

Comparative transcriptome analysis uncovers cell wall reorganization and repressed cell division during cotton fiber initiation

Wenyuan Liu

Linyi University

Yanjia Lv

Linyi University

Xiaoyue Li

Linyi University

Zongqin Feng

Linyi University

Lichen Wang (✉ lichenwang@126.com)

Linyi University

Research Article

Keywords: Lint fiber initiation, Transcriptome sequencing, Xuzhou142 fibreless mutant (Xu142fl), MML, EXPA

Posted Date: June 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-607693/v1>

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

[Read Full License](#)

Abstract

Background

Tetraploid cotton plants serve as prime natural fiber source for the textile industry. Although various omics studies have revealed molecular basis for fiber development, a better understanding of transcriptional regulation mechanism regulating lint fiber initiation is necessary to meet global natural fiber demand.

Conclusions

Here, we aimed to perform transcriptome sequencing to identify DEGs (differentially expressed genes) in ovules of the cotton variety Xu142 and its fibreless mutant *Xu142fl* during early lint fiber initiation period. Totally, 5516 DEGs including 1840 upregulated and 3676 downregulated were identified. GO enrichment analysis revealed that the downregulated DEGs mainly associated with biological processes such as transcription related biosynthesis and metabolism, organic cyclic compound biosynthesis and metabolism, photosynthesis, and plant cell wall organization, with molecular functions involving transcription related binding, organic cyclic compound binding, and dioxygenase activity, while the upregulated DEGs were associated with DNA replication and phospholipid biosynthetic related processes. Among the 490 DEGs annotated as transcription factor genes 86.5% were downregulated in the mutant including the *Malvaceae*-specific *MMLs*, expression patterns of which were confirmed during the central period of lint fiber initiation. Investigation of the 20 genes enriched in the cell wall organization revealed that 17 were EXPA coding genes. Overall, we suggest that lint fiber initiation is a complicated process involving cooperation of multiple transcription factor families, which might ultimately lead to the reorganization of the cell wall and terminated cell division of the differentiating fiber initials.

Background

Cotton plants serve as the largest natural fiber source for the global textile industry [1]. Mature cotton seeds are covered with adherent fuzz and spinnable lint fibers. Cotton fibers development can be classified into four overlapping stages: initiation, elongation, thickening of the secondary cell wall, and maturation [2]. Lint fibers starts initiation from – 3dpa and continue to 3 dpa [3]. The initiation of lint fiber involves multiple regulators such as the upstream transcription factors and the downstream expansins proteins [4–6].

The transcriptional mechanisms have been widely explored by various researchers, which established the model in cotton that fiber initiation mimics the hair trichome initiation in *Arabidopsis* involving the MBW complex consisting of the R2R3 MYB protein GL1, the bHLH protein GL3, and the WD-repeat containing protein TTG1, which controls the expression of the downstream HD-ZIP transcription factor gene *GL2* [7]. In cotton, the *GL1* homologous R2R3-MYB transcription factor genes *GhMYB25-like* and *GhMYB25* have been implicated in regulating fiber initiation and elongation respectively [8, 9], and renamed as *MYB-*

MIXTA-like 3 (MML3) and *MML7* respectively lately [10]. Previously, the *N1* gene (*GhMML3_A12*) in *N1 naked seed mutant (N1NSM)* and the *fi3* gene (*GhMML4_D12*) in *Xuzhou142 fibreless mutant (Xu142fl)* have been isolated through a map based cloning method respectively [11, 12]. Totally 10 MMLs in *G.raimondii* have been classified into one lineage as the *Malvaceae*-specific 9th subfamily R2R3-MYBs that regulates epidermal cell differentiation [12], different from the 15th subfamily which regulates leaf hair trichome development in *Arabidopsis* according to evolution analysis [13], and they contain a signature protein motif and are highly expressed during the lint fiber initiation period [10]. Other homologous transcription factor genes in the MBW pathway such as *GL3* homolog bHLH transcription factor gene *GhDEL65* [14] and *GL2* homologs *GaHOX1* and *GhHOX3* also contribute to lint fiber development [15, 16].

