *G. barbadense* | *GbGAUT01* | *GOBAR_AA01456* | At_S0001 | 1 016 334 | 1 024 307 | 1 989 | 662 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GbGAUT02* | *GOBAR_AA00301* | At_S0001 | 9 613 451 | 9 616 141 | 1 605 | 534 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GbGAUT03* | *GOBAR_AA01219* | At_S0002 | 23 157 092 | 23 162 812 | 1 812 | 603 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GbGAUT04* | *GOBAR_AA02321* | At_S0004 | 74 825 334 | 74 826 761 | 1 428 | 475 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GbGAUT05* | *GOBAR_AA25346* | At_S0005 | 14 497 863 | 14 501 682 | 1 887 | 628 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GbGAUT06* | *GOBAR_AA13922* | At_S0005 | 93 596 407 | 93 601 574 | 1 752 | 583 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GbGAUT07* | *GOBAR_AA10943* | At_S0006 | 107 027 362 | 107 031 375 | 1 731 | 576 | *AT2G38650.1* | GAUT7 | GAUT-A |

**Table 1** Chromosomal locus ID and length of the encoded galacturonosyltransferase proteins in cotton *(Continued)*

| cotton species | Gene model | ID | Chromosomes | Start | End | CDS length / bp | Protein length / aa | Arabidopsis ortholog | Arabidopsis protein description | classification |
|---|---|---|---|---|---|---|---|---|---|---|
| | *GbGAUT08* | *GOBAR_AA06647* | At_S0007 | 89 670 406 | 89 674 210 | 1 800 | 599 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT09* | *GOBAR_AA01514* | At_S0007 | 75 555 669 | 75 557 947 | 1 674 | 557 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GbGAUT10* | *GOBAR_AA37671* | At_S0007 | 15 099 056 | 15 101 716 | 1 710 | 569 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GbGAUT11* | *GOBAR_AA14268* | At_S0007 | 46 927 015 | 46 929 189 | 1 590 | 529 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GbGAUT12* | *GOBAR_AA04077* | At_S0008 | 11 065 254 | 11 067 714 | 1 134 | 377 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT13* | *GOBAR_AA34869* | At_S0009 | 73 319 234 | 73 324 201 | 1 830 | 609 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GbGAUT14* | *GOBAR_AA07870* | At_S0009 | 2 874 750 | 2 880 650 | 1 539 | 512 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GbGAUT15* | *GOBAR_AA36945* | At_S0011 | 13 994 303 | 13 999 030 | 1 656 | 551 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT16* | *GOBAR_AA25184* | At_S0011 | 7 886 584 | 7 890 058 | 1 950 | 649 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GbGAUT17* | *GOBAR_AA35174* | At_S0011 | 6 702 894 | 6 713 442 | 4 899 | 1 632 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GbGAUT18* | *GOBAR_AA22791* | At_S0012 | 13 348 599 | 13 350 991 | 1 683 | 560 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GbGAUT19* | *GOBAR_AA03897* | At_S0013 | 14 455 836 | 14 460 722 | 2 064 | 687 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GbGAUT20* | *GOBAR_AA05453* | At_S0013 | 77 398 681 | 77 401 498 | 1 539 | 512 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GbGAUT21* | *GOBAR_AA14759* | At_S0386 | 114 937 | 118 369 | 1 497 | 498 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT22* | *GOBAR_DD05031* | Dt_S0001 | 933 831 | 940 348 | 2 067 | 688 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GbGAUT23* | *GOBAR_DD13138* | Dt_S0001 | 9 304 556 | 9 307 362 | 1 347 | 448 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GbGAUT24* | *GOBAR_DD33622* | Dt_S0004 | 44 183 238 | 44 186 775 | 1 602 | 533 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GbGAUT25* | *GOBAR_DD37379* | Dt_S0004 | 6 278 421 | 6 283 494 | 1 851 | 616 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GbGAUT26* | *GOBAR_DD20347* | Dt_S0004 | 6 480 662 | 6 485 936 | 1 758 | 585 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GbGAUT27* | *GOBAR_DD28302* | Dt_S0005 | 16 373 039 | 16 381 357 | 2 811 | 936 | *AT5G47780.1* | GAUT4 | GAUT-A |
| | *GbGAUT28* | *GOBAR_DD32865* | Dt_S0006 | 52 531 030 | 52 535 058 | 1 608 | 535 | *AT5G47780.1* | GAUT4 | GAUT-A |
| | *GbGAUT29* | *GOBAR_DD34963* | Dt_S0007 | 51 871 252 | 51 875 198 | 1 800 | 599 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT30* | *GOBAR_DD05509* | Dt_S0007 | 26 427 319 | 26 430 903 | 2 088 | 695 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GbGAUT31* | *GOBAR_DD33929* | Dt_S0007 | 43 934 795 | 43 936 661 | 1 476 | 491 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GbGAUT32* | *GOBAR_DD27569* | Dt_S0007 | 31 847 170 | 31 849 350 | 1 695 | 564 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GbGAUT33* | *GOBAR_DD14647* | Dt_S0007 | 13 131 930 | 13 134 606 | 1 710 | 569 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GbGAUT34* | *GOBAR_DD04306* | Dt_S0008 | 44 255 956 | 44 259 633 | 2 019 | 672 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GbGAUT35* | *GOBAR_DD07609* | Dt_S0008 | 8 776 221 | 8 778 456 | 1 134 | 377 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT36* | *GOBAR_DD34943* | Dt_S0008 | 56 586 461 | 56 590 364 | 1 743 | 580 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT37* | *GOBAR_DD05442* | Dt_S0011 | 11 746 553 | 11 750 934 | 1 761 | 586 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GbGAUT38* | *GOBAR_DD22761* | Dt_S0011 | 2 807 344 | 2 808 777 | 1 434 | 477 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GbGAUT39* | *GOBAR_DD24858* | Dt_S0011 | 7 360 685 | 7 363 733 | 1 485 | 494 | *AT2G46480.1* | GAUT2 | GAUT-A |

**Table 1** Chromosomal locus ID and length of the encoded galacturonosyltransferase proteins in cotton *(Continued)*

| cotton species | Gene model | ID | Chromosomes | Start | End | CDS length / bp | Protein length / aa | Arabidopsis ortholog | Arabidopsis protein description | classification |
|---|---|---|---|---|---|---|---|---|---|---|
| | *GbGAUT40* | *GOBAR_DD38263* | Dt_S0011 | 6 083 354 | 6 086 957 | 1 536 | 511 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GbGAUT41* | *GOBAR_DD34814* | Dt_S0012 | 8 771 758 | 8 774 141 | 1 683 | 560 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GbGAUT42* | *GOBAR_DD11145* | Dt_S0013 | 46 541 744 | 46 544 946 | 1 272 | 423 | *AT5G54690.1* | GAUT12 | GAUT-C |
| *G. hirsutum* | *GhGAUT01* | *Gh_A01G0568* | A01 | 9 996 974 | 9 999 674 | 1 605 | 534 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GhGAUT02* | *Gh_A02G0896* | A02 | 32 207 939 | 32 213 457 | 1 824 | 607 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GhGAUT03* | *Gh_A04G0911* | A04 | 57 533 849 | 57 537 350 | 1 602 | 533 | *AT2G20810.1* | GAUT10 | GAUT-B |
| | *GhGAUT04* | *Gh_A05G3203* | A05 | 83 801 743 | 83 806 499 | 1 602 | 533 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GhGAUT05* | *Gh_A05G3268* | A05 | 85 630 826 | 85 638 852 | 1 719 | 572 | *AT3G58790.1* | GAUT15 | GAUT-C |
| | *GhGAUT06* | *Gh_A05G1520* | A05 | 15 501 936 | 15 516 418 | 2 805 | 934 | *AT5G47780.1* | GAUT4 | GAUT-A |
| | *GhGAUT07* | *Gh_A06G1385* | A06 | 95 717 993 | 95 722 006 | 1 863 | 620 | *AT5G47780.1* | GAUT4 | GAUT-A |
| | *GhGAUT08* | *Gh_A07G1907* | A07 | 74 675 761 | 74 679 567 | 1 800 | 599 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GhGAUT09* | *Gh_A07G1602* | A07 | 62 859 121 | 62 861 400 | 1 674 | 557 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GhGAUT10* | *Gh_A07G1480* | A07 | 44 247 043 | 44 249 325 | 1 776 | 591 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GhGAUT11* | *Gh_A07G0842* | A07 | 14 511 134 | 14 513 794 | 1 710 | 569 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GhGAUT12* | *Gh_A07G1344* | A07 | 33 278 890 | 33 282 504 | 2 046 | 681 | *AT3G61130.1* | GAUT1 | GAUT-A |
| | *GhGAUT13* | *Gh_A08G1146* | A08 | 80 396 030 | 80 399 699 | 2 019 | 672 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GhGAUT14* | *Gh_A11G1169* | A11 | 14 257 384 | 14 261 704 | 1 797 | 598 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GhGAUT15* | *Gh_A11G0310* | A11 | 2 866 274 | 2 869 499 | 1 602 | 533 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GhGAUT16* | *Gh_A11G0648* | A11 | 6 318 816 | 6 322 420 | 1 602 | 533 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GhGAUT17* | *Gh_A13G0573* | A13 | 13 490 010 | 13 493 804 | 2 046 | 681 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GhGAUT18* | *Gh_A13G1199* | A13 | 65 040 755 | 65 043 553 | 1 605 | 534 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GhGAUT19* | *Gh_D01G0577* | D01 | 7 904 964 | 7 913 279 | 1 605 | 534 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GhGAUT20* | *Gh_D02G1116* | D02 | 31 936 862 | 31 942 590 | 1 827 | 608 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GhGAUT21* | *Gh_D04G1424* | D04 | 45 645 539 | 45 649 073 | 1 602 | 533 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GhGAUT22* | *Gh_D04G0402* | D04 | 6 351 785 | 6 357 064 | 1 674 | 557 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GhGAUT23* | *Gh_D04G0340* | D04 | 5 179 695 | 5 186 785 | 1 389 | 462 | *AT3G58790.1* | GAUT15 | GAUT-C |
| | *GhGAUT24* | *Gh_D05G1691* | D05 | 15 200 861 | 15 204 140 | 1 842 | 613 | *AT5G47780.1* | GAUT4 | GAUT-A |
| | *GhGAUT25* | *Gh_D06G1729* | D06 | 56 469 689 | 56 473 720 | 1 863 | 620 | *AT5G47780.1* | GAUT4 | GAUT-A |
| | *GhGAUT26* | *Gh_D07G2130* | D07 | 51 568 886 | 51 572 831 | 1 800 | 599 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GhGAUT27* | *Gh_D07G1799* | D07 | 42 869 876 | 42 872 172 | 1 674 | 557 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GhGAUT28* | *Gh_D07G0913* | D07 | 11 891 038 | 11 893 713 | 1 710 | 569 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GhGAUT29* | *Gh_D07G1583* | D07 | 30 120 469 | 30 122 649 | 1 695 | 564 | *AT3G25140.1* | GAUT8 | GAUT-B |

