Journal of Cotton Research

RESEARCH Open Access

Global identification of genes associated with xylan biosynthesis in cotton fiber

CHEN Feng[1], GUO Yanjun[1], CHEN Li[1], GAN Xinli[1], LIU Min[1], LI Juan[1] and XU Wenliang[1,2*]

Abstract

Background: Mature cotton fiber secondary cell wall comprises largely of cellulose (> 90%) and small amounts of xylan and lignin. Little is known about the cotton fiber xylan biosynthesis by far.

Results: To comprehensively survey xylan biosynthetic genes in cotton fiber, we identified five IRX9, five IRX10, one IRX14, six IRX15, two FRA8, one PARVUS, eight GUX, four GXM, two RWA, two AXY9, 13 TBL genes by using phylogenetic analysis coupled with expression profile analysis and co-expression analyses. In addition, we also identified two GT61 members, two GT47 members, and two DUF579 family members whose homologs in Arabidopsis were not functionally characterized. These 55 genes were regarded as the most probable genes to be involved in fiber xylan biosynthesis. Further complementation analysis indicated that one IRX10 like and two FRA8 related genes were able to partially recover the irregular xylem phenotype conferred by the xylan deficiency in their respective Arabidopsis mutant. We conclude that these genes are functional orthologs of respective genes that are implicated in GX biosynthesis.

Conclusion: The list of 55 cotton genes presented here provides not only a solid basis to uncover the biosynthesis of xylan in cotton fiber, but also a genetic resource potentially useful for future studies aiming at fiber improvement via biotechnological approaches.

Keywords: Cotton fiber, Secondary cell wall, Xylan biosynthesis, Expression profile, Co-expression

Background

Mature cotton fiber secondary cell wall (SCW) is composed mainly of cellulose (> 90%) and small amounts of noncellulosic polymers, such as xylan and lignin (Haigler et al. 2012; Han et al. 2013). Fiber cell wall composition not only defines fiber morphogenesis but also affects fiber quality and quantity (Haigler et al. 2012). For instance, the xyloglucan endo-transglycosylase/hydrolase (XTH) proteins are capable of degrading xyloglucan irreversibly, and cleaving and transferring chain ends between molecules. Transgenic cotton plants overexpressing *GhXTH1* exhibited approximately two-fold higher XET activity and 15% ~ 20% longer fiber compared with wild type cotton (Lee et al. 2010). In addition, Avci et al. (2013) found that 14 to 21 days post-anthesis (DPA) fibers of *Gossypium barbadense* barely contain loosely-bound xyloglucan whereas the *Gossypium hirsutum* fibers contain lots of xyloglucan via glycome profiling (Avci et al. 2013). Li et al. (2013) even speculated that xyloglucan might affect fiber elongation negatively by comparing xyloglucan content in *G. barbadense* and *G. hirsutum* (Li et al. 2013). Moreover, Han et al. (2013) found that overexpression of *WLIM1a* encoding a LIM-domain protein in cotton upregulated expression of the PAL-box genes and resulted in higher amounts of lignin production in transgenic fiber cells, leading to a desirable improvement in both fiber fineness and strength. The authors also proposed that lignin is likely to be a key determinant of cotton fiber quality (Han et al. 2013). Manipulating genes involved in cell wall component biosynthesis can alter fiber quality and quantity and

* Correspondence: wenliangxu@mail.ccnu.edu.cn
[1]Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan 430079, China
[2]Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou 450001, China

represents a valuable and efficient approach to genetically improve fiber quality. Therefore it is crucial to understand the structure of fiber cell walls and how different wall components are synthesized. Currently, a number of genes involved in fiber secondary cell wall cellulose and lignin synthesis have been identified (Fan et al. 2009; Han et al. 2013; Li et al. 2013). In addition, glycome profiling and immunohistochemistry techniques have exhibited the presence of xylans in 10 to 24 DPA cotton fibers in both *G. barbadense* and *G. hirsutum* (Avci et al. 2013). Moreover, a gel-state 2D-NMR method revealed the structure of 4-*O*-methylglucuronoxylan from the cotton linters (Kim and Ralph 2014). However, our knowledge of xylan biosynthesis in cotton fiber remains limited. The function of xylan in fiber development is barely understood.

Glucuronoxylan (GX) is a major hemicellulosic component in dicot secondary cell wall. GX is made of a linear chain of β-1,4-linked xylosyl residues, some of which are substituted with side chains, such as glucuronic acid (GlcA), methylglucuronic acid (MeGlcA), and also decorated with *O*-acetyl groups (Scheller and Ulvskov 2010). In addition, dicot xylan has a tetrasaccharide unit β-d-Xyl-(1,3)-α-l-Rha-(1,2)-α-d-GalA-(1,4)-d-Xyl at its reducing end named reducing end sequence (RES) (Scheller and Ulvskov 2010). GX synthesis requires the coordinated actions of a suite of various enzymes, including glycosyltransferases, acetyl transferases, and methyl transferases. Great progress has been made during the last decade in identifying genes involved in the GX biosynthesis in the model dicot plant Arabidopsis. So far 35 genes have been shown to participate in xylan synthesis. Among this set of genes, *IRX9*/ *IRX9L*, *IRX10*/*IRX10L*, *IRX14*/*IRX14L* and possibly *IRX15*/*IRX15L* were demonstrated to be required for backbone elongation (Peña et al. 2007; Brown et al. 2007; Brown et al. 2009; Wu et al. 2010; Brown et al. 2011; Jensen et al. 2011; Jensen et al. 2014; Urbanowicz et al. 2014). *FRA8*/*F8H*, *IRX8* and *PARVUS* were shown to be responsible for reducing end synthesis (Zhong et al. 2005; Peña et al. 2007; Brown et al. 2007; Lee et al. 2007; Lee et al. 2009a). GUX1-5 were glucuronyltransferases that add GlcA to xylan (Mortimer et al. 2010; Rennie et al. 2012). GXM1-3 are methyltransferases catalyzing 4-*O*-methylation of GlcA side chains on xylan (Lee et al. 2012). Last but not least, *ESK1* (*TBL29*), *TBL3*, *TBL28*, *TBL30–35*, *AXY9*, and *RWA1* to *RWA5* have been shown to be involved in GX acetylation (Lee et al. 2011; Manabe et al. 2011; Urbanowicz et al. 2014; Schultink et al. 2015; Zhong et al. 2017). Many studies have shown that most proteins responsible for the GX biosynthesis are highly conserved among lower plants such as *Physcomitrella* and higher plants such as Arabidopsis and Populus (Hörnblad et al. 2013; Zhou et al. 2006; Zhou et al. 2007).

A comprehensive understanding of xylan biosynthesis in cotton fiber will require identification of all the genes involved just like what have achieved in Arabidopsis. In this report, we applied a bioinformatics approach aimed at identifying candidate genes as many as possible putatively involved in fiber xylan biosynthesis. Fifty-five genes were shown to be strong candidates responsible for xylan synthesis by phylogenetic analysis and expression pattern analysis. Three of these genes were further demonstrated to function in xylan biosynthesis by complementation analysis. The study of xylan synthesis in cotton fiber with unusual cell wall composition will undoubtedly shed valuable light on manipulating plant cell walls with enhanced productivity and biotechnological potential.

Materials and methods

Identification of candidate genes

The published *G. hirsutum* genome sequence databases offer unique opportunities to identify genes involved in xylan biosynthesis in upland cotton (Li et al. 2015; Zhang et al. 2015). The known 35 Arabidopsis proteins involved in xylan synthesis were used to blast search the *G. hirsutum* genome database [https://www.cottongen.org/, *G.hirsutum* NBI proteins (70478)]. In addition, some conserved protein domains (such as DUF579, DUF231, Casp1) were also used as a query to blast search the upland cotton genome database. The obtained sequences were sequentially passed through phylogenetic analysis, expression profile analysis and co-expression analysis.

Phylogenetic analysis

The multiple sequence alignments were performed by using clustal W. A neighbor-joining phylogenetic tree of the complete amino acid sequences of homologous proteins in a same GT family or containing the same domains from Arabidopsis and cotton was then constructed using MEGA7.0 with 1 000 bootstraps. The phylogenetic tree was displayed using iTOL (http://itol.embl.de/) with default settings.

Expression analyses of cotton fiber xylan synthesis associated genes using RNA-seq data

To investigate the spatial and temporal expression profiles of the identified genes, we analyzed their transcript levels using RNA-seq data of 10 different tissues and 9 different developmental stages of ovules and fibers, high throughput RNA-seq data of 9 samples from leaves, roots and stems of 2-week-old plants; petals, torus, pistils and stamens; ovules from − 3, − 1, 0, 1, 3, 5, 10, 20 and 25 DPA; fibers from 5, 10, 20 and 25 DPA were downloaded from http://www.ncbi.nlm.nih.gov/SRA: PRJNA248163 or CottonGen database (https://www.cottongen.org/). Expression level of each gene was calculated using FPKM, and those expressions with an FPKM < 1 were not considered. Because the expression level of the corresponding genes

was too low to be related and probably are not essential for xylan biosynthesis. Heat maps and hierarchical clustering were generated with Omicshare (http://www.omicshare.com/tools/Home/Soft/heatmap).

Co-expression analysis

Co-expression analysis was performed using a database of co-expression networks with functional modules for diploid and polyploidy Gossypium (http://structuralbiology.cau.edu.cn/gossypium/).

Plant material and growth conditions

Arabidopsis seeds were surface-sterilized and planted as described previously (Xu et al. 2013).

Reverse transcription polymerase chain reaction (RT-PCR) analysis

Total RNA was extracted from 6-week-old Arabidopsis stems as previously described (Xu et al. 2013). A two-step RT-PCR procedure was performed according to the method described previously (Li et al. 2014).

Complementation of Arabidopsis *fra8* and *irx10* mutant

The open reading frames of *GhGT47A1, GhGT47B1* and *GhGT47B2* were amplified by PCR and cloned into pBI121 vector. The resulting constructs were transferred into *irx10* and *fra8* mutants respectively by floral dip method. Seeds were harvested and stored at 4 °C. Positive first-generation (T1) transgenic plants were screened on MS medium containing hygromycin. Homozygous transgenic plants were selected for further analysis.