**Table 1** Chromosomal locus ID and length of the encoded galacturonosyltransferase proteins in cotton *(Continued)*

| cotton species | Gene model | ID | Chromosomes | Start | End | CDS length / bp | Protein length / aa | Arabidopsis ortholog | Arabidopsis protein description | classification |
|---|---|---|---|---|---|---|---|---|---|---|
| | *GhGAUT30* | *Gh_D07G1453* | D07 | 24 699 109 | 24 702 694 | 2 031 | 676 | *AT3G61130.1* | GAUT1 | GAUT-A |
| | *GhGAUT31* | *Gh_D08G1429* | D08 | 46 983 870 | 46 990 546 | 2 235 | 744 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GhGAUT32* | *Gh_D09G2061* | D09 | 47 942 831 | 47 947 845 | 1 818 | 605 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GhGAUT33* | *Gh_D11G1324* | D11 | 12 768 969 | 12 772 925 | 1 794 | 597 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GhGAUT34* | *Gh_D11G0365* | D11 | 3 115 363 | 3 118 680 | 1 602 | 533 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GhGAUT35* | *Gh_D11G0880* | D11 | 7 605 876 | 7 610 151 | 2 031 | 676 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GhGAUT36* | *Gh_D11G0760* | D11 | 6 527 861 | 6 531 464 | 1 602 | 533 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GhGAUT37* | *Gh_D12G0542* | D12 | 9 891 906 | 9 894 305 | 1 614 | 537 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GhGAUT38* | *Gh_D13G0555* | D13 | 7 521 289 | 7 525 063 | 2 046 | 681 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GhGAUT39* | *Gh_D13G1495* | D13 | 46 583 202 | 46 585 967 | 1 599 | 532 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GhGAUT40* | *Gh_A09G2268* | scaffold2280_A09 | 33 334 | 38 396 | 1 821 | 606 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GhGAUT41* | *Gh_A11G3052* | scaffold2738_A11 | 31 739 | 35 214 | 2 031 | 676 | *AT2G46480.1* | GAUT2 | GAUT-A |
| G. raimondii | *GrGAUT01* | *Gorai.001G104800* | Chr01 | 11 739 260 | 11 742 665 | 1 713 | 570 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GrGAUT02* | *Gorai.001G172800* | Chr01 | 24 959 043 | 24 963 513 | 2 034 | 677 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GrGAUT03* | *Gorai.001G188500* | Chr01 | 30 394 865 | 30 398 560 | 1 779 | 592 | *AT3G25140.1* | GAUT8 | GAUT-B |
| | *GrGAUT04* | *Gorai.001G205700* | Chr01 | 40 061 128 | 40 063 974 | 1 677 | 558 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GrGAUT05* | *Gorai.001G243400* | Chr01 | 48 406 329 | 48 410 709 | 1 803 | 600 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GrGAUT06* | *Gorai.002G013600* | Chr02 | 881 738 | 888 088 | 1 854 | 617 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GrGAUT07* | *Gorai.002G082100* | Chr02 | 10 361 510 | 10 364 854 | 1 608 | 535 | *AT5G54690.1* | GAUT12 | GAUT-C |
| | *GrGAUT08* | *Gorai.004G074600* | chr04 | 8 330 249 | 8 333 832 | 1 098 | 365 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GrGAUT09* | *Gorai.004G155600* | Chr04 | 44 036 606 | 44 041 193 | 2 022 | 673 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GrGAUT10* | *Gorai.004G219100* | Chr04 | 55 255 869 | 55 260 007 | 1 746 | 581 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GrGAUT11* | *Gorai.005G129100* | Chr05 | 31 303 031 | 31 309 906 | 1 830 | 609 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GrGAUT12* | *Gorai.006G236100* | Chr06 | 48 377 701 | 48 383 955 | 1 833 | 610 | *AT2G38650.1* | GAUT7 | GAUT-A |
| | *GrGAUT13* | *Gorai.007G040000* | Chr07 | 2 785 971 | 2 790 164 | 1 605 | 534 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GrGAUT14* | *Gorai.007G081500* | Chr07 | 5 820 168 | 5 825 079 | 1 605 | 534 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GrGAUT15* | *Gorai.007G093500* | Chr07 | 6 832 789 | 6 837 085 | 2 034 | 677 | *AT2G46480.1* | GAUT2 | GAUT-A |
| | *GrGAUT16* | *Gorai.007G142400* | Chr07 | 11 853 941 | 11 858 955 | 1 797 | 598 | *AT1G06780.1* | GAUT6 | GAUT-A |
| | *GrGAUT17* | *Gorai.008G059800* | Chr08 | 9 388 299 | 9 391 237 | 1 686 | 561 | *AT3G02350.1* | GAUT9 | GAUT-B |
| | *GrGAUT18* | *Gorai.010G190800* | Chr10 | 54 465 897 | 54 470 520 | 1 866 | 621 | *AT2G30575.1* | GAUT5 | GAUT-A |
| | *GrGAUT19* | *Gorai.012G042500* | Chr12 | 5 361 960 | 5 369 902 | 1 611 | 536 | *AT3G58790.1* | GAUT15 | GAUT-C |

**Table 1** Chromosomal locus ID and length of the encoded galacturonosyltransferase proteins in cotton *(Continued)*

| cotton species | Gene model | ID | Chromosomes | Start | End | CDS length / bp | Protein length / aa | Arabidopsis ortholog | Arabidopsis protein description | classification |
|---|---|---|---|---|---|---|---|---|---|---|
| | *GrGAUT20* | *Gorai.012G049900* | Chr12 | 6 493 884 | 6 500 082 | 1 677 | 558 | *AT3G01040.1* | GAUT13 | GAUT-C |
| | *GrGAUT21* | *Gorai.012G131400* | Chr12 | 30 042 007 | 30 046 372 | 1 605 | 534 | *AT1G18580.1* | GAUT11 | GAUT-B |
| | *GrGAUT22* | *Gorai.013G063700* | Chr13 | 7 098 252 | 7 102 442 | 2 049 | 682 | *AT4G38270.1* | GAUT3 | GAUT-A |
| | *GrGAUT23* | *Gorai.013G164000* | Chr13 | 44 518 254 | 44 522 187 | 1 605 | 534 | *AT5G54690.1* | GAUT12 | GAUT-C |

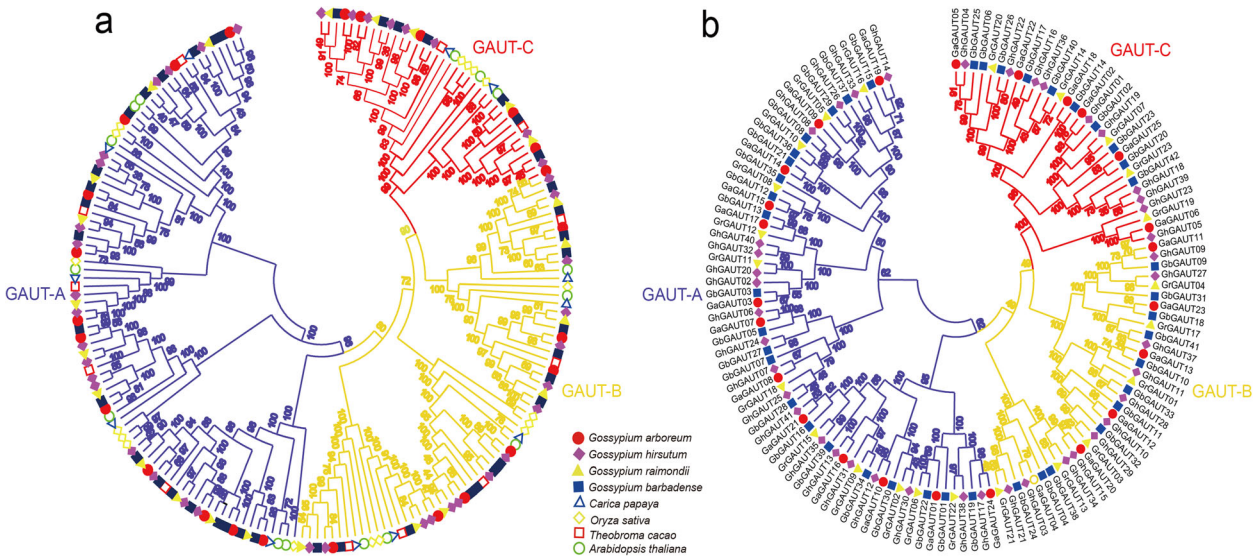


**Fig. 1** Phylogenetic tree of the galacturonosyltransferase (GAUT) gene family**a.** Neighbor-joining phylogenetic tree of galacturonosyltransferase (GAUT) genes from eight species. **b:** Neighbor-joining phylogenetic tree of galacturonosyltransferase (GAUT) family form the four cotton species. They comparised three types of protein sequences by multiple sequence alignment of the 187 GAUT proteins for eight species. From the four cotton species, GAUT gene family were classified into GAUT-A (62 proteins), GAUT-B (38 proteins), and GAUT-C (31 proteins)

*AtGAUT3*, *AtGAUT11*, *AtGAUT12*, and *AtGAUT13*, respectively, and were also segment repeats. Genes in diploid *Gossypium* species corresponding to the above repeated genes were shown in Fig. 3. According to their relative order in *Gossypium*, GAUT genes were categorized into five groups (1 to 5). Only group 4 belongs to *GAUT-A*, all other groups were members of *GAUT-C*.

### Analysis of *GAUTs* transcription 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 *Gossypium* species, *GaGAUT02* had the highest transcript level at 15 DPA but almost no transcription during other periods. Other members of group 1 (Fig. 3) were also barely transcribed 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 transcript level during the late stage (20 to 25 DPA) of fiber development (Fig. 4). Group 3 members *GhGAUT16*, *GhGAUT36*, *GrGAUT14*, and *GaGAUT22* were not transcribed at all during the fiber developmental period. Among group 4 genes, *GhGAUT17* and *GhGAUT38* had their highest transcript level during fiber developmental from 5 to 10 DPA, and *GrGAUT22* and *GaGAUT24* had the peak transcript level at 0 and 15 DPA, respectively. All members of group 5 except for *GaGAUT25* were transcribed at 15 DPA, and the remaining genes had almost no transcription during fiber development. In four cotton species, six genes, namely *GaGAUT08*, *GaGAUT12*, *GaGAUT13*, *GrGAUT03*, *GrGAUT18*, and *GhGAUT25*, had peak FPKM values greater than 40. The expression peak of these six genes all occurred before 15 DPA, which suggested that the GAUT gene family play a important 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* which belonged to *GAUT-A*, and *GhGAUT10*, *GhGAUT11* and *GhGAUT29* which belonged to *GAUT-B*. And all had FPKM values greater than 10 (Tables 2,3,4, and 5). The qRT-PCR analysis revealed that the GAUT genes had an important influence on fiber development before 15 DPA. From 5 to 30 DPA, the overall transcript levels of *GhGAUT08* and *GhGAUT10* were higher in high fiber-strength materials than those in low strength materials. At 5 DPA, the transcript levels of *GhGAUT11* and *GhGAUT29* were higher in low fiber-strength materials than those in high strength materials, with the opposite trend from 10 to 30 DPA. The transcript level of *GhGAUT25* was higher in low fiber-strength materials than high strength materials from 5 to 10 DPA, with the reverse pattern after 15 DPA. In six high strength materials, peak GAUT expression was from 10 to 15 DPA (Fig. 5),

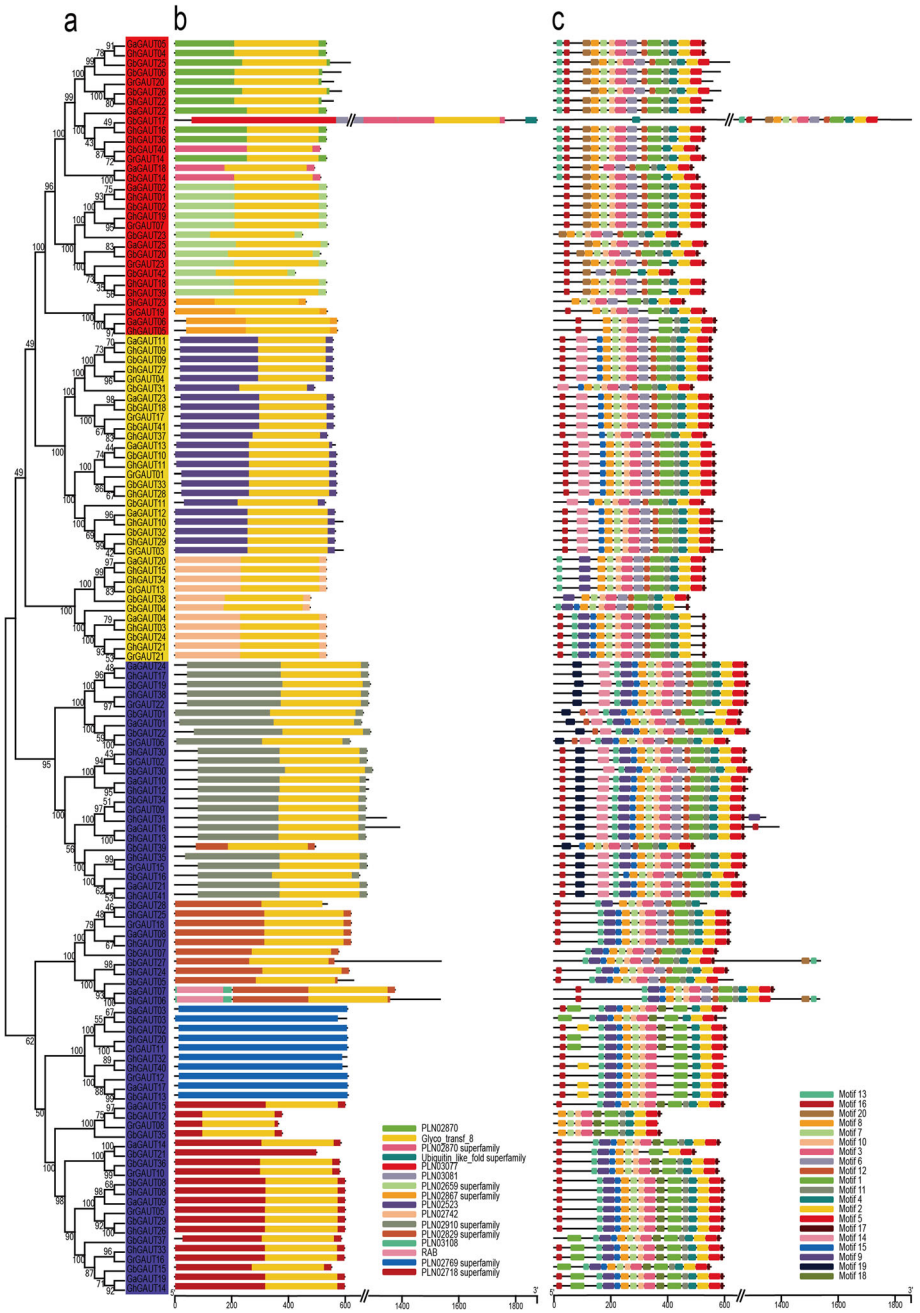


**Fig. 2** The stucture analysis of galacturonosyltransferase (GAUT) gene family. **a:** Phylogenetic tree of the galacturonosyltransferase (GAUT) gene family. **b:** Analysis of the genomic sequences of the GAUT gene family domains. **c:** Analysis of the genomic sequences of the GAUT gene family motifs

whereas the period of highest expression in six low-strength materials was 5 to 10 DPA (Fig. 5).

Since the GAUT gene family affect 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 15 DPA, while the low strength materials appeared at 10 DPA (Fig. 6). This result was similar to the pattern of gene transcription.

## Analysis of the *cis*-elements in the promoter regions GAUT genes

To investigate the potential reasons for different expression patterns among GAUT genes, we analyzed the promoter region of the GAUT gene family in upland cotton. This analysis was performed because *cis*-elements (Fig. 7) can affect gene expression (Higo et al. 1999). The 32 *cis*-elements were detected in the

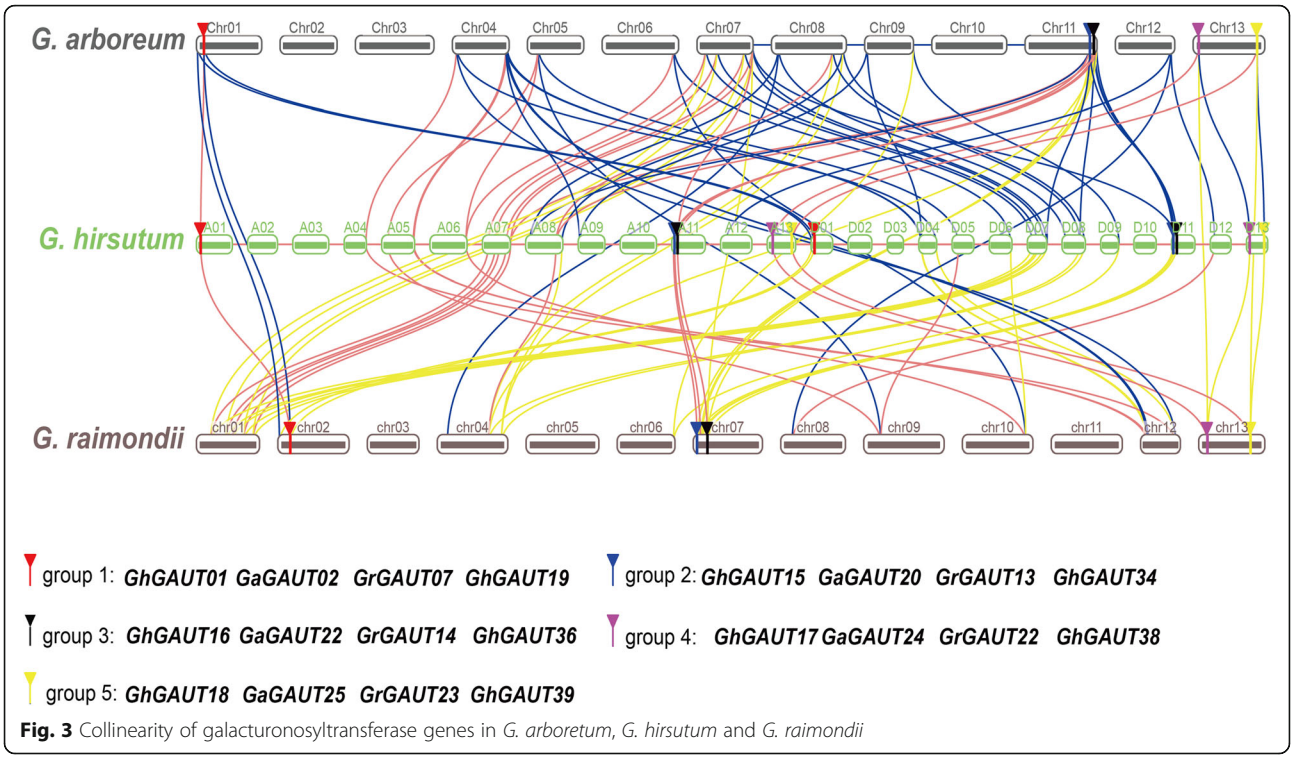


**Fig. 3** Collinearity of galacturonosyltransferase genes in *G. arboretum*, *G. hirsutum* and *G. raimondii*

promoter region of the 41 GAUT genes in upland cotton. The 32 *cis*-elements contained motifs that belonged to the regulatory sequences that responsive to anaerobicity, abiotic stress and hormone signaling. CGTCA and TGACG elements 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 (CCATCTTTTT) is a salicylic acid (SA) responsive element (Herrera-Vásquez et al. 2015). AuxRR-core (GGTCCAT) and the TGA-element (AACGAC) are related to the responsiveness to auxin (Herrera-Vásquez 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).

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

For a more in-depth study of *GhGAUTs* transcript levels induced by abiotic stress, the transcription patterns of five *GhGAUT* genes in cotton with NaCl, PEG, abscisic acid (ABA), naphthylacetic acid (NAA), salicylic acid (SA), and methyl jasmonate (MeJA) treatments were analyzed by qRT-PCR (Fig. 8). We examined the effects of various hormones on the transcription of the five *GhGAUT* genes. We observed that within 1 h after all treatments, the relative transcript levels of these five genes were rapidly increased and then decreased after 24 h. The peak transcription levels of the up-regulated gene came between 3 h and 12 h, except for *GhGAUT11*. *GhGAUT11* did not respond to the treatment of three hormones ABA, SA, and MeJA. *GhGAUT29* responded to all stresses and hormone treatments, and the detected transcript levels were higher. *GhGAUT08* responded to ABA and SA treatments with the higher transcript levels. *GhGAUT25* also responded to two stresses and four hormone treatments, but the transcript levels were the highest under the treatment of PEG and ABA, and the response peaks came at 6 h and 12 h after treatments, respectively. *GhGAUT10* also responded to the treatments of ABA, SA, and MeJA, with the peak transcript levels at 6 h, 12 h, and 12 h; the response levels to NACl, PEG and NAA treatment was low at 3 h and 6 h, respectively. 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

Researches on the GAUT gene family currently focused on *Arabidopsis* (Sterling et al. 2006; Cantarel et al.