Sectioning of stems and roots and microscopy analysis

Freehand cross-sections of 6-week-old stems and roots were conducted as previously described (Li et al. 2014). For lignin staining, sections were stained with phloroglucinol-HCl as described previously (Huang et al. 2016).

Results

Identification of genes involved in fiber xylan backbone and RES synthesis

In Arabidopsis, members of glycosyltransferase family 43 (GT43) and GT47 are responsible for synthesis of the xylan backbone: IRX9 and IRX14 from GT43 and IRX10 from GT47 (Brown et al. 2007). Each of them has a single partially redundant paralog: IRX9L, IRX14L and IRX10L, respectively (Brown et al. 2009; Wu et al. 2010). By far only IRX10 and IRX10L have been demonstrated to possess β-(1,4)-xylosyltransferase activity in vitro (Urbanowicz et al. 2014; Jensen et al. 2014). IRX9 (L) and/or IRX14 (L) were proposed to act as accessory proteins and play primarily structural roles in xylan synthase complex (XSC) (Zeng et al. 2016).

We first focused on GT43 family. The GT43 family contains 18 members in cotton genome (Fig. 1a). A phylogenetic analysis of GT43 family members showed that the tree was divided into four subclades, in which *Gh_A09G1418*, *Gh_D09G1426* (*GhGT43A*, Li et al. 2014) were clustered with IRX9 and comprised a subclade, while *Gh_D04G2324*, *Gh_A06G0232*, *Gh_D02G0864*, *Gh_A02G0815*, *Gh_D01G1084* and *Gh_A01G1031* were mostly closely related to IRX9L and constituted another subclade. IRX14 and IRX14L were grouped with *Gh_A08G0365*, *Gh_D11G1966*, *Gh_A11G1802*, *Gh_D03G0229* and *Gh_A02G1486* and comprised a subclade. The remaining 5 genes comprised an additional subclade (Fig. 1a). Among the 13 cotton IRX9 and IRX14 genes, two IRX9 genes (*Gh_A09G1418* and *Gh_D09G1426*) and one IRX9L (*Gh_A06G0232*) were specifically expressed in 20 DPA fibers, two IRX9L genes (*Gh_D01G1084* and *Gh_D06G2324*) were predominantly expressed in 20 DPA fibers (Fig. 2a, Additional file 1: Fig. S1). Co-expression analysis revealed that the five 20 DPA fiber specifically or preferentially expressed genes were co-expressed with SCW-associated genes (Additional file 9: Table S1). Thus they represent strong candidates. One (*Gh_D03G0229*) of the five IRX14/IXR14L genes was preferentially expressed in ovules, but during the fiber developmental stage, it was predominantly expressed in 20 DPA fibers and also co-expressed with SCW genes (Fig. 2b, Additional file 1: Fig. S1, Additional file 9: Table S1).These analyses demonstrated that cotton fiber may also employ IRX9 and IRX14-like genes to synthesize xylan backbone.

Next, we focused on the GT47 family. As shown in Fig. 1b, IRX10 subclade with 5 cotton genes and FRA8 subclade containing 4 cotton genes constitute a clade. Though IRX10 and FRA8 are closely related to each other, they perform completely different functions, as IRX10 is involved in xylan backbone synthesis while FRA8 participates in RES biosynthesis in Arabidopsis (Zhong et al. 2005; Brown et al. 2007). These two distinct enzyme activities within the same phylogenetic clade highlight the need for detailed investigation of specific function of every member in the same GT47 subgroup. Two (*Gh_D02G1793* and *Gh_A03G1353*, thereafter named as *GhGT47A1* and *GhGT47A2*, respectively) of the five IRX10L genes were specifically and mostly strongly expressed in 20 DPA fibers and another IRX10L gene (*Gh_D12G1919*) was predominantly transcribed in 20 DPA fibers (Fig. 2a, Additional file 2: Fig. S2). In addition, two IRX10L genes (*Gh_A12G1292* and *Gh_D12G1414*) were preferentially expressed in leaves, but during the fiber developmental stage, they were dominantly expressed in 20 DPA fibers (Fig. 2b, Additional file 2: Fig. S2). Moreover, all of the five IRX10 genes were correlated with SCW associated genes

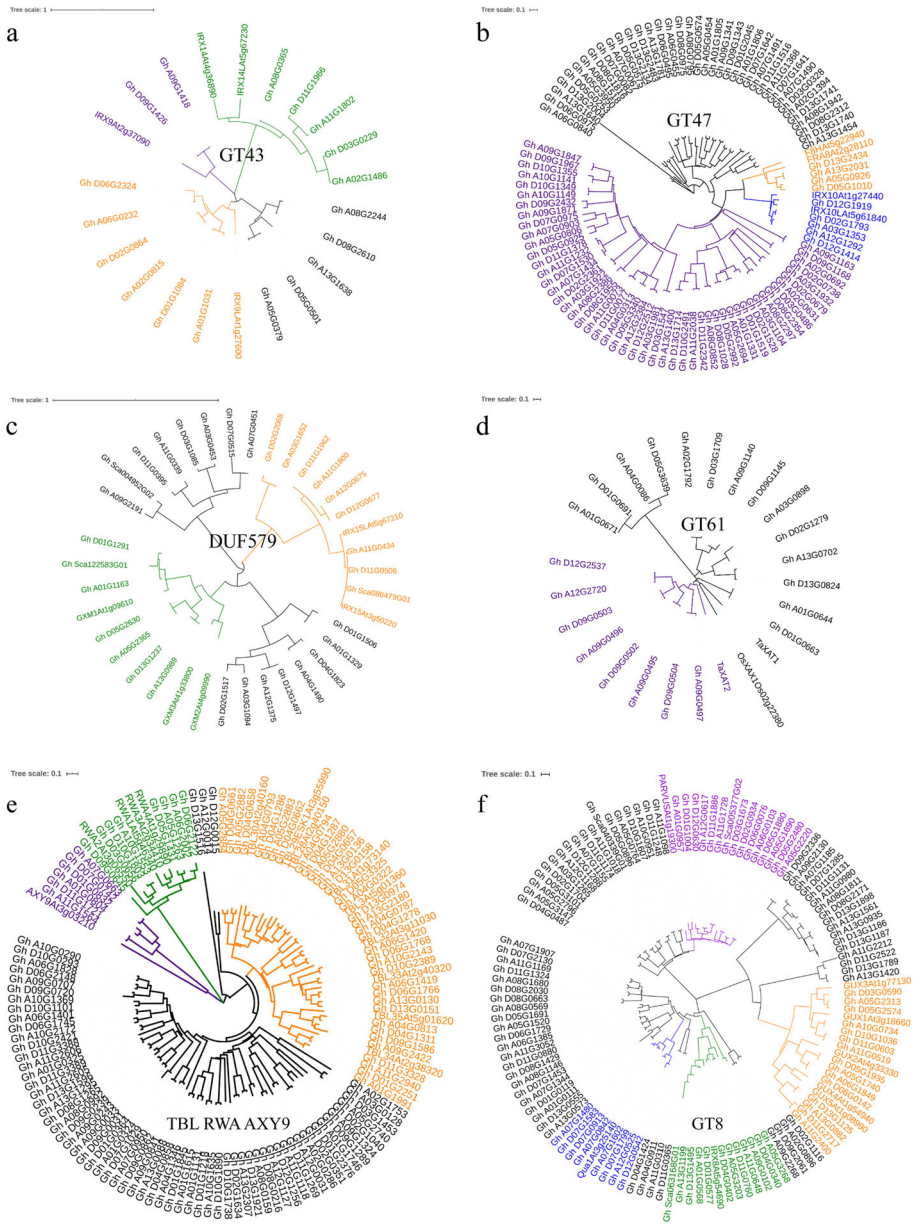

Fig. 1 Phylogenetic analysis of putative cotton fiber xylan synthesis-associated proteins and 35 Arabidopsis homologs. **a** Phylogenetic tree of Arabidopsis and cotton GT43 family members. Several xylan-associated Arabidopsis GT43 members and their cotton homologs are highlighted in different color. **b** Phylogenetic tree of Arabidopsis and cotton GT47 family members. Several xylan-associated Arabidopsis GT47 members and their cotton homologs are highlighted in different color. **c** Phylogenetic tree of Arabidopsis and cotton DUF579-domain containing proteins. Several xylan-associated Arabidopsis DUF579-domain containing proteins and their cotton homologs are highlighted in different color. **d** Phylogenetic tree of Arabidopsis and cotton GT61 family members. **e** Phylogenetic tree of Arabidopsis and cotton TBL, RWA and AXY9 family members. The cotton TBL family comprises 100 genes in three subclades, with 9 additional genes homologous to reduced wall O-acetylation (RWA) and 6 related to Arabidopsis xyloglucan acetyl transferase9 (AXY9). **f** Phylogenetic tree of Arabidopsis and cotton GT8 family members. Several xylan-associated Arabidopsis GT8 members and their cotton homologs are highlighted in different color

(Fig. 3a, Additional file 9: Table S1). Thus, they represent strong candidates for fiber xylan biosynthesis.