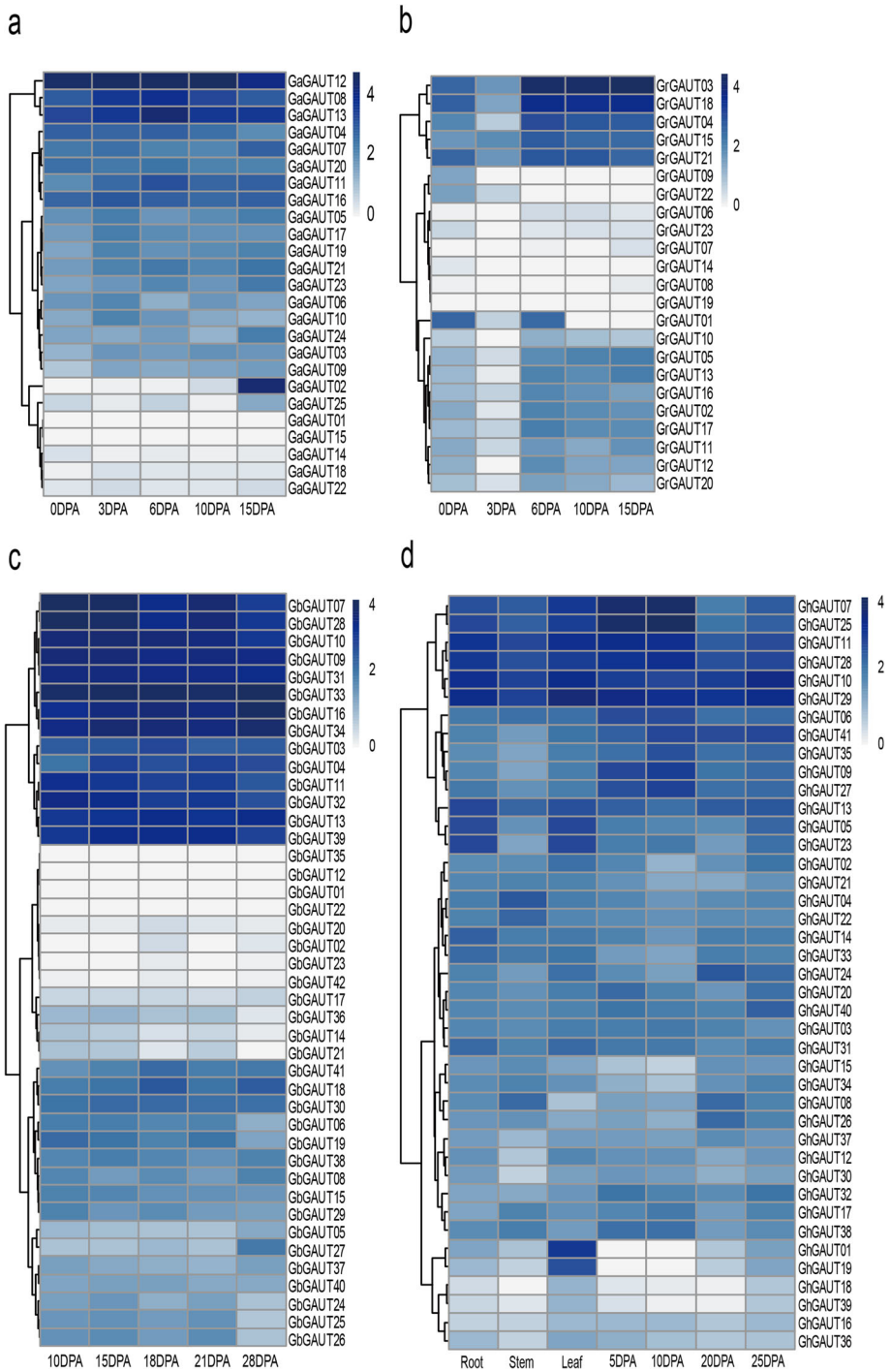


**Fig. 4** Heatmap of the galacturonosyltransferase gene transcript levels based on transcriptome analysis. **a**: galacturonosyltransferase gene transcript levels during fiber development in *G. arboreum* (0 to 15 DPA). **b**: Galacturonosyltransferase gene transcript levels during fiber development in *G. raimondii* (0 to 15 DPA). **c**: Galacturonosyltransferase gene transcript levels during fiber development in *G. arboreum* (10 to 28 DPA). **d**: Galacturonosyltransferase gene transcript levels in the root, stem, leaf, and fiber tissues (5 to 30 DPA) from *G. hirsutum*

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 systematically evolutionary analysis has not yet been conducted. Among 131 GAUT members in cotton, 20 are homologous to *AtGAUT6*; no gene in cotton is homologous to *AtGAUT14*, and no *AtGAUT5* homolog was present in *G. hirsutum*.

**Table 2** The FPKM value of GAUT genes in *G. arboreum*

| Gene model | ID | 0 DPA | 3 DPA | 6 DPA | 10 DPA | 15 DPA |
|---|---|---|---|---|---|---|
| *Ga01G0140* | *GaGAUT01* | 0.04 | 0.00 | 0.02 | 0.05 | 0.03 |
| *Ga01G0796* | *GaGAUT02* | 0.04 | 0.15 | 0.19 | 0.65 | 79.57 |
| *Ga03G0997* | *GaGAUT03* | 2.65 | 5.51 | 4.86 | 6.06 | 5.33 |
| *Ga04G0446* | *GaGAUT04* | 17.91 | 16.45 | 17.27 | 13.14 | 6.62 |
| *Ga04G1634* | *GaGAUT05* | 6.43 | 9.64 | 5.77 | 6.85 | 9.92 |
| *Ga04G1718* | *GaGAUT06* | 5.40 | 8.04 | 3.06 | 5.33 | 4.01 |
| *Ga05G1886* | *GaGAUT07* | 11.15 | 13.59 | 8.43 | 7.89 | 17.68 |
| *Ga06G1924* | *GaGAUT08* | 20.44 | 39.23 | 49.70 | 28.73 | 19.21 |
| *Ga07G2432* | *GaGAUT09* | 1.47 | 3.84 | 3.56 | 4.07 | 5.00 |
| *Ga07G1617* | *GaGAUT10* | 3.65 | 8.38 | 5.67 | 3.41 | 2.82 |
| *Ga07G2025* | *GaGAUT11* | 7.05 | 14.31 | 23.29 | 16.95 | 17.30 |
| *Ga07G1793* | *GaGAUT12* | 97.22 | 104.53 | ##### | 122.70 | 59.66 |
| *Ga07G1005* | *GaGAUT13* | 29.80 | 39.13 | 75.55 | 40.56 | 39.00 |
| *Ga08G2216* | *GaGAUT14* | 0.60 | 0.15 | 0.31 | 0.12 | 0.31 |
| *Ga08G0724* | *GaGAUT15* | 0.00 | 0.00 | 0.03 | 0.00 | 0.03 |
| *Ga08G1537* | *GaGAUT16* | 16.38 | 21.55 | 18.07 | 15.19 | 17.47 |
| *Ga09G2452* | *GaGAUT17* | 4.84 | 9.73 | 6.85 | 6.38 | 6.36 |
| *Ga09G0159* | *GaGAUT18* | 0.17 | 0.48 | 0.34 | 0.46 | 0.38 |
| *Ga11G2709* | *GaGAUT19* | 4.08 | 8.46 | 6.26 | 6.86 | 8.21 |
| *Ga11G3730* | *GaGAUT20* | 13.06 | 10.67 | 12.06 | 8.77 | 9.09 |
| *Ga11G3192* | *GaGAUT21* | 4.77 | 8.87 | 10.25 | 7.52 | 11.48 |
| *Ga11G3319* | *GaGAUT22* | 0.42 | 0.76 | 0.46 | 0.36 | 0.76 |
| *Ga12G2518* | *GaGAUT23* | 4.01 | 5.25 | 7.88 | 5.31 | 10.94 |
| *Ga13G0584* | *GaGAUT24* | 3.92 | 3.59 | 4.68 | 2.84 | 9.19 |
| *Ga13G1828* | *GaGAUT25* | 0.84 | 0.26 | 1.16 | 0.16 | 3.59 |

The GAUT gene family encodes galacturonosyltransferases involved in the synthesis of pectin, a compound related to cell wall formation. The conserved domain of the GAUT protein is H-DN-A-SVV-S-V-H-T-F. In tomato, GAUT1 protein had been predicted to be a membrane protein with a single N-terminal transmembrane helix and a major globular domain in the Golgi cavity, however, there was no report about the function of pectin synthesis and its role in fiber development. Analyses of the GAUT gene family have shown that all members share a conserved glycosyltransferase family 8 domain (Godoy et al. 2013; Vasco et al. 2011). In addition, GAUT proteins in the four species of cotton have a PLN02769 domain. 14 *GAUT* genes are homologous to *GAUT7*, two of which are present in *G.raimondii*, and 4 in other three species.

### The possible role of the GAUT gene family in fiber development

The formation of bast fiber is an important quality trait in ramie. The development of ramie fiber affecting the rate of hemp formation and ultimately the economic value of ramie directly (Chen et al. 2014). Analysis of the cDNA sequence of ramie GalAT, a key pectin biosynthetic homolog of GAUT4, has showed that most *GAUT4* transcript 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 is the main factor affecting cotton fiber quality (Haigler et al. 2012). GbGAUT1, a high galacturonic acid (HG) GAUT protein contains a conserved glycosyl transferase family 8 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 six high-strength-fiber materials, the highest expression was from 10 to 15 DPA, when the periods of fiber secondary wall thickening and fiber elongation overlap (Ji 2011). In the low-

**Table 3** The FPKM value of GAUT genes in *G. raimondii*

| Gene model | ID | 0 DPA | 3 DPA | 6 DPA | 10 DPA | 15 DPA |
|---|---|---|---|---|---|---|
| *Gorai.001G104800* | *GrGAUT01* | 13.13 | 0.94 | 11.94 | 0.00 | 0.00 |
| *Gorai.001G172800* | *GrGAUT02* | 2.95 | 0.41 | 7.31 | 6.03 | 5.36 |
| *Gorai.001G188500* | *GrGAUT03* | 13.30 | 4.76 | 72.76 | 70.32 | 87.30 |
| *Gorai.001G205700* | *GrGAUT04* | 6.77 | 1.11 | 20.09 | 17.12 | 15.52 |
| *Gorai.001G243400* | *GrGAUT05* | 2.35 | 0.60 | 5.99 | 7.27 | 8.03 |
| *Gorai.002G013600* | *GrGAUT06* | 0.16 | 0.00 | 0.57 | 0.59 | 0.36 |
| *Gorai.002G082100* | *GrGAUT07* | 0.00 | 0.00 | 0.19 | 0.00 | 0.53 |
| *Gorai.004G074600* | *GrGAUT08* | 0.13 | 0.00 | 0.09 | 0.04 | 0.20 |
| *Gorai.004G155600* | *GrGAUT09* | 3.27 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Gorai.004G219100* | *GrGAUT10* | 1.12 | 0.07 | 2.82 | 1.91 | 1.35 |
| *Gorai.005G129100* | *GrGAUT11* | 3.01 | 0.81 | 4.85 | 2.88 | 5.34 |
| *Gorai.006G236100* | *GrGAUT12* | 2.61 | 0.00 | 5.66 | 3.57 | 3.26 |
| *Gorai.007G040000* | *GrGAUT13* | 2.20 | 0.27 | 7.31 | 6.00 | 7.97 |
| *Gorai.007G081500* | *GrGAUT14* | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Gorai.007G093500* | *GrGAUT15* | 4.58 | 5.83 | 16.21 | 12.30 | 11.55 |
| *Gorai.007G142400* | *GrGAUT16* | 2.17 | 0.88 | 6.24 | 5.33 | 3.80 |
| *Gorai.008G059800* | *GrGAUT17* | 2.08 | 0.91 | 8.11 | 5.76 | 5.78 |
| *Gorai.010G190800* | *GrGAUT18* | 14.82 | 3.20 | 41.49 | 37.46 | 40.83 |
| *Gorai.012G042500* | *GrGAUT19* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Gorai.012G049900* | *GrGAUT20* | 1.93 | 0.56 | 3.75 | 3.08 | 2.07 |
| *Gorai.012G131400* | *GrGAUT21* | 12.74 | 4.86 | 18.23 | 16.99 | 13.68 |
| *Gorai.013G063700* | *GrGAUT22* | 3.62 | 0.99 | 0.00 | 0.00 | 0.00 |
| *Gorai.013G164000* | *GrGAUT23* | 0.74 | 0.00 | 0.33 | 0.48 | 0.45 |

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 is well known, 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 and catalyzes pectin deesterification to produce pectate and methanol (Fan 2013; Li et al. 2016). According to previous studies, PME levels increased during fiber development, with the lowest enzyme activity found in high-strength-fiber cultivars (Fan 2013; Li 2016). By comparing multiple transcriptome datasets, Guo et al. (2007) identified two genes related to pectin esterase, which is involved in the hydrolysis of pectin to gelatinic acid, and the expression of these genes were sharply upregulated starting at 12 DPA. 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 anthors found that the amount 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-mentioned 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 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 15 DPA. Gene expression in the six materials with

**Table 4** The FPKM value of GAUT genes in *G. barbadense*

| Gene model | ID | 10 DPA | 15 DPA | 18 DPA | 21 DPA | 28 DPA |
|---|---|---|---|---|---|---|
| *GOBAR_AA01456* | *GbGAUT01* | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| *GOBAR_AA00301* | *GbGAUT02* | 0.04 | 0.01 | 0.50 | 0.04 | 0.28 |
| *GOBAR_AA01219* | *GbGAUT03* | 10.48 | 10.76 | 14.62 | 9.76 | 10.01 |
| *GOBAR_AA02321* | *GbGAUT04* | 6.93 | 15.84 | 12.14 | 14.85 | 12.65 |
| *GOBAR_AA25346* | *GbGAUT05* | 1.54 | 1.43 | 1.19 | 1.21 | 2.19 |
| *GOBAR_AA13922* | *GbGAUT06* | 5.59 | 5.50 | 4.78 | 4.91 | 1.95 |
| *GOBAR_AA10943* | *GbGAUT07* | 45.27 | 41.40 | 22.30 | 31.44 | 17.28 |
| *GOBAR_AA06647* | *GbGAUT08* | 4.44 | 3.20 | 4.10 | 3.13 | 4.90 |
| *GOBAR_AA01514* | *GbGAUT09* | 30.41 | 28.78 | 25.56 | 26.60 | 25.26 |
| *GOBAR_AA37671* | *GbGAUT10* | 29.24 | 31.37 | 28.77 | 27.95 | 17.48 |
| *GOBAR_AA14268* | *GbGAUT11* | 20.94 | 18.81 | 15.18 | 17.15 | 11.50 |
| *GOBAR_AA04077* | *GbGAUT12* | 0.02 | 0.01 | 0.02 | 0.01 | 0.00 |
| *GOBAR_AA34869* | *GbGAUT13* | 17.90 | 19.76 | 23.95 | 19.18 | 22.65 |
| *GOBAR_AA07870* | *GbGAUT14* | 1.31 | 0.85 | 0.46 | 0.67 | 0.18 |
| *GOBAR_AA36945* | *GbGAUT15* | 5.22 | 4.46 | 3.95 | 3.97 | 3.52 |
| *GOBAR_AA25184* | *GbGAUT16* | 21.04 | 26.67 | 29.70 | 26.49 | 41.24 |
| *GOBAR_AA35174* | *GbGAUT17* | 0.62 | 0.64 | 0.70 | 0.54 | 0.81 |
| *GOBAR_AA22791* | *GbGAUT18* | 5.62 | 6.42 | 11.06 | 6.51 | 10.14 |
| *GOBAR_AA03897* | *GbGAUT19* | 7.78 | 6.91 | 5.00 | 6.68 | 2.60 |
| *GOBAR_AA05453* | *GbGAUT20* | 0.26 | 0.18 | 0.48 | 0.27 | 0.22 |
| *GOBAR_AA14759* | *GbGAUT21* | 1.23 | 1.04 | 0.33 | 0.90 | 0.01 |
| *GOBAR_DD05031* | *GbGAUT22* | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| *GOBAR_DD13138* | *GbGAUT23* | 0.07 | 0.04 | 0.26 | 0.06 | 0.17 |
| *GOBAR_DD33622* | *GbGAUT24* | 2.93 | 3.29 | 1.99 | 2.91 | 1.30 |
| *GOBAR_DD37379* | *GbGAUT25* | 3.39 | 3.86 | 3.01 | 3.67 | 1.15 |
| *GOBAR_DD20347* | *GbGAUT26* | 3.66 | 4.18 | 3.24 | 4.01 | 1.26 |
| *GOBAR_DD28302* | *GbGAUT27* | 1.32 | 1.17 | 1.54 | 1.16 | 5.85 |
| *GOBAR_DD32865* | *GbGAUT28* | 42.92 | 37.70 | 23.37 | 29.94 | 18.16 |
| *GOBAR_DD34963* | *GbGAUT29* | 5.21 | 3.51 | 4.02 | 3.19 | 2.97 |
| *GOBAR_DD05509* | *GbGAUT30* | 6.76 | 7.91 | 7.69 | 7.43 | 7.44 |
| *GOBAR_DD33929* | *GbGAUT31* | 26.33 | 27.23 | 25.11 | 24.79 | 23.75 |
| *GOBAR_DD27569* | *GbGAUT32* | 25.99 | 22.78 | 16.21 | 18.16 | 12.44 |
| *GOBAR_DD14647* | *GbGAUT33* | 37.89 | 41.23 | 39.75 | 36.68 | 45.19 |
| *GOBAR_DD04306* | *GbGAUT34* | 24.68 | 32.42 | 28.71 | 26.97 | 35.18 |
| *GOBAR_DD07609* | *GbGAUT35* | 0.05 | 0.00 | 0.01 | 0.02 | 0.00 |
| *GOBAR_DD34943* | *GbGAUT36* | 1.69 | 1.75 | 1.29 | 1.46 | 0.29 |
| *GOBAR_DD05442* | *GbGAUT37* | 2.76 | 2.21 | 2.04 | 1.86 | 2.93 |
| *GOBAR_DD22761* | *GbGAUT38* | 4.72 | 5.52 | 4.74 | 3.81 | 4.89 |
| *GOBAR_DD24858* | *GbGAUT39* | 17.74 | 23.75 | 22.63 | 23.55 | 15.60 |
| *GOBAR_DD38263* | *GbGAUT40* | 2.53 | 2.71 | 2.83 | 2.34 | 2.29 |
| *GOBAR_DD34814* | *GbGAUT41* | 3.99 | 5.13 | 8.02 | 5.42 | 5.95 |
| *GOBAR_DD11145* | *GbGAUT42* | 0.16 | 0.03 | 0.24 | 0.04 | 0.09 |