Though the FRA8 subclade contains 4 cotton homologs (Fig. 1b), only two (*Gh_A13G2031* and *Gh_D13G2434*, hereafter named as *GhGT47B1* and *GhGT47B2*, respectively) of them were specifically expressed in 20 DPA fiber and co-expressed with SCW genes (Fig. 2a, Fig. 3b, Additional file 2: Fig. S2, Additional file 9: Table S1). In Arabidopsis, except for IRX10(L) and FRA8(L) in GT47 family, it appears that no other GT47 members function in xylan synthesis. However, we found that *Gh_A13G1640* and *Gh_*

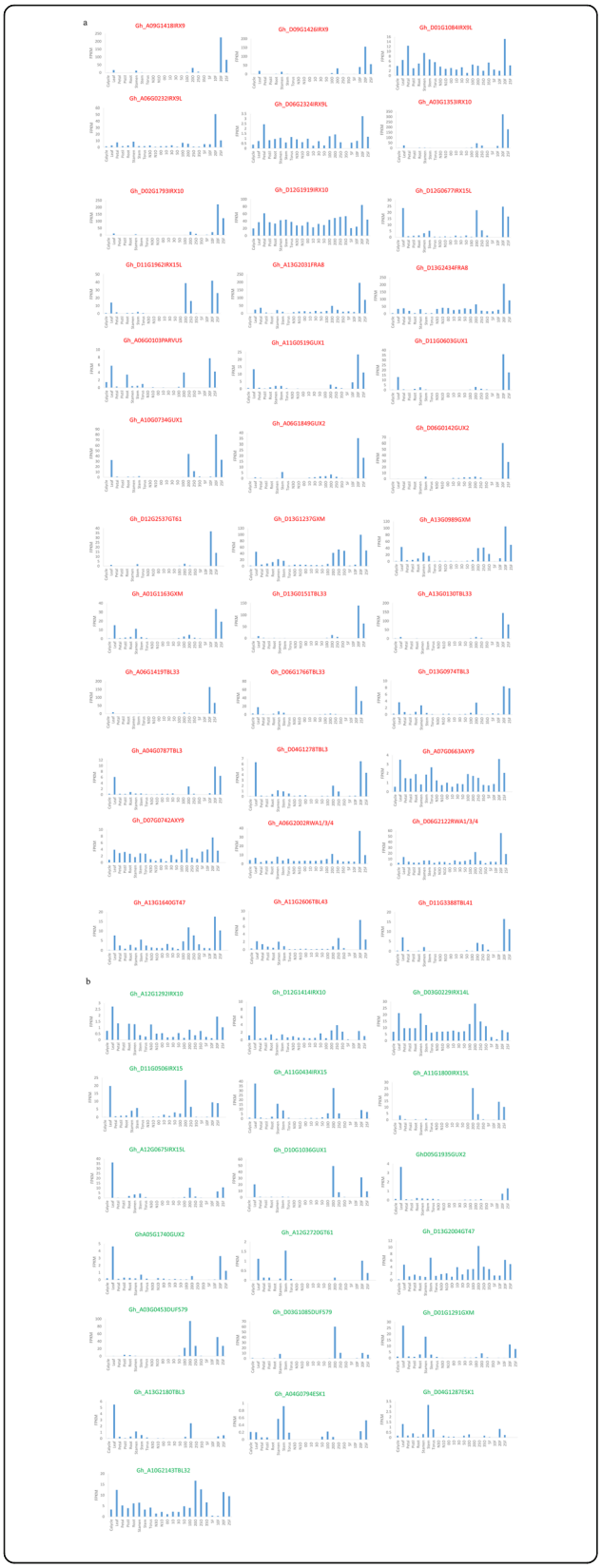

Fig. 2 Expression profiles of 55 fiber xylan synthesis-associated candidates in different cotton tissues. **a** 36 fiber xylan synthesis-associated candidates showed greatest expression in 20 DPA fiber. **b** 19 fiber xylan synthesis-associated candidates were expressed highly in 20 DPA fiber. -3O:Ovule in 3 days before-anthesis; −1O:Ovule in 1 day before-anthesis; 0O:Ovule in anthesis; 1O:Ovule in 1 day post-anthesis (DPA); 3O:Ovule in 3 DPA; 5O:Ovule in 5 DPA; 10O:Ovule in 10 DPA; 20O:Ovule in 20 DPA; 25O:Ovule in 25 DPA; 35O:Ovule in 35 DPA; 5F:Fiber in 5 DPA; 10F:Fiber in 10 DPA; 20F:Fiber in 20 DPA; 25F:Fiber in 25 DPA

D13G2004 located in a different clade within the GT47 tree were not only expressed highly in 20 DPA fibers (Fig. 2a and b, Additional file 2: Fig. S2) but also co-expressed with SCW genes (Additional file 9: Table S1). Due to the lack of functional validation, by now we can't determine whether *Gh_A13G1640* and *Gh_D13G2004* participate in backbone synthesis or RES synthesis. They may represent novel candidates. This finding might reflect the specificity of xylan synthesis in cotton fiber, as compared with Arabidopsis stem.

As for the reducing end sequence synthesis, in addition to FRA8 (L) from GT47, two other members from GT8, IRX8 and PARVUS, were reported to be involved (Lee et al. 2007; Peña et al. 2007). In cotton genome, 115 genes were annotated as GT8 family members (Fig. 1f). A phylogenetic analysis displayed that the tree was divided into three clades. IRX8 together with 13 cotton homologs were in one clade. PARVUS and 15 cotton homologs were in another clade (Fig. 1f). Among the 13 cotton IRX8 genes, none of the genes was expressed highly in 20 DPA fibers, and thus they are not likely to participate in xylan synthesis in fiber cell (Additional file 6: Fig. S6). Within the PARVUS subclade, only one gene (*Gh_A06G0103*) was preferentially expressed in 20 DPA fiber and co-expressed with SCW genes (Fig. 2a, Additional file 6: Fig. S6, Additional file 9: Table S1) and thus regarded as a strong candidate. It seems that cotton fiber employs different genes to synthesize the RES or contains a different RES structure in relative to that of Arabidopsis stem.

Identification of fiber xylan modification genes

Xylan can be modified by the addition of GlcA and MeGlcA side chains by xylan glucuronosyltransferases of the GlcA substitution of xylan (GUX) family which belong to the GT8 family (Mortimer et al. 2010; Rennie et al. 2012; Mortimer et al. 2015). In Arabidopsis, there exist five *GUX* genes. Although GUX3 and GUX5 are closely related to GUX1, GUX2 and GUX4, only GUX1, GUX2 and GUX4 have xylan α-glucuronosyltransferase activity, while GUX3 is involved in primary cell wall polymer synthesis (Rennie et al. 2012; Mortimer et al. 2015). Within the GT8 phylogenetic tree, the 5 Arabidopsis GUX proteins and 15 cotton homologs

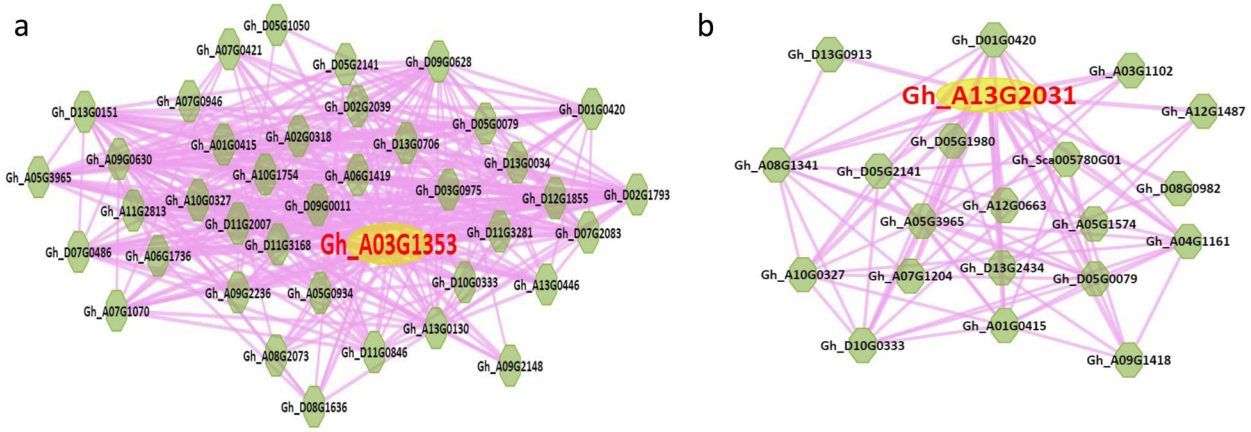

Fig. 3 Co-expressed genes with *GhGT47A1* (*Gh_A03G1353*) and *GhGT47B1* (*Gh_A13G2031*). Protein in yellow elliptic box represents query protein. Proteins in green hexagon boxes represent interaction proteins. Pink interaction line indicates protein has interaction and positive co-expression relationship with target protein. In (**a**), *Gh_D07G2083*, *Gh_D05G0079*, *Gh_A05G3965*, *Gh_D10G0333* and *Gh_A10G0327* were annotated as homologs of secondary cell wall cellulose synthase, *Gh_D02G1793* was annotated as IRX10 homolog, *Gh_A13G0130*, *Gh_D13G0151* and *Gh_A06G1419* were annotated as homologs of DUF231 domain proteins. In (**b**), *Gh_D05G0079*, *Gh_A05G3965*, *Gh_A10G0327* and *Gh_D10G0333* were annotated as homologs of secondary cell wall cellulose synthase, *Gh_A09G1418* was annotated as IRX9 homolog, *Gh_D13G2434* was annotated as FRA8 homolog, *Gh_A07G1204* was annotated as MYB61 homolog

constituting one subclade were in the third clade (Fig. 1f). The GUX subclade was further divided into four subgroups, including *Gh_A10G0734*, *Gh_D10G1036*, *Gh_D11G0603*, *Gh_A11G0519* were clustered with GUX1 and comprised one subclade; *Gh_D05G1935*, *Gh_A05G1740*, *Gh_A06G1849*, *Gh_D06G0142* were grouped with GUX2 and constituted another subclade (Fig. 1f). GUX3 subclade contains three cotton proteins, while GUX4/GUX5 subclade contains 4 cotton proteins (Fig. 1f). It is very interesting to find that genes grouped with GUX1 and GUX2 were all predominantly or highly expressed during a stage of secondary cell wall thickening stage (Fig. 2a and b, Additional file 6: Fig. S6). Moreover, the 8 cotton GUX1/2 related genes all co-expressed with SCW biosynthesis-associated genes (Additional file 9: Table S1). These results are consistent with what has been shown in Arabidopsis, where GUX1 and GUX2 are SCW specific (Rennie et al. 2012). It should be noted that two cotton GUX3-like genes (Fig. 1f), *Gh_D03G0590* and *Gh_D05G2574*, also showed very high-level expression during fiber secondary cell wall thickening stage (Additional file 7: Fig. S7). In addition, a pair of homologous genes (*Gh_A13G1420* and *Gh_D13G1789*) in another subclade next to the GUX subclade also exhibited high expression in 20 DPA fibers (Fig.1f, Additional file 7: Fig. S7). Though these genes were not co-expressed with SCW genes, they might be plausible candidates of xylan synthesis. This finding is in contrast to GUX3, as it has been shown to participate in primary cell wall xylan biosynthesis. Mutations of both *GUX1* and *GUX2* lead to only acetylated xylan with scarcely detectable GlcA modification, indicating that xylan GlcA modification in Arabidopsis was primarily catalyzed by GUX1 and GUX2. However, in cotton fiber, though GUX1 and GUX2-like genes are highly or preferentially expressed in 20 DPA fibers and co-expressed with SCW genes. Genes in the GUX3 subclade also showed high-level expression in 20 DPA fibers suggesting that they may play some roles in xylan modification even though functional verification is needed (Additional file 6: Fig. S6). These results suggest that cotton fiber may employ different genes for xylan GlcA modification.