**Table 5** The FPKM value of GAUT genes in *G. hirsutum*

| Gene model | ID | root | stem | leaf | 5 DPA fiber | 10 DPA fiber | 20 DPA fiber | 25 DPA fiber |
|---|---|---|---|---|---|---|---|---|
| *Gh_A01G0568* | *GhGAUT01* | 2.69 | 1.36 | 18.75 | 0.07 | 0.02 | 1.06 | 2.84 |
| *Gh_A02G0896* | *GhGAUT02* | 4.50 | 4.39 | 7.51 | 5.01 | 1.96 | 4.07 | 6.83 |
| *Gh_A04G0911* | *GhGAUT03* | 4.95 | 4.39 | 5.79 | 6.02 | 6.15 | 5.31 | 4.12 |
| *Gh_A05G3203* | *GhGAUT04* | 5.91 | 12.21 | 5.92 | 5.05 | 3.80 | 4.57 | 4.90 |
| *Gh_A05G3268* | *GhGAUT05* | 13.72 | 3.94 | 15.45 | 6.10 | 4.66 | 4.36 | 9.24 |
| *Gh_A05G1520* | *GhGAUT06* | 6.67 | 7.43 | 7.86 | 14.36 | 14.49 | 7.63 | 8.40 |
| *Gh_A06G1385* | *GhGAUT07* | 12.62 | 10.45 | 18.91 | 37.20 | 41.63 | 5.77 | 10.38 |
| *Gh_A07G1907* | *GhGAUT08* | 4.47 | 8.20 | 1.24 | 3.04 | 2.59 | 8.15 | 5.55 |
| *Gh_A07G1602* | *GhGAUT09* | 4.67 | 2.65 | 5.92 | 14.99 | 17.92 | 7.29 | 8.58 |
| *Gh_A07G1480* | *GhGAUT10* | 22.20 | 16.95 | 23.86 | 17.97 | 15.28 | 17.72 | 26.25 |
| *Gh_A07G0842* | *GhGAUT11* | 20.23 | 14.72 | 23.33 | 24.01 | 23.32 | 11.06 | 13.40 |
| *Gh_A07G1344* | *GhGAUT12* | 3.98 | 1.08 | 4.70 | 4.15 | 4.10 | 2.43 | 3.75 |
| *Gh_A08G1146* | *GhGAUT13* | 15.55 | 9.08 | 12.80 | 9.81 | 7.43 | 10.44 | 12.05 |
| *Gh_A11G1169* | *GhGAUT14* | 10.10 | 5.59 | 6.22 | 5.34 | 3.61 | 6.04 | 5.68 |
| *Gh_A11G0310* | *GhGAUT15* | 3.80 | 4.30 | 2.65 | 1.33 | 0.78 | 3.85 | 3.78 |
| *Gh_A11G0648* | *GhGAUT16* | 0.84 | 0.79 | 1.68 | 1.59 | 1.71 | 1.21 | 1.61 |
| *Gh_A13G0573* | *GhGAUT17* | 2.71 | 5.42 | 4.07 | 5.08 | 6.14 | 3.43 | 5.11 |
| *Gh_A13G1199* | *GhGAUT18* | 0.55 | 0.06 | 1.83 | 0.32 | 0.24 | 0.17 | 1.15 |
| *Gh_D01G0577* | *GhGAUT19* | 1.80 | 0.79 | 13.13 | 0.04 | 0.02 | 0.96 | 2.56 |
| *Gh_D02G1116* | *GhGAUT20* | 4.95 | 3.94 | 6.05 | 8.15 | 5.38 | 3.73 | 7.89 |
| *Gh_D04G1424* | *GhGAUT21* | 5.00 | 5.29 | 5.20 | 4.14 | 2.46 | 2.38 | 3.89 |
| *Gh_D04G0402* | *GhGAUT22* | 5.31 | 9.34 | 4.77 | 4.25 | 4.04 | 4.31 | 4.52 |
| *Gh_D04G0340* | *GhGAUT23* | 15.34 | 2.59 | 14.94 | 5.79 | 6.26 | 3.44 | 7.55 |
| *Gh_D05G1691* | *GhGAUT24* | 5.12 | 3.32 | 7.47 | 4.25 | 2.83 | 11.93 | 8.28 |
| *Gh_D06G1729* | *GhGAUT25* | 14.69 | 10.26 | 18.32 | 41.09 | 48.94 | 7.24 | 9.88 |
| *Gh_D07G2130* | *GhGAUT26* | 3.78 | 3.86 | 2.44 | 2.93 | 2.05 | 8.21 | 5.23 |
| *Gh_D07G1799* | *GhGAUT27* | 6.53 | 3.87 | 5.70 | 13.20 | 16.98 | 7.38 | 8.29 |
| *Gh_D07G0913* | *GhGAUT28* | 19.85 | 12.38 | 17.22 | 20.97 | 22.64 | 13.23 | 14.94 |
| *Gh_D07G1583* | *GhGAUT29* | 24.72 | 16.46 | 31.64 | 27.44 | 22.77 | 20.91 | 24.89 |
| *Gh_D07G1453* | *GhGAUT30* | 3.22 | 0.79 | 3.07 | 3.75 | 3.27 | 2.25 | 3.05 |
| *Gh_D08G1429* | *GhGAUT31* | 8.65 | 5.42 | 8.31 | 6.68 | 6.18 | 4.61 | 5.79 |
| *Gh_D09G2061* | *GhGAUT32* | 2.59 | 2.34 | 3.55 | 7.04 | 5.11 | 4.61 | 7.16 |
| *Gh_D11G1324* | *GhGAUT33* | 8.25 | 6.34 | 6.75 | 3.42 | 2.60 | 5.80 | 5.32 |
| *Gh_D11G0365* | *GhGAUT34* | 4.49 | 5.13 | 3.89 | 2.06 | 1.26 | 3.99 | 5.13 |
| *Gh_D11G0880* | *GhGAUT35* | 4.60 | 2.63 | 5.80 | 7.97 | 12.90 | 7.35 | 9.40 |
| *Gh_D11G0760* | *GhGAUT36* | 1.77 | 0.85 | 2.79 | 2.11 | 1.45 | 1.11 | 1.39 |
| *Gh_D12G0542* | *GhGAUT37* | 3.29 | 1.62 | 3.31 | 3.31 | 2.95 | 4.44 | 3.81 |
| *Gh_D13G0555* | *GhGAUT38* | 4.33 | 5.66 | 3.37 | 7.61 | 7.61 | 3.31 | 4.24 |
| *Gh_D13G1495* | *GhGAUT39* | 0.67 | 0.32 | 1.90 | 0.41 | 0.11 | 0.17 | 1.12 |
| *Gh_A09G2268* | *GhGAUT40* | 4.82 | 4.45 | 5.48 | 7.20 | 5.97 | 5.25 | 10.16 |
| *Gh_A11G3052* | *GhGAUT41* | 5.37 | 3.17 | 6.77 | 9.65 | 15.14 | 13.75 | 14.55 |

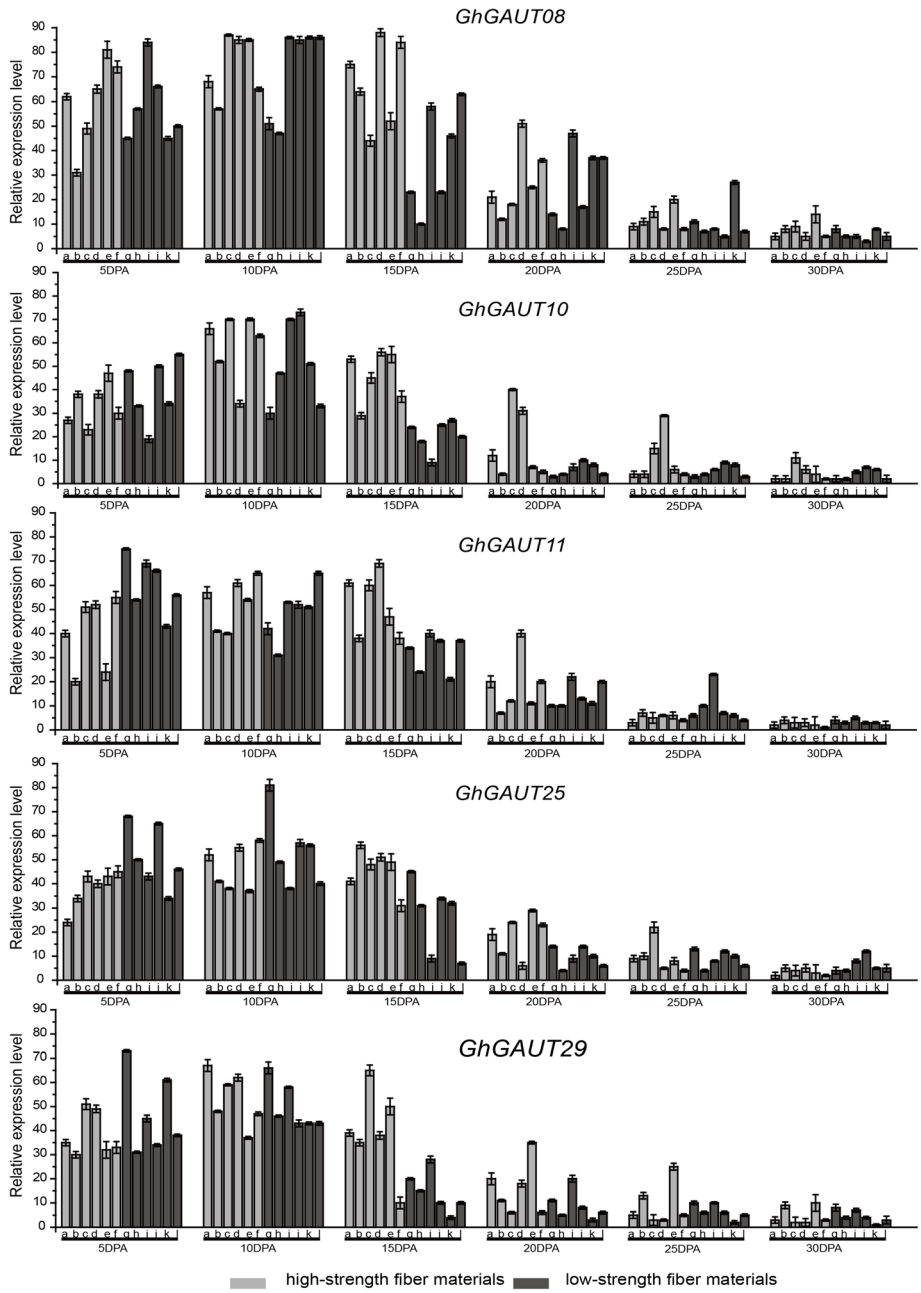


**Fig. 5** Analysis of 5 *GhGAUT* transcript levels in 6 high-strength fiber materials and 6 low-strength fiber materials. (a:0–153, b:69262, c:69272, d:69328, e:69327, f:69307, g: sGK9708, h:69312, i:69404, j:69305, k:69412, l:69362)

high-strength fiber were concentrated between 10 and 15 DPA, corresponding to the beginning of the fiber secondary wall thickening period and part of the fiber elongation and thickening phase. In contrast, GAUT gene expression in six materials with low-strength fiber was at the highest level from 5 to 10 DPA, which was the fiber elongation period. The result lays the foundation for future researches 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 integrated in HMMER 3.0 (Finn et al. 2011) to search for the glycosyl transferase family 83 (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)

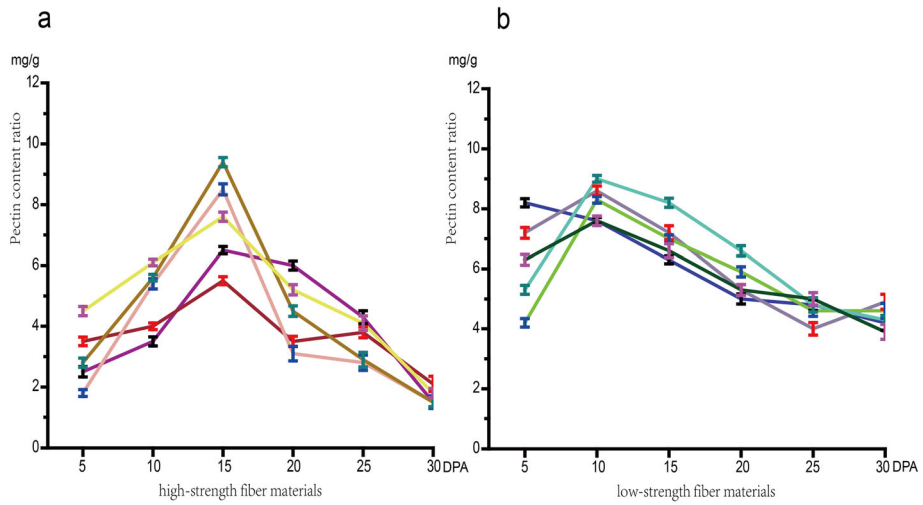


**Fig. 6** The pectin content of different samples. **a**: 5 high-strength fiber materials; **b**: 5 low-strength fiber materials