Moreover, the GlcA substituents of xylan are often methylated by glucuronoxylan methyltransferases (GXMs) belonging to the DUF579 family (Lee et al. 2012). This type of methylation occurs exclusively to SCW xylan. The DUF579 family includes 32 members in cotton (Fig. 1c). Phylogenetic analysis showed that Arabidopsis GXMs were closely related to 7 cotton homologs and comprised one subclade (Fig. 1c). Four (*Gh_D13G1237*, *Gh_A13G0989*, *Gh_A01G1163* and *Gh_D01G1291*) of the seven GXM genes displayed predominant or high-level expression in 20 DPA fibers and were correlated with SCW genes (Fig. 2a and b, Additional file 5: Fig. S5, Additional file 9: Table S1), so they are likely to be methyltransferases and their enzymatic activities await further studies.

IRX15 and IRX15L also belong to the DUF579 family. However, it is apparent that they perform different functions from that of GXMs. The mutation of GXM genes led to a specific defect in GlcA methylation on xylan while simultaneous mutations of *IRX15* and *IRX15L* caused reduced xylan level and chain length, decreased xylosyltransferase activity without affecting the GlcA methylation of xylan (Brown et al. 2011; Jensen et al.

2011). IRX15/IRX15L and 9 cotton homologs constitute another subclade in the DUF579 family (Fig. 1c). Of the 9 cotton IRX15 genes, two (*Gh_D12G0677* and *Gh_D11G1962*) were preferentially and four (*Gh_D11G0506*, *Gh_A11G0434*, *Gh_A11G1800* and *Gh_A12G0675*) were highly expressed in 20 DPA fibers and the six genes were all co-expressed with SCW genes (Fig. 2a and b, Additional file 5: Fig. S5, Additional file 9: Table S1). Therefore they represent candidates for xylan synthesis. Very interestingly, two additional genes (*Gh_A03G0453* and *Gh_D03G1085*) located in a separate subclade were also highly expressed in 20 DPA fibers and co-expressed with SCW-associated genes (Fig. 1c, Fig. 2b, Additional file 5: Fig. S5, Additional file 9: Table S1). This implies that *Gh_A03G0453* and *Gh_D03G1085* may be novel fiber SCW xylan biosynthesis related genes. Whether they have a function similar to GXM or IRX15 awaits further studies.

In addition to the sugar side groups, xylans can have high degree of acetylation (Scheller and Ulvskov 2010). So far three different protein families have been shown to be involved in xylan *O*-acetylation. One is the TRICHOME-BIREFRINGENCE-LIKE (TBL) protein family which includes 46 members in Arabidopsis. Nine of them are involved in the *O*-acetylation of xylan. TBL28, ESK1/TBL29, TBL3, TBL31 and TBL34 catalyzed xylan 2-*O*- and 3-*O*-monoacetylation and 2,3-di-*O*-acetylation with differential positional preference (Urbanowicz et al. 2014; Zhong et al. 2017); TBL30 catalyzed 2-*O*- and 3-*O*-monoacetylation(Zhong et al. 2017); TBL35 performed 2,3-di-*O*-acetylation and TBL32 and TBL33 carried out 3-*O*-acetylation of 2-*O*-GlcA-substituted xylosyl residues (Zhong et al. 2017). In rice, 66 TBL genes were identified (Gao et al. 2017), 14 of which show xylan 2-*O*- and 3-*O*-acetyltransferase activity. In poplar 64 TBLs have been identified, 12 of which are shown to *O*-acetylate xylan. In Arabidopsis ESK1/TBL29 appears to be the principle TBL for xylan acetylation in Arabidopsis (Zhong et al. 2017). However, in cotton two ESK1-like genes showed the greatest expression in stems whereas moderate expression in 20 DPA fibers suggesting they might play minor roles in fibers (Fig. 1e, Fig. 2b, Additional file 3: Fig. S3). Additionally, 4 TBL3-like genes (*Gh_D13G0974*, *Gh_A04G0787*, *Gh_D04G1278* and *Gh_A13G2180*), 4 TBL33 related genes (*Gh_D13G0151*, *Gh_A13G0130*, *Gh_A06G1419* and *Gh_D06G1766*) and one TBL32-like gene (*Gh_A10G2143*) were most probable candidates for fiber xylan acetylation as they were not only expressed strongly in 20 DPA fibers but also co-expressed with SCW biosynthesis genes (Fig. 1e, Fig. 2a and b, Additional file 3: Fig. S3, Additional file 9: Table S1). Genes in TBL30, TBL31, TBL34 and TBL35 subclades were expressed extremely low in 20 DPA fibers and excluded as candidates (Additional file 3: Fig. S3). It should be noted that the 4 TBL33 subclade genes showed the highest and specific expression in 20 DPA fibers, suggesting they might be the major enzymes for xylan acetylation (Fig. 1e, Fig. 2a, Additional file 3: Fig. S3). It is also interesting to find that two genes (*Gh_A11G2606* and *Gh_D11G3388*) located outside of the TBL subclade were predominantly transcribed in 20 DPA fibers and co-expressed with SCW genes (Figs. 1e and 2a, Additional file 3: Fig. S3), indicating that they may be new players in SCW xylan-modification. These findings suggest that fiber xylan acetylation pattern might be different to that of Arabidopsis stem as the involved genes are different. It needs further investigation on whether other TBLs are required for this acetylation process. In summary, in cotton genome we identified 100 TBL genes and found that 13 of them were strong candidates of xylan acetyltransferase (Additional file 3: Fig. S3), although enzymatic activity and specificity of their encoded enzymes are currently lacking.

A second protein family involved in xylan *O*-acetylation is represented by ALTERED XYLOGLUCAN 9 (AXY9). Contrary to the substrate specificity of the 9 TBL proteins, AXY9 seems to be non-specific in polysaccharide *O*-acetylation as *axy9* mutant plants display reductions in the acetylation of both xylan and xyloglucan (Schultink et al. 2015). In addition, biochemical studies showed that recombinant AXY9 did not exhibit any acetyltransferase activity, suggesting that AXY9 might act as an intermediate transferring acetyl groups used later by TBL proteins. We identified 6 AXY9 like genes in cotton, only two (*Gh_D07G0742* and *Gh_A07G0663*) of which displayed 20 DPA fiber preferential expression correlated with SCW genes (Figs. 1e and 2a). REDUCED WALL O-ACETYLATION (RWA) is the third protein family involved in xylan *O*-acetylation (Manabe et al. 2011). There are 4 RWA genes in Arabidopsis and simultaneous mutations of the four genes result in an overall reduction in the acetylation of both xylan and xyloglucan (Lee et al. 2011). Similarly, RNAi-silencing of the 4 RWA orthologs in hybrid aspen leads to reduced wood xylan and xyloglucan *O*-acetylation, suggesting the conservative functions of RWA proteins among plant species (Pawar et al. 2017). We identified 8 RWA related genes in cotton (Fig. 1e). *Gh_A10G1553*, *Gh_D10G1803* and *Gh_D03G0656* were mostly closely related to RWA2 whereas RWA1/3/4 together with *Gh_D06G2122*, *Gh_A06G2002*, *Gh_A05G1263* and *Gh_D05G1431* was grouped together (Fig. 1e). Two (*Gh_A10G1553* and *Gh_D10G1803*) of the three cotton RWA2 genes were co-expressed with primary cell wall genes, however, they were highly expressed in 20 DPA fibers, and thus they could be fiber SCW xylan-modification genes (Fig. 1e, Additional file 4: Fig. S4, Additional file 8: Fig. S8, Additional file 9: Table S9).

Two of the four cotton RWA1/3/4 genes (*Gh_A06G2002* and *Gh_D06G2122*) were predominantly expressed during fiber secondary cell wall and co-expressed with SCW genes (Figs. 1e and 2a, Additional file 4: Fig. S4), so they could represent the predominant RWA proteins that involved in fiber xylan acetylation. The other two cotton RWA1/3/4 genes (*Gh_A05G1263* and *Gh_D05G1431*) were highly transcribed in 20 DPA fibers but not co-expressed with SCW genes (Fig. 1e, Additional file 4: Fig. S4, Additional file 7: Fig.S7). They could be involved in xylan acetylation. It should be noted that RWAs were proposed to be responsible for the translocation of acetyl-groups across the membrane to supply the substrate for AXY9 and TBL protein families, although no experimental data has been provided yet (Lee et al. 2011; Manabe et al. 2011; Schultink et al. 2015).

Very interestingly, we identified one gene (*Gh_D12G2537*) from GT61 family that is not only preferentially expressed in 20 DPA fibers but also co-expressed with xylan backbone synthesis genes IRX9 and IRX9L (Figs. 1d and 2a, Additional file 9: Table S1), so it may represent a strong candidate involved in xylan modification. *Gh_A12G2720* and *Gh_D12G2537* is a homoeologous pair in allotetraploid cotton. *Gh_A12G2720* was preferentially expressed in stems, but during the fiber developmental stage, it was predominantly expressed in 20 DPA fibers (Fig. 2b). Moreover, it was also co-expressed with SCW related genes (Additional file 9: Table S1), and thus it can be considered as a candidate. *Gh_D12G2537* was most closely related to OsXAT3, a xylan arabinosyltransferase (XAT) from the GT61 family, responsible for the addition of arabinose residues to the xylan backbone to produce arabinoxylan in rice (Anders et al. 2012). This implies that fiber SCW xylan may be substituted with arabinose, in contrast with Arabidopsis that does not contain Arabinoxylan.

Combining all of the evidences from phylogenetic, expression profiling and co-expression analyses, in total we identified 55 genes as most probable fiber xylan biosynthesis genes (Table 1, Additional file 10: Table S2, Additional file 11: Table S3).