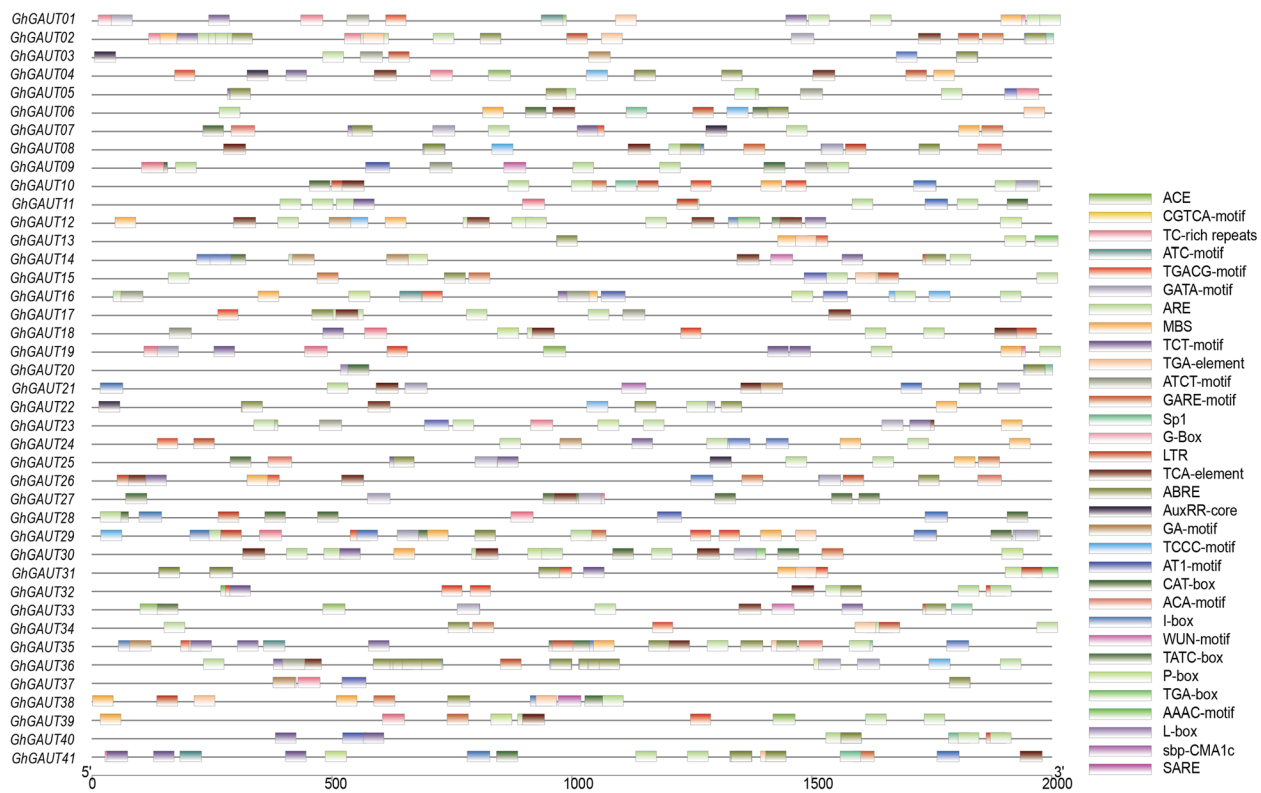


**Fig. 7** Analysis of *cis*-elements in *GhGAUT* genes. 32 *cis*-elements in upland cotton genes were analyzed, namely, elements responsive to anaerobic conditions, different hormones (Me-JA, GA, SA, ABA and IAA), and abiotic stress conditions, including drought, low temperature and other stress conditions

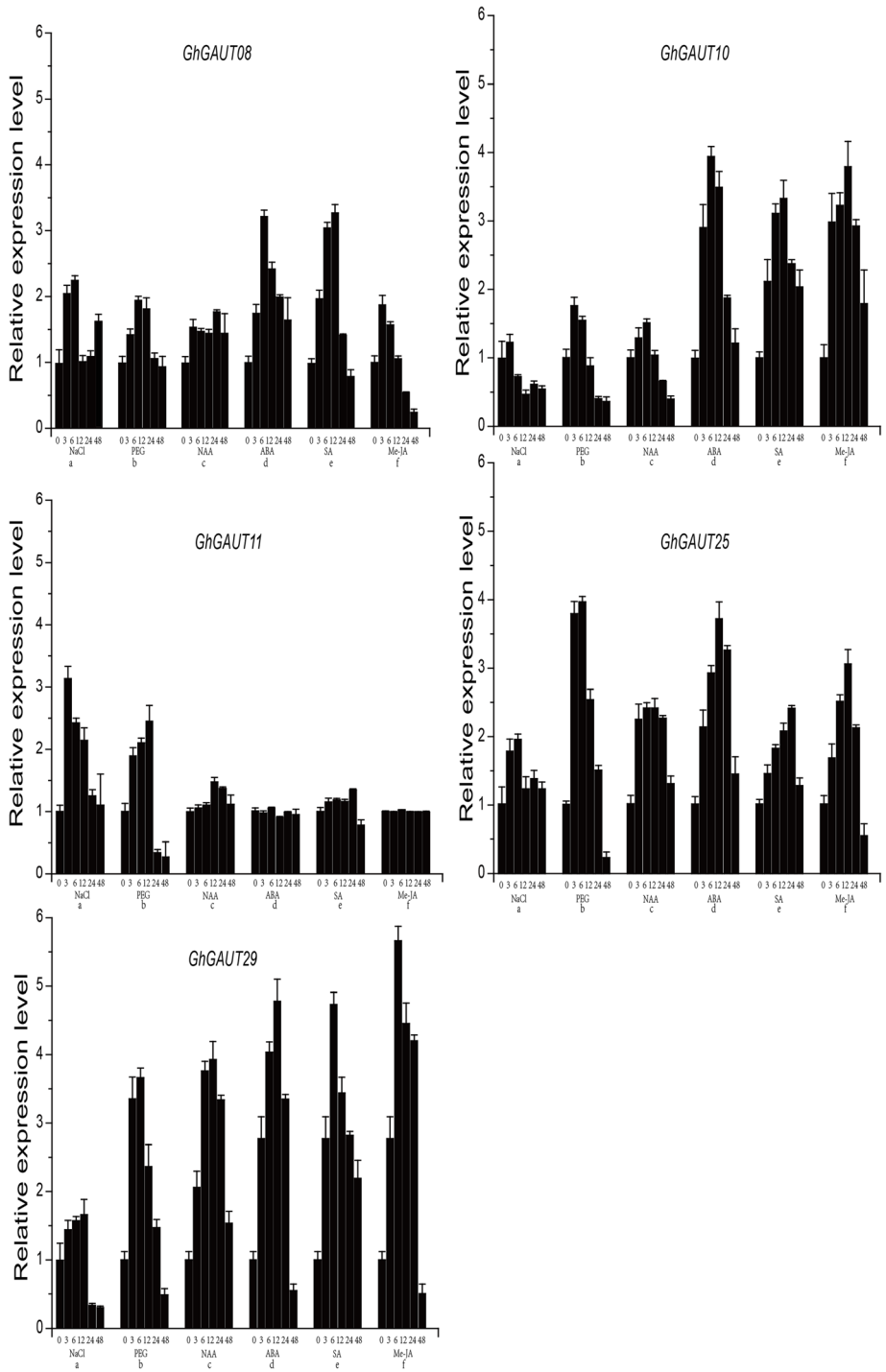


**Fig. 8** Analysis of *GhGAUT*s transcript levels with abiotic stress and phytohormone treatments according to a qRT-PCR assay. **a**: 200 μmol·L$^{-1}$ NaCl treatment; **b**: 20% PEG 6000 treatment; **c**: 100 μmol·L$^{-1}$ NAA treatment; **d**: 100 μmol·L$^{-1}$ ABA treatment; **e**: 100 μmol·L$^{-1}$ SA treatment; **f**: 100 μmol·L$^{-1}$ MeJA treatment

(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. 2015) (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

**Table 6** Details of primers for quantitative real-time PCR

| Gene model | ID | Forward primer | Reverse primer |
|---|---|---|---|
| *Gh_A07G1907* | *GhGAUT08* | CTGCGGTTCTATCTGCCAGACG | TCTTTGCCAAAAACGGGTCCGA |
| *Gh_A07G1480* | *GhGAUT10* | CTTGCTGCATCCGTAGTGGTGA | ACCTGCATTGCCCCAAGATTCA |
| *Gh_A07G0842* | *GhGAUT11* | TGGTGCACTTCAACGGGAACAT | TGCAGGCCTGTACAAACTCCAG |
| *Gh_D06G1729* | *GhGAUT25* | TAGGGTCGGCTCCTTCGTCAAT | CAATGGGTTCTTTCAAGGCGGC |
| *Gh_D07G1583* | *GhGAUT29* | AAGGAGGTCCGTTTTGGCACTC | CTTTCGAGCGTAGGAGGCGTAG |

with *AtGAUT* sequences as the query sequences and an e-value threshold of $1 \times 10^{-5}$. After multiple sequence alignment in ClustalW program (Thompson et al. 2003), incomplete sequences were 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 the exons and introns of GAUT genes was retrieved from the GFF3 file, and the exon/intron structures were visualized using the Gene Structure Display Server 2.0 (Hu et al. 2015). Conserved domains were predicted using a conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler et al. 2010). 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 were carried out using MapChart 2.2 (Voorrips 2002).

### Promoter region and collinearity analysis of four cotton species

The promoter region of 41 *GhGAUT* genes were analyzed. Analysis of *cis*-acting elements was carried out using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) (Lescot et al. 2002), The 32 *cis*-elements were detected in the promoter region of the 41 GAUT genes in upland cotton. The 32 *cis*-elements contain motifs that have characteristic of regulatory sequences in response to anaerobicity, abiotic stress and hormone signaling. Repetitive elements in the GAUT gene family were identified by collinear analysis using the entire BLAST array (e-value = $1 \times 10^{-5}$) in the MCScan (Tang et al. 2008).

### Plant materials

Plants were grown using standard field management practices in Anyang, China. One hundred and ninty-six RIL populations were constructed with 0-153 and 9708 as parents, and 5 lines with high fiber strength and 5 lines with low fiber strength were screened. 0-153 were a high-value parents in fiber characters and 9708 were a low-value parents. Plant materials consisted 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). Cotton bolls were sampled every 5 days from 0 to 30 DPA. After collection, fibers were separated from the cotton bolls with a sterile knife, immediately froze in liquid nitrogen, and stored at 80 °C, another part of fiber samples were dried at 45 °C for determination of pectin content. RNA was extracted from cotton fiber tissues, reverse-transcribed into cDNA. The cDNA was stored at −20 °C for subsequent 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. 2015), *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 and Salzberg 2012) with the library-type and fr-unstranded parameters. The Cufflinks program was used to calculate the transcript levels of the genes in the reference genome (Trapnell et al. 2010). Gene transcript levels (Tables 2,3,4, and 5) were visualized using a normalization method based on $\log_2$ (FPKM + 1) in the pheatmap (https://CRAN.R-project.org/package=pheatmap). qRT-PCR assays of selected genes were performed using specific designed primers (Table 6) on a Roche 480 II PCR system. Gene transcript levels were calculated using the $2^{-\Delta\Delta C}{}_t$ method, and the experimental design included three biological replicates and three technical replicates (Livak and Schmittgen 2001; Pfaffl 2001).

#### Abbreviations

GAUT: Galacturonosyltransferase; GAUTs: Galacturonosyltransferases; PLN: Phospholamban; DPA: Days post-anthesis; HG: Homogalacturonan; RG II: Rhamnogalacturonan II; HG:GalAT: HGα-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

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42397-021-00099-z.

**Additional file 1 Fig. S1**: The protein sequences of 187 GAUT genes from eight species: *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium raimondii*, *Arabidopsis thaliana*, *Oryza sativa*, *Carica papaya*, and *Theobroma cacao*.

**Additional file 2 Fig. S2:** Logos of 20 motifs for four cotton species according to the MEME suite.

### Acknowledgments

Not applicable.

### 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 final version of the manuscript.

### 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.

### Availability of data and materials

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

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no conflict of interest.

### Author details

[1]State Key Laboratory of Cotton Biology; Key Laboratory of Biological and Genetic Breeding of Cotton, Ministry of Agriculture and Rural Affairs, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China. [2]School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450000, China.

Received: 24 January 2021 Accepted: 30 June 2021
Published online: 02 August 2021

### References

Argout X, Salse J, Aury JM, et al. The genome of *Theobroma cacao*. Nat Genet. 2011;43(2):101–8. https://doi.org/10.1038/ng.736.

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. Proc Natl Acad Sci. 2011;108(50):20225–30. https://doi.org/10.1073/pnas.1112816108.

Bailey TL, Williams N, Misleh C, et al. MEME: discovering and analyzing DNA and protein sequence motiefs. Nucleic Acid Res. 2006;34(suppl_2):369–73. https://doi.org/10.1093/nar/gkl198.

Basyuni M, Wati R, Sulistiyono N, et al. Isolation and analysis of a multifunctional triterpene synthase KcMS promoter region from mangrove plant kandelia candel. IOP Conference Series: Earth and Environmental Science. Medan, Indonesia; 2018, 130(1): 012013. doi: https://doi.org/10.1088/1755-1315/130/1/012013.

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. Mol Plant. 2009;2(5):1000–14. https://doi.org/10.1093/mp/ssp062.

Campoy JA, Le Dantec L, Barreneche T, et al. New insights into fruit firmness and weight control in sweet cherry. Plant Mol Biol Report. 2015;33(4):783–96. https://doi.org/10.1007/s11105-014-0773-6.

Cantarel BL, Coutinho PM, Rancurel C, et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acid Res. 2008;37(suppl_1):233–8. https://doi.org/10.1093/nar/gkn663.

Cao J, Schneeberger K, Ossowski S, et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet. 2011;43(10):956–63. https://doi.org/10.1038/ng.911.

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. https://doi.org/10.1186/1471-2164-15-919.

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 Sci. 2018;270:257–67. https://doi.org/10.1016/j.plantsci.2018.02.017.

Chi JN, Han GY, Wang XF, et al. Isolation and molecular characterization of a novel homogalacturonan galacturonosyl-transferase gene (*GbGAUT1*) from *Gossypium barbadense*. Afr J Biotechnol. 2009;8(19):4755–64. https://doi.org/10.4314/ajb.v8i19.65164.

Du HL, Yu Y, Ma YF, et al. Sequencing and *de novo* assembly of a near complete indica rice genome. Nat Commun. 2017;8(1):15324. https://doi.org/10.1038/ncomms15324.

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. Nat Genet. 2018;50(6):796–802. https://doi.org/10.1038/s41588-018-0116-x.

Durand C, Vicré-Gibouin M, Follet-Gueye ML, et al. The organization pattern of root border-like cells of Arabidopsis is dependent on cell wall homogalacturonan. Plant Physiol. 2009;150(3):1411–21. https://doi.org/10.1104/pp.109.136382.

Fan SM. The relationship between the cellulose synthesis and cotton fiber quality. Kaifeng, China: Henan University; 2013.

Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39(suppl_2):29–37. https://doi.org/10.1093/nar/gkr367.

Godoy F, Bermúdez L, Lira BS, et al. Galacturonosyltransferase 4 silencing alters pectin composition and carbon partitioning in tomato. J Exp Bot. 2013;64(8):2449–66. https://doi.org/10.1093/jxb/ert106.

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 Res. 2007;17(5):422–3. https://doi.org/10.1038/sj.cr.7310150.

Haigler CH, Betancur L, Stiff MR, et al. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Front Plant Sci. 2012;3:104. https://doi.org/10.3389/fpls.2012.00104.

Harholt J, Suttangkakul A, Scheller HV. Biosynthesis of pectin. Plant Physiol. 2010;153(2):384–95. https://doi.org/10.1104/pp.110.156588.

Herrera-Vásquez A, Salinas P, Holuigue L. Salicylic acid and reactive oxygen species interplay in the transcriptional control of defense genes expression. Front Plant Sci. 2015;6:171. https://doi.org/10.3389/fpls.2015.00171.

Higo K, Ugawa Y, Iwamoto M, et al. Plant *cis*-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res. 1999;27(1):297–300. https://doi.org/10.1093/nar/27.1.297.

Hu B, Jin J, Guo AY, et al. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7. https://doi.org/10.1093/bioinformatics/btu817.

Hyodo H, Terao A, Furukawa J, et al. Tissue specific localization of pectin–$Ca^{2+}$ cross-linkages and pectin methyl-esterification during fruit ripening in tomato (*Solanum lycopersicum*). PLoS One. 2013;8(11):e78949. https://doi.org/10.1371/journal.pone.0078949.

Ji ZB. Gene expression profile analyses of upland cotton (*Gossypium hirsutum*) fiber during secondary cell wall deposition. Wuhan, China: Huazhong Agricultural University; 2011.

Kikuchi N, Kwon YD, Gotoh M, et al. Comparison of glycosyltransferase families using the profile hidden Markov model. Biochem Biophys Res Commun. 2003;310(2):574–9. https://doi.org/10.1016/j.bbrc.2003.09.031.

Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36. https://doi.org/10.1186/gb-2013-14-4-r36.

Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357–60. https://doi.org/10.1038/nmeth.1923.

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 Res. 2002;30(1):325–7. https://doi.org/10.1093/nar/30.1.325.

Li FG, Fan GY, Lu CR, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–30. https://doi.org/10.1038/nbt.3208.

Li WJ. Phylogeny analysis of PMEs gene family and function of GhJAZ in cotton. Kaifeng, China: Henan University; 2016.

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. https://doi.org/10.1186/s12864-016-3365-z.

Liu JX, Yu CM, Tang SW, et al. Cloning and expression of key enzyme gene *GalAT* in ramie pectin biosynthesis. Agric Sci China. 2009;8(6):664–70. https://doi.org/10.1016/S1671-2927(08)60262-X.

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. Sci Rep. 2015;5(1):14139. https://doi.org/10.1038/srep14139.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_T$ method. Methods. 2001;25(4):402–8. https://doi.org/10.1006/meth.2001.1262.

Marchler BA, Lu S, Anderson JB, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acid Res. 2010;39(suppl_1):225–9. https://doi.org/10.1093/nar/gkq1189.

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*. J Integr Plant Biol. 2014;56(4):388–99. https://doi.org/10.1111/jipb.12149.

Mohnen D. Pectin structure and biosynthesis. Curr Opin Plant Biol. 2008;11(3):266–77. https://doi.org/10.1016/j.pbi.2008.03.006.

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):613–22. https://doi.org/10.1007/s00425-005-0008-z.

Paterson AH, Wendel JF, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature. 2012;492(7429):423–7. https://doi.org/10.1038/nature11798.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. 2001;29(9):e45. https://doi.org/10.1093/nar/29.9.e45.

Porto MS, Pinheiro MP, Batista VG, et al. Plant promoters: an approach of structure and function. Mol Biotechnol. 2014;56(1):38–49. https://doi.org/10.1007/s12033-013-9713-1.

Ridley BL, O'Neill MA, Mohnen D. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry. 2001;57(6):929–67. https://doi.org/10.1016/S0031-9422(01)00113-3.

Scheller HV, Jensen JK, Sørensen SO, et al. Biosynthesis of pectin. Physiol Plant. 2007;129(2):283–95. https://doi.org/10.1104/pp.110.156588.

Sterling JD, Atmodjo MA, Inwood SE, et al. Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. Proceed Natl Acad Sci. 2006;103(13):5236–41. https://doi.org/10.1073/pnas.0600120103.

Sun FD, Zhang JH, Wang SF, et al. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. Mol Breed. 2012;30(1):569–82. https://doi.org/10.1007/s11032-011-9645-z.

Tang H, Bowers JE, Wang X, et al. Synteny and collinearity in plant genomes. Science. 2008;320(5875):486–8. https://doi.org/10.1126/science.1153917.

Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using Clustal W and Clustal X. Curr Protoc Bioinformatics. 2003;1(1):2–3. https://doi.org/10.1002/0471250953.bi0203s00.

Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5. https://doi.org/10.1038/nbt.1621.

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. https://doi.org/10.1186/s12864-015-1708-9.

Vasco M, De Stefano R, Di Matteo A, et al. Candidate genes controlling fruit quality in a tomato introgression line tolerant to water deficit. In: Proceedings of the Joint Meeting AGI-SIBV-SIGA. Assisi, Italy; 2011. p. 19–22.

Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93(1):77–8. https://doi.org/10.1093/jhered/93.1.77.

Wang KB, Wang ZW, Li FG, et al. The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet. 2012;44(10):1098–103. https://doi.org/10.1038/ng.2371.

Wang L, Wang W, Wang YQ, et al. *Arabidopsis* galacturonosyltransferase (GAUT) 13 and GAUT14 have redundant functions in pollen tube growth. Mol Plant. 2013;6(4):1131–48. https://doi.org/10.1093/mp/sst084.

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. Wuhan, China: Huazhong Agricultural University; 2012.

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 Eastern Europe. 2014;5(107):28–33.

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. Int J Plant Sci. 2009;170(8):979–89. https://doi.org/10.1086/605122.

Willats WG, Knox JP, Mikkelsen JD. Pectin: new insights into an old polymer are starting to gel. Trends Food Sci Technol. 2006;17(3):97–104. https://doi.org/10.1016/j.tifs.2005.10.008.

Yin Y, Chen H, Hahn MG, et al. Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. Plant Physiol. 2010a;153(4):1729–46. https://doi.org/10.1104/pp.110.154229.

Yin Y, Mohnen D, Xu Y, et al. Glycosyltransferases of the GT8 family. Ann Plant Rev Plant Polysaccharides Biosynthesis Bioeng. 2010b;41:167–211. https://doi.org/10.1002/9781444391015.ch6.

Zhang TZ, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531–7. https://doi.org/10.1038/nbt.3207.

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 Sci. 2017;57(2):774–88. https://doi.org/10.2135/cropsci2016.06.0544.

Zhao L, Ding Q, Zeng J, et al. Immunohistochemical study of cell wall pectin during cotton fiber development. J Chin Electron Microsc Soc. 2012;31(2):179–83 (in Chinese).

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. https://doi.org/10.3390/genes10020119.