GhGT47A1 (*Gh_D02G1793*) can genetically complement the *irx10* phenotype

The Arabidopsis GT mutant complementation approach has been successfully used to explore the functions of several corresponding GT homologs in xylan synthesis in poplar (Zhou et al. 2006; Lee et al. 2009b). We also employed this method to study the functions of *GhGT43A* and *GhGT43C* previously (Li et al. 2014). Accordingly, in this study we applied this approach to establish function of candidate genes. We first selected one fiber-preferential gene *GhGT47A1* and used the complementation analysis to investigate if GhGT47A has similar function as its Arabidopsis ortholog IRX10.

We transformed the full-length *GhGT47A1* cDNA into the Arabidopsis *irx10* mutant to see whether *GhGT47A1* can restore the *irx10* phenotypes. Transgenic lines were tested for the presence of the *GhGT47A1* transgene in a homozygous *irx10* background. *GhGT47A1* gene-specific primers were used to check the presence of *GhGT47A1* mRNA in the transgenic Arabidopsis lines (Fig. 4c). The absence of *IRX10* transcript in *irx10* mutant and the transgenic line was also confirmed (Fig. 4c). Homozygous *irx10* mutant displayed dark-green leaves, slightly shorter stature. *GhGT47A1* expression in *irx10* mutant recovered the leaf color and the plant height almost identical to the wild type after 6-week growth (Fig. 4b). However, we observed that after transplantation into soil for 3 weeks from one-week plate-grown seedlings, the inflorescence stem height of *irx10* mutant is significant shorter than that of the wild type, whereas the *irx10-GhGT47A*1 complemented plant shows an intermediate height between *irx10* mutant and the wild type (Fig. 4a). Most obviously, more than 60% of siliques in *irx10* mutant are shrunken and sterile while only 20% of the complemented siliques are sterile (Fig. 4d). It should be noted that the number of siliques was not significantly different.

Cross sections of basal stems from *irx10* mutant exhibit a moderate irregular xylem phenotype (Additional file 8: Fig. S8b and 8e). This phenotype was reported previously (Brown et al. 2009). We did not find any apparent difference between the complemented plants and the *irx10* mutant, as stems of the *GhGT47A1* complemented plants also exhibit a mild irregular xylem phenotype (Additional file 8: Fig. S8c and 8f). However, in sharp contrast with the stem phenotype, the *irx10* roots display an obvious irregular xylem phenotype in which most of the xylem vessels are almost completely collapsed and distorted relative to the wild type (Fig. 4e and f). Compared with the *irx10* mutant, the shape of root xylem vessels from *GhGT47A1* complemented plants is very close to the wild type (Fig. 4e, f and g). Meanwhile, compared with the wild type, *irx10* showed much weaker red phloroglucinol staining, whereas the *GhGT47A1*-complemented *irx10* showed the intermediate degree of staining, which is weaker than the wild type but stronger than *irx10* (Fig. 4h, i and j).

Collectively, *GhGT47A1* expression in *Arabidopsis irx10* mutant is able to recover the root irregular xylem phenotype. These results demonstrate that *GhGT47A1* can largely complement the xylan deficiency of *irx10* mutant which suggests that GhGT47A1 is functionally conserved with IRX10 and performs a similar biochemical function as IRX10. Taken together, our results demonstrate that GhGT47A1 most likely performs a similar

Table 1 Fifty-five candidates involved in fiber xylan synthesis in upland cotton

Function	Arabidopsis Gene names[a]	G.hirsutum Gene ID	Sequence identity /%[b]	GT family or domain	References[c]
Backbone	IRX9At2g37090	*Gh_D09G1426*	61.71	GT43	Brown et al. 2007; Peña et al. 2007; Brown et al. 2009; Wu et al. 2010; Jensen et al. 2014; Urbanowicz et al. 2014
		Gh_A09G1418	61.43	GT43	
	IRX9L At1g27600	*Gh_A06G0232*	62.13	GT43	
		Gh_D01G1084	60.05	GT43	
		Gh_D06G2324	63.61	GT43	
	IRX10At1g27440/ IRX10L At5g61840	*Gh_D12G1919*	90.93	GT47	
		Gh_D02G1793	90.80	GT47	
		Gh_A03G1353	90.55	GT47	
		Gh_D12G1414	89.38	GT47	
		Gh_A12G1292	87.16	GT47	
	IRX14At4g36890/ IRX14L At5g67230	*Gh_D03G0229*	69.72	GT43	
	IRX15At3g50220/ IRX15L At5g67210	*Gh_D11G0506*	70.61	DUF579	Brown et al. 2011; Jensen et al. 2011
		Gh_A11G0434	68.75	DUF579	
		Gh_A11G1800	68.71	DUF579	
		Gh_D11G1962	69.40	DUF579	
		Gh_D12G0677	69.61	DUF579	
		Gh_A12G0675	69.03	DUF579	
Reducing end	FRA8At2g28110/F8H At5g22940	*Gh_A13G2031*	67.57	GT47	Brown et al. 2007; Lee et al. 2007; Lee et al. 2009a; Zhong et al. 2005
		Gh_D13G2434	67.04	GT47	
	PARVUS/GATL1At1g19300	*Gh_A06G0103*	80.56%	GT8	
Side chains (Glucuronosyltransferases)	GUX1At3g18660	*Gh_A10G0734*	65.51	GT8	Mortimer et al. 2010; Rennie et al. 2012
		Gh_D10G1036	65.51	GT8	
		Gh_A11G0519	66.77	GT8	
		Gh_D11G0603	66.36	GT8	
	GUX2At4g33330	*Gh_A06G1849*	62.01	GT8	
		Gh_D06G0142	62.12	GT8	
		Gh_D05G1935	63.71	GT8	
		Gh_A05G1740	63.82	GT8	
GlcA methyltransferases	GXM1At1g09610/ GXM2At4g09990/ GXM3 At1g338001	*Gh_A01G1163*	70.61	DUF579	Lee et al. 2012
		Gh_D01G1291	69.93	DUF579	
		Gh_D13G1237	64.87	DUF579	
		Gh_A13G0989	63.80	DUF579	
Acetylation (acetyltransferases)	ESK1/TBL29At3g55990	*Gh_D04G1287*	68.50	DUF231	Urbanowicz et al. 2014; Zhong et al. 2017
		Gh_A04G0794	68.35	DUF231	
	TBL33At2g40320	*Gh_D06G1766*	79.34	DUF231	
		Gh_A06G1419	78.87	DUF231	
		Gh_D13G0151	76.51	DUF231	
		Gh_A13G0130	76.51	DUF231	
	TBL3At5g01360	*Gh_A13G2180*	66.51	DUF231	
		Gh_D13G0974	66.35	DUF231	
		Gh_D04G1278	66.59	DUF231	
		Gh_A04G0787	66.35	DUF231	
	TBL32At3g11030	*Gh_A10G2143*	63.66	DUF231	

Table 1 Fifty-five candidates involved in fiber xylan synthesis in upland cotton *(Continued)*

Function	Arabidopsis Gene names[a]	G.hirsutum Gene ID	Sequence identity /%[b]	GT family or domain	References[c]
	TBL43 At2G30900	*Gh_A11G2606*	52.00	DUF231	
	TBL41 At3G14850	*Gh_D11G3388*	57.00	DUF231	
	At5g46340RWA1/At2g34410RWA3/ At1g29890RWA4	*Gh_A06G2002*	82.90	Casp1	Lee et al. 2011; Manabe et al. 2011;
		Gh_D06G2122	85.14	Casp1	
	AXY9At3g03210	*Gh_D07G0742*	40.47	Casp1	Schultink et al. 2015
		Gh_A07G0663	40.18	Casp1	
Side chain(Arabinosyltransferase)	TaXAT2 (*Triticum aestivum*, wheat)	*Gh_D12G2537*	39.80	GT61	Anders et al. 2012
		Gh_A12G2720	39.71	GT61	
Unknown function	At5G25820	*Gh_A13G1640*	46%	GT47	
Unknown function	At4G32790	*Gh_D13G2004*	72%	GT47	
Unknown function	At4G24910	*Gh_D03G1085*	51%	DUF579	
Unknown function	At4G24910	*Gh_A03G0453*	57%	DUF579	

[a] All genes are from Arabidopsis, except TaXAT2 that is from wheat
[b] The value shows the sequence identity between the Arabidopsis protein and the closely related cotton protein
[c] The references denote representative studies on xylan biosynthesis in Arabidopsis

catalytic function as IRX10 which has been shown to be involved in the synthesis of GX backbone in Arabidopsis (Brown et al. 2009). So, it is plausible that GhGT47A1 is also involved in GX biosynthesis in cotton fiber.

Both *GhGT47B1* (*Gh_A13G2031*) and *GhGT47B2* (*Gh_D13G2434*) have the ability to partially rescue the *fra8* mutant phenotypes

Gh_A13G2031 and *Gh_D13G2434* is a homoeologous pair in allotetraploid cotton, in which the former located on chromosome A13, and the latter located on its corresponding homoeologous chromosome D13. Their deduced proteins were mostly closely related to AtFRA8 which has been demonstrated to be involved in RES synthesis (Zhong et al. 2005; Brown et al. 2007). To test whether GhGT47B1 and GhGT47B2 are functionally equivalent to FRA8, we expressed *GhGT47B1* and *GhGT47B2* individually in the *fra8* mutant plants. The *fra8* mutant showed a thin and much shorter stature compared with the wild type (Fig. 5a). It seems that expression of both *GhGT47B1* and *GhGT47B2* individually in *fra8* could partially restore the plant size (Fig. 5a and b). Further cross section staining of stems showed that *fra8* mutation caused deformation of both xylem vessels and interfascicular fibers in stems (Fig. 5c and d). Though we did not find that the xylem vessels of the stems from these *GhGT47B1* or *GhGT47B2* complemented *fra8* lines became round and open, it was found that interfascicular fibers in these complemented lines showed normal shape as those of the wild type (Fig. 5c-f). These results show that both *GhGT47B1* and *GhGT47B2* are able to partially rescue the *fra8* phenotypes conferred by xylan deficiency, and indicate that both GhGT47B1 and GhGT47B2 are functional orthologs of FRA8 and probably involved in fiber SCW xylan synthesis, although biochemical activity assay is required to determine their exact roles.

Discussion

Cotton fibers employ a relatively conserved xylan biosynthesis machinery, nevertheless with many fiber-specific features

Mature cotton fiber is composed mainly of cellulose (> 90%) with a small amount of hemicellulose and lignin (Haigler et al. 2012; Kim and Ralph 2014). It is currently accepted that, when xylan arrives at the cell wall, it coats and crosslinks the cellulose microfibrils, and also makes hydrophobic pockets to facilitate the monolignols polymerization (Wierzbicki et al. 2019). These features of xylan render xylan a critical component of cell wall structure and composition. Xylan has been shown to be required for the normal assembly and mechanical strength of secondary walls. As researches performed in Arabidopsis showed that reduction in xylan content often causes a severe decrease in cellulose deposition, secondary cell wall thickening and mechanical strength, which suggests that xylan is highly correlated with cellulose synthesis (Brown et al. 2007). However, xylan biosynthesis, a largely unexplored area in cotton fiber, little is known about the genes involved in. Hence, a full understanding of the genes required for xylan synthesis will shed light on fiber secondary cell wall synthesis and enable us to genetically improve fibers with altered xylan.

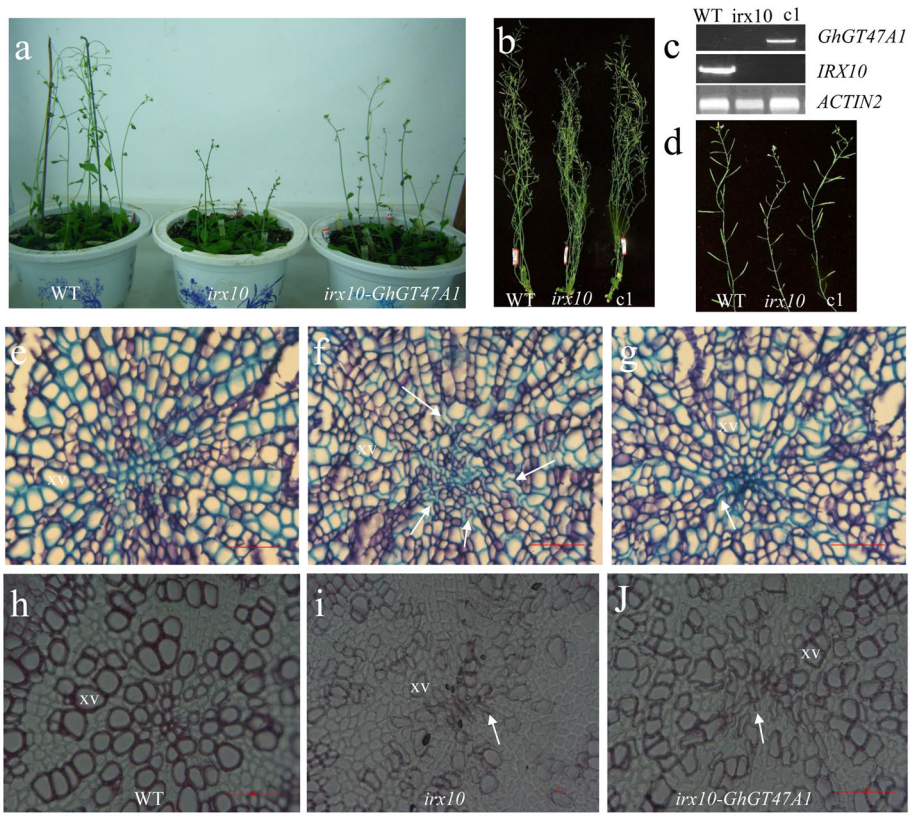

Fig. 4 Expression of *GhGT47A1* in Arabidopsis *irx10* mutant. **a** 4-week-old soil-grown wild-type, *irx10*, *GhGT47A1* complemented *irx10* plants. **b** 6-week-old soil-grown wild-type, *irx10*, *GhGT47A1* complemented *irx10* plants. C1 represents *GhGT47A1* complemented *irx10* plants. **c** RT-PCR analysis of expression of the *GhGT47A1*, *IRX10* in inflorescence stems of wild type, *irx10*, *GhGT47A1* complemented *irx10* plants. C1 represents *GhGT47A1* complemented *irx10* plants. The expression level of *ACTIN2* gene was used as an internal control. **d** Representative image of a branch carrying siliques from different genotypes. C1 represents *GhGT47A1* complemented *irx10* plants. **e-g** Toluidine-blue staining of cross sections of roots from wild type (E), *irx10* (F), and *GhGT47A1* complemented *irx10* plants(G). xv denotes xylem vessel, white arrows indicate collapsed xylem vessels. Scale bar = 50 μm. **h-j** Phloroglucinol staining of cross sections of roots from wild type (H), *irx10* (I), and *GhGT47A1* complemented *irx10* plants (J) . xv denotes xylem vessel, white arrows indicate collapsed xylem vessels. WT represents wild type, *irx10* represents *irx10* mutant, c1 represents *GhGT47A1* complemented *irx10* line. Scale bar = 50 μm

Extensive reports have shown that xylan is synthesized by evolutionarily conserved enzyme machinery in monocots and dicots (Hörnblad et al. 2013). Several studies have identified a number of genes involved in cell wall synthesis from different plant species via a bioinformatics method. For instances, 27 Arabidopsis sequences showing similarity to either the GT-A or the GT-B fold that have not yet been classified in the CZAY database were obtained by adopting a simple bioinformatics approach (Egelund et al. 2004). In addition, all the poplar wood-associated GTs have close homologs in Arabidopsis (Aspeborg et al. 2005). Sequences of annotated cell wall-related genes of Arabidopsis were used in a BLAST search to find the putative orthologs sequences in rice and maize (Sado et al. 2009). Also, Voxeur et al. (2012) identified 26 genes potentially encoding GT candidates for RG-II synthesis based on co-expression analysis and homologies with GTs of known functions (Voxeur et al. 2012). In this study, in order to gain a full overview of xylan-synthesis related genes in cotton fiber, we applied a similar bioinformatics approach which has already been successfully used to identify GhGT43A and GhGT43C in cotton, and found that the cotton genome nearly contains homologs of all reported genes involved in biosynthesis of GX backbone, RES and GX modification including methylation and acetylation in Arabidopsis. We provide further experimental evidence that *GhGT47A1* and *GhGT47B1/B2* are able to complement the respective Arabidopsis *irx10* and *fra8* mutants showing that they might have similar function as their Arabidopsis counterparts. Future work will be the experimental verification to provide direct evidence of their participation in fiber xylan formation.

Our results indicate that cotton fiber possesses a suit of biosynthetic genes similar to that of Arabidopsis which suggests that molecular machinery underlying xylan structure is largely conserved between cotton fiber and Arabidopsis stem, except that cotton fiber does not necessarily

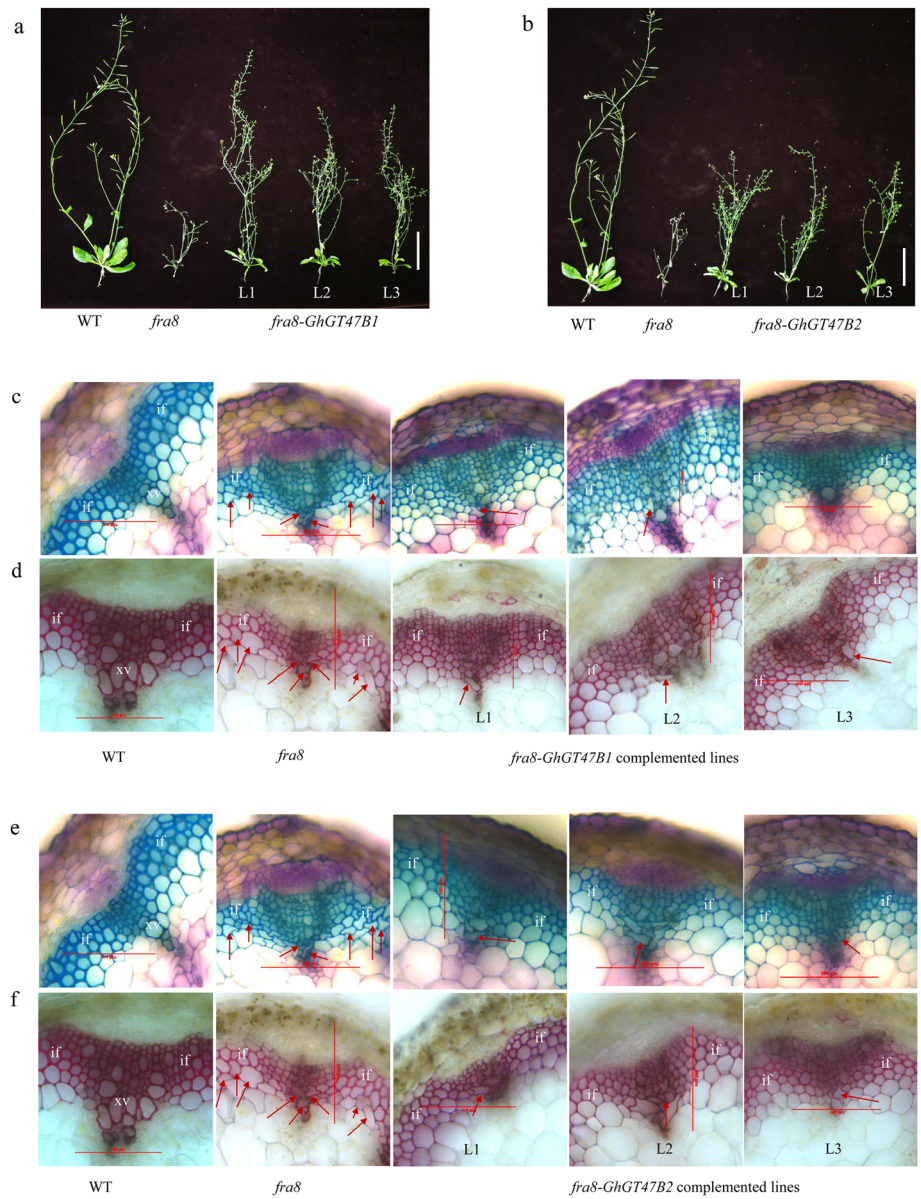

Fig. 5 Expression of *GhGT47B1* and *GhGT47B2* individually in Arabidopsis *fra8* mutant. **a** 6-week-old plants of wild type, *fra8* mutant and three independent *GhGT47B1*-complemented *fra8* lines. **b** 6-week-old plants of wild type, *fra8* mutant and three independent *GhGT47B2*-complemented *fra8* lines. **c** Toluidine blue staining of stem cross sections from 6-week-old wild type, *fra8* mutant and three independent *GhGT47B1*-complemented *fra8* lines. **d** Phloroglucinol staining of stem cross sections from 6-week-old wild type, *fra8* mutant and three independent *GhGT47B1*-complemented *fra8* lines. **e** Toluidine blue staining of stem cross sections from 6-week-old wild type, *fra8* mutant and three independent *GhGT47B2*-complemented *fra8* lines. **f** Phloroglucinol staining of stem cross sections from 6-week-old wild type, *fra8* mutant and three independent *GhGT47B2*-complemented *fra8* lines. WT represents wild type, *fra8* shows the *fra8* mutant, L1, L2, L3 denote *GhGT47B*-complemented *fra8* lines. Red arrows indicate deformed xylem vessels and interfascicular fibers. xv, xylem vessel, if, interfascicular fibers. Scale bar = 5 cm (a, b) and 100 μm (c, d, e, f) respectively

use homologs of IRX8, GUX3/4/5 and TBL28/30/34/35, but prefer to use orthologs of TBL41/43 with unknown function, and novel GT47 and DUF579-domain protein members. Furthermore, two GT61 family members encoding arabinosyltransferase seem to be involved in fiber xylan synthesis. These differences may reflect the specificity of xylan synthesis in cotton fiber.

It is known that proteins involved in xylan backbone elongation physically interact with one another, creating the xylan synthase complex (XSC). XSC composition differs among plant kingdom and affects xylan modification (Wierzbicki et al. 2019). Whether cotton fiber employs similar mechanism for xylan backbone elongation is unknown. Whether proteins involved in reducing end

synthesis or xylan modification also interact with each other or are part of the XSC still remains elusive. Our findings imply that in addition to a common biosynthetic mechanism of xylan synthesis, cotton fiber may also use specific groups of genes in secondary cell wall xylan formation. Since cotton fiber has specific cell wall components, it is rational to propose that some GTs or proteins are specifically involved in the generation of fiber-specific cell wall components. Our results also showed that in a specific gene family, some members may be the functional orthologs of xylan-synthesis related enzymes, some members may have divergent functions though they are phylogenetically close to each other. This functional conservation and divergence in xylan biosynthesis also existed in rice GT43 family members since complementation experiments revealed that some rice GT43 family members are able to rescue the *irx9* or *irx14* mutant phenotype but some are not. One possibility is that this group of members might have evolved to perform different catalytic activities.

Cotton fiber xylan biosynthesis genes can fall into major and minor function sets

Xylan biosynthetic genes can be separated into major and minor functionally redundant pairs, in which pair one gene is functionally dominant in Arabidopsis. *IRX9*, *IRX10*, *IRX14*, and *FRA8* make up the major function genes whereas their corresponding paralogs *IRX9L*, *IRX10L*, *IRX14L*, and *F8H* comprise minor function genes (Brown et al. 2007; Peña et al. 2007; Brown et al. 2009). Generally, the minor function genes were universally expressed in most tissues at a lower level, whereas the major function homologs were predominantly expressed in the stems and hypocotyls at a much higher level. Higher expression levels of major function genes might account for their functional dominance. The differences in their expression patterns might contribute to their differential involvement in GX biosynthesis in different secondary wall-forming cell types. Gene redundancy seems very common in GX biosynthesis in Arabidopsis which is likely to ensure the correct making of xylan. However, *IRX9*, *IRX10*, *IRX14* and *FRA8* play major roles in GX synthesis, and functional dominance of these genes may be due to their higher levels of expression or their higher activities of enzyme. In consistent with these previous results, our list of 36 strong candidates in Fig. 2a may play a major role while the list of 19 genes in Fig. 2b may play a major role in xylan synthesis. The difference in expression pattern may account for their differential roles in xylan formation.

Cotton fiber xylan biosynthesis is associated with cellulose and lignin synthesis

Qua1, *parvus-3*/*gatl1* and *irx8*/*gaut12* mutants all display alterations in biosynthesis of both xylan and pectin (Orfila et al. 2005; Brown et al. 2007; Lee et al. 2007). This complex pleiotropic phenotype makes it difficult to define the primary functions of the three genes. Lignin content may be directly or indirectly affected when xylan biosynthesis is disturbed. It has been reported that *irx7*, *irx8* (*GAUT12*) and *irx9* stems which are deficient in xylan synthesis may also have lignin defects (Zhong et al. 2005; Brown et al. 2007; Peña et al. 2007). Our work together with Brown et al. (2007) and Brown et al. (2009)'s results revealed that *irx10* stems had decreased xylan content as well as reduced amounts of lignin, which further supported that xylan and lignin synthesis are associated in some cell types and tissues. Acetyl- and methylglucuronic acid decorations of xylan affects its interaction with cellulose and lignin (Pauly and Ramírez 2018). The xylan-modification patterns required for the interaction with cellulose are different from the patterns required for the interaction with lignin. Co-expression analysis revealed that there possibly exist specific combinations of xylan-modification genes to establish the patterns required for each interaction. For instance, GXM1 (*Gh_A01G1163*, *Gh_D13G1237*), GUX1 (*Gh_D11G0603*, *Gh_A11G0519*) and TBL3 (*Gh_D13G0974*, *Gh_D04G1278* and *Gh_A04G0787*) and a new TBL-like gene (*Gh_D11G3388*) are likely to set up a xylan-modification pattern suitable for lignin interaction, as these genes were co-expressed with lignin biosynthesis genes (F5H (*Gh_D11G1805*), laccase (*Gh_D04G1243*); Additional file 9: Table S1).

Complementation of Arabidopsis GT mutants with corresponding cotton GTs could provide an alternative method to unravel the functional roles of cotton GTs (Zhou et al. 2006; Lee et al. 2009b). Since cotton transformation is a tedious and time consuming work, we took advantage of the Arabidopsis mutants deficient in xylan formation to clarify the functions of cotton homologs. Whether mutation of *IRX10* or *FRA8*, a generally mutants in these genes have reduced xylan content often leading to a typical irx (irregular collapsed xylem vessels) phenotype (Zhong et al. 2005; Brown et al. 2009). Our results showed that three fiber-specific genes (*GhGT47A1*, *GhGT47B1* and *GhGT47B2*) were capable of partially rescuing the irx phenotype of respective mutant. Together with our previous results about *GhGT43A* and *GhGT43C* (Li et al. 2014), it is conceivable that other genes in the list might indeed function in fiber xylan biosynthesis. Enormous advances have been made in identifying individual genes involved in xylan synthesis. However, many questions about how individual proteins work together to make a functional xylan remain unanswered. Characterization of genes involved in xylan synthesis will help understand the complex process of fiber secondary cell wall formation. Our primary objective is to identify a few promising candidate genes involved in fiber xylan biosynthesis and gain a better

understanding of molecular basis of fiber secondary cell wall synthesis. Our list of candidate genes is apparently far from being complete. We cannot rule out the possibility that we have omitted some important genes involved in xylan biosynthesis. Identification of only the closest Arabidopsis homologs cannot define the specific genes that are responsible for the differences in the xylan structures between stem and fiber. Our increasing knowledge of xylan biosynthetic mechanisms will provide new opportunities to manipulate the structures of xylan, their relative abundance and their interactions with cellulose and lignin within the wall. Understanding gene function in xylan biosynthesis will undoubtedly create new tools for the breeding of cotton with enhanced productivity, quality and biotechnological potential.

Conclusion

In this study, we identify that a total of 55 genes probably involved in the xylan biosynthesis in cotton fiber. We further demonstrate that *GhGT47A1*, *GhGT47B1* and *GhGT47B2* are able to partially complement the GX biosynthesis defects caused by the *irx10* and *fra8* mutation respectively. These findings provide a solid basis to uncover the biosynthesis of xylan in cotton fiber.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s42397-020-00063-3

Additional file 1: Figure S1. Hierarchical clustering of expression profiles of 18 cotton GT43 family members in different cotton tissues.

Additional file 2: Figure S2. Hierarchical clustering of expression profiles of 60 cotton GT47 family members in different cotton tissues.

Additional file 3: Figure S3. Hierarchical clustering of expression profiles of 100 cotton TBL members in different cotton tissues.

Additional file 4: Figure S4. Hierarchical clustering of expression profiles of 9 cotton RWA genes in different cotton tissues.

Additional file 5: Figure S5. Hierarchical clustering of expression profiles of 32 cotton DUF579 family genes in different cotton tissues.

Additional file 6: Figure S6. Hierarchical clustering of expression profiles of 115 cotton GT8 family genes in different cotton tissues.

Additional file 7: Figure S7. Expression profiles of 13 putative fiber xylan synthesis-associated candidates in different cotton tissues.

Additional file 8: Figure S8. Cross sections of stems from wild type, *irx10*, and *GhGT47A1* complemented *irx10* plants.

Additional file 9: Table S1. Annotation of co-expression genes.

Additional file 10: Table S2. Protein domain search by pfam analysis.

Additional file 11: Table S3. Expression data of 55 candidates involved in fiber xylan biosynthesis.

Acknowledgments

Not applicable.

Authors' contributions

Xu WL conceived and designed the study; Chen F, Guo YJ, Chen L, Gan XL, Liu M performed the experiments and analyzed the data; Li J analyzed the data; Xu WL wrote the manuscript. All authors read and approved the final version of manuscript.

Funding

This work was supported by National Natural Science Foundation of China, China [grant numbers: 31671735; 31970516], Hubei provincial Natural Science Foundation, China [grant number: 2016CFA071], and self-determined research funds of Central China Normal University from the colleges' basic research and operation of Ministry of Education, China (CCNU18TS021).

Availability of data and materials

No other data related to this study is available at this time.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 31 March 2020 Accepted: 21 July 2020
Published online: 05 September 2020

References

Anders N, Wilkinson MD, Lovegrove A, et al. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc Natl Acad Sci U S A. 2012; 109(3):989–93. https://doi.org/10.1073/pnas.1115858109.

Aspeborg H, Schrader J, Coutinho PM, et al. Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid Aspen. Plant Physiol. 2005;137:983–97. https://doi.org/10.1104/pp.104.055087.

Avci U, Pattathil S, Singh B, et al. Cotton fiber cell walls of *Gossypium hirsutum* and *Gossypium barbadense* have differences related to loosely-bound xyloglucan. PLoS One. 2013;8(2):e56315. https://doi.org/10.1371/journal.pone.0056315.

Brown DM, Goubet F, Wong VW, et al. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J. 2007;52:1154–68. https://doi.org/10.1111/j.1365-313X.2007.03307.x.

Brown DM, Wightman R, Zhang Z, et al. Arabidopsis genes *IRREGULAR XYLEM*(*IRX15*) and *IRX15L* encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. Plant J. 2011; 66(3):401–13. https://doi.org/10.1111/j.1365-313X.2011.04501.x.

Brown DM, Zhang Z, Stephens E, et al. Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J. 2009;57:732–46. https://doi.org/10.1111/j.1365-313X.2008.03729.x.

Egelund J, Skjot M, Geshi N, et al. A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes. Plant Physiol. 2004;136:2609–20. https://doi.org/10.1104/pp.104.042978.

Fan L, Shi WJ, Hu WR, et al. Molecular and biochemical evidence for phenylpropanoid synthesis and presence of wall-linked phenolics in cotton fibers. J Integr Plant Biol. 2009;51(7):626–37. https://doi.org/10.1111/j.1744-7909.2009.00840.x.

Gao Y, He C, Zhang D, et al. Two trichome birefringence-like proteins mediate xylan acetylation, which is essential for leaf blight resistance in rice. Plant Physiol. 2017;173:470–81. https://doi.org/10.1104/pp.16.01618.

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.

Han LB, Li YB, Wang HY, et al. The dual functions of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. Plant Cell. 2013;25: 4421–38. https://doi.org/10.1105/tpc.113.116970.

Hörnblad E, Ulfstedt M, Ronne H, et al. Partial functional conservation of IRX10 homologs in *physcomitrella patens* and *Arabidopsis thaliana* indicates an evolutionary step contributing to vascular formation in land plants. BMC Plant Biol. 2013;13(1):3. https://doi.org/10.1186/1471-2229-13-3.

Huang JF, Chen F, Wu SY, et al. Cotton *GhMYB7* is predominantly expressed in developing fibers and regulates secondary cell wall biosynthesis in transgenic Arabidopsis. Sci China Life Sci. 2016;59(2):194–205. https://doi.org/10.1007/s11427-015-4991-4.

Jensen JK, Johnson NR, Wilkerson CG. *Arabidopsis thaliana* IRX10 and two related proteins from psyllium and *Physcomitrella patens* are xylan xylosyltransferases. Plant J. 2014;80(2):207–15. https://doi.org/10.1111/tpj.12641.

Jensen JK, Kim H, Cocuron JC, et al. The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in Arabidopsis. Plant J. 2011;66:387–400. https://doi.org/10.1111/j.1365-313X.2010.04475.x.

Kim H, Ralph J. A gel-state 2D-NMR method for plant cell wall profiling and analysis: a model study with the amorphous cellulose and xylan from ball-milled cotton linters. RSC Adv. 2014;4(15):7549–60. https://doi.org/10.1039/c3ra46338a.

Lee C, Teng Q, Huang W, et al. The poplar GT8E and GT8F glycosyltransferases are functional orthologs of Arabidopsis PARVUS involved in glucuronoxylan biosynthesis. Plant Cell Physiol. 2009b;50(11):1982–7. https://doi.org/10.1093/pcp/pcp131.

Lee C, Teng Q, Huang WL, et al. The F8H glycosyltransferase is a functional paralog of FRA8 involved in glucuronoxylan biosynthesis in Arabidopsis. Plant Cell Physiol. 2009a;50:812–27. https://doi.org/10.1093/pcp/pcp025.

Lee C, Teng Q, Zhong R, et al. The four *Arabidopsis REDUCED WALL ACETYLATION* genes are expressed in secondary wall-containing cells and required for the acetylation of xylan. Plant Cell Physiol. 2011;52(8):1289–301. https://doi.org/10.1093/pcp/pcr075.

Lee C, Teng Q, Zhong R, et al. Three Arabidopsis DUF579 domain-containing GXM proteins are methyltransferases catalyzing 4-*O*-methylation of glucuronic acid on xylan. Plant Cell Physiol. 2012;53(11):1934–49. https://doi.org/10.1093/pcp/pcs138.

Lee C, Zhong R, Richardson EA, et al. The *PARVUS* gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant Cell Physiol. 2007;48:1659–72. https://doi.org/10.1093/pcp/pcm155.

Lee J, Burns TH, Light GG, et al. Xyloglucan endotransglycosylase/hydrolasegenes in cotton and their role in fiber elongation. Planta. 2010;232:1191–205. https://doi.org/10.1007/s00425-010-1246-2.

Li A, Xia T, Xu W, et al. An integrative analysis of four CESA isoforms specific for fiber cellulose production between *Gossypium hirsutum* and *Gossypium barbadense*. Planta. 2013;237(6):1585–97. https://doi.org/10.1007/s00425-013-1868-2.

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

Li L, Huang J, Qin L, et al. Two cotton fiber-associated glycosyltransferases, GhGT43A1 and GhGT43C1, function in hemicellulose glucuronoxylan biosynthesis during plant development. Physiol Plantarum. 2014;152:367–79. https://doi.org/10.1111/ppl.12190.

Manabe Y, Nafisi M, Verhertbruggen Y, et al. Loss-of-function mutation of *REDUCED WALL ACETYLATION2* in Arabidopsis leads to reduced cell wall ACET YLATION and increased resistance to *Botrytis cinerea*. Plant Physiol. 2011;155(3):1068–78. https://doi.org/10.1104/pp.110.168989.

Mortimer JC, Miles GP, Brown DM, et al. Absence of branches from xylan in Arabidopsis *gux* mutants reveals potential for simplification of lignocellulosic biomass. Proc Natl Acad Sci U S A. 2010;107:17409–14. https://doi.org/10.1073/pnas.1005456107.

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.

Pauly M, Ramírez V. New insights into wall polysaccharide *O*-acetylation. Front Plant Sci. 2018;9:1210. https://doi.org/10.3389/fpls.2018.01210.

Pawar PM, Ratke C, Balasubramanian VK, et al. Downregulation of RWA genes in hybrid aspen affects xylan acetylation and wood saccharification. New Phytol. 2017;214:1491–505. https://doi.org/10.1111/nph.14489.

Peña MJ, Zhong R, Zhou GK, et al. Arabidopsis *irregular xylem8* and *irregular xylem9*: implications for the complexity of glucuronoxylan biosynthesis. Plant Cell. 2007;19:549–63. https://doi.org/10.1105/tpc.106.049320.

Rennie EA, Hansen SF, Baidoo EEK, et al. Three members of the Arabidopsis glycosyltransferase family 8 are xylan glucuronosyltransferases. Plant Physiol. 2012;159:1408–17. https://doi.org/10.1104/pp.112.200964.

Sado PE, Tessier D, Vasseur M, et al. Integrating genes and phenotype: a wheat–*Arabidopsis*–rice glycosyltransferase database for candidate gene analyses. Funct Integr Genomics. 2009;9(1):43–58. https://doi.org/10.1007/s10142-008-0100-0.

Scheller HV, Ulvskov P. Hemicelluloses. Annu Rev Plant Biol. 2010;61:263–89. https://doi.org/10.1146/annurev-arplant-042809-112315.

Schultink A, Naylor D, Dama M, et al. The role of the plant-specific ALTERED XYLOGLUCAN9 protein in Arabidopsis cell wall polysaccharide *O*-acetylation. Plant Physiol. 2015;167:1271–83. https://doi.org/10.1104/pp.114.256479.

Urbanowicz BR, Peña MJ, Moniz HA, et al. Two Arabidopsis proteins synthesize acetylated xylan *in vitro*. Plant J. 2014;80(2):197–206. https://doi.org/10.1111/tpj.12643.

Voxeur A, Andre A, Breton C, Lerouge P. Identification of putative Rhamnogalacturonan-II specific glycosyltransferases in Arabidopsis using a combination of bioinformatics approaches. PLoS One. 2012;7(12):e51129. https://doi.org/10.1371/journal.pone.0051129.

Wierzbicki MP, Christie N, Pinard D, et al. A systems genetics analysis in *Eucalyptus* reveals coordination of metabolic pathways associated with xylan modification in wood-forming tissues. New Phytol. 2019;223:1952–72. https://doi.org/10.1111/nph.15972.

Wu AM, Hörnblad E, Voxeur A, et al. Analysis of the Arabidopsis *IRX9/IRX9-L* and *IRX14/IRX14-L* pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. Plant Physiol. 2010;153:542–54. https://doi.org/10.1104/pp.110.154971.

Xu WL, Zhang DJ, Wu YF, et al. Cotton *PRP5* gene encoding a proline-rich protein is involved in fiber development. Plant Mol Biol. 2013;82:353–65. https://doi.org/10.1007/s11103-013-0066-8.

Zeng W, Lampugnani ER, Picard KL, et al. Asparagus IRX9, IRX10, and IRX14A are components of an active xylan backbone synthase complex that forms in the Golgi apparatus. Plant Physiol. 2016;171:93–109. https://doi.org/10.1104/pp.15.01919.

Zhang T, 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;33:531–7. https://doi.org/10.1038/nbt.3207.

Zhong R, Cui D, Ye ZH. Regiospecific acetylation of xylan is mediated by a group of DUF231-containing *O*-acetyltransferases. Plant Cell Physiol. 2017;58:2126–38. https://doi.org/10.1093/pcp/pcx147.

Zhong R, Peña MJ, Zhou GK, et al. Arabidopsis Fragile Fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell. 2005;17:3390–408. https://doi.org/10.1105/tpc.105.035501.

Zhou GK, Zhong R, Himmelsbach DS, et al. Molecular characterization of PoGT8D and PoGT43B, two secondary wall-associated glycosyltransferases in poplar. Plant Cell Physiol. 2007;48(5):689–99. https://doi.org/10.1093/pcp/pcm037.

Zhou GK, Zhong R, Richardson EA, et al. The poplar glycosyltransferase GT47C is functionally conserved with *Arabidopsis Fragile Fiber8*. Plant Cell Physiol. 2006;47(9):1229–40. https://doi.org/10.1093/pcp/pcj093.