The down-stream biological events regulating cell wall reorganization and biosynthesis employ various proteins like expansins, sucrose synthases, and tubulins [17–19]. Expansins are the first identified cell-wall-loosening proteins [20], which function by weakening the noncovalent bonds between cell wall matrix polymers to promote slippage of cellulose microfibrils and cause cell wall relaxation and cell extension [21], and constitute a large multigene family of four groups: alpha-expansin (EXPA), beta-expansin (EXPB), expansin-like A and expansin-like B. EXPAs were firstly speculated to involve in fiber development since the isolation of two alpha-expansin cDNAs from the developing fiber of *G.hirsutum* [22], then two homologous fiber-specific EXPAs coding genes-*GbEXPA2* from the D subgenome and *GbEXPATR* from the A subgenome of *G.barbadense* were cloned and genetically characterized to participate in the process of lint fiber elongation, evidenced by the facts that silencing of *GbEXPA* in *G.hirsutum* resulted into shorter fibers with thicker cell walls while overexpressing *GbEXPATR* induced longer, finer, and stronger fibers coupled with significantly thinner cell walls [23]. It has been reported that overexpression of *GhEXPA8* significantly improves fiber length and micronaire value [24], while co-expression of *GhRDL1* and *GhEXPA1* led to longer fiber [25]. However, the correlation between expansins proteins and lint fiber initiation at the transcriptome level still need to be investigated.

Despite various studies strongly suggest the key role of *Xu142* and *Xu142fl* in investigating lint fiber initiation; however, transcriptional regulation mechanisms of *Xu142* and *Xu142fl* regulating fiber initiation, in particular, with a focus on the changing upstream transcription regulations and the downstream cell wall reorganization during early lint fiber initiation in cotton are poorly understood. In this study, we compared the whole genome transcription changes between *Xu142* and *Xu142fl* at early fiber initiation stages and identified 3676 downregulated and 1840 upregulated genes. Gene Ontology (GO) enrichment analysis revealed that the downregulated genes were mainly involved in the biological processes related to transcription, organic cyclic compound biosynthesis and metabolism, photosynthesis, response to chitin and plant cell wall organization, and had molecular functions such as transcription related binding, organic cyclic compound binding and dioxygenase activity, while the upregulated DEGs were enriched in DNA replication and phospholipid biosynthesis related processes. Classification of the transcription factor genes revealed that various transcription factor families involved in lint fiber initiation, including the members belonging to the MBW complex. Examination of the cell wall organization related genes revealed 17 EXPA coding genes, one pectinesterase (PE) coding gene, and two

COBL10 coding genes, which may contribute to cell loosening and reorganization during lint fiber initiation. Overall, these findings provide preliminary data which suggest importance of introducing the essential transcription factor genes or *EXPAs* for improving lint fiber initiation rate.

Results

Overview of the comparative transcriptome sequencing using ovules of *Xu142fl* and *Xu142* at early fiber initiation stage

In order to study *Xu142fl* and *Xu142* in the context of fiber initiation, we first performed transcriptome sequencing by mixing ovules at -3 and -1 DPA of *Xu142fl* and *Xu142* respectively, before obvious fiber initials could be observed from the epidermis of the wild type cotton seeds under optical microscope, by setting 3 biological replicates for each variety. After removing adapter contamination and low quality tags, a total of 66.12-72.19 million clean reads were generated from each library, with clean read ratios between 92.51%-93.74%, and ~ 95% of the clean reads can be mapped to cotton TM-1 genome (Table 1). Besides this, 49381 novel transcripts were identified, including 36093 candidate protein coding and 13288 noncoding transcripts and 5604 novel genes were predicted.

After calculating the expression levels of each gene in each sample; Pearson correlation (R^2) was calculated based on the whole gene expression profile between each sample pair among the total 6 samples. The result showed that the correlations between biological repeats were 0.966-0.993 for the mutant pairs, and 0.971-0.996 for the wild type pairs, but lower (0.921-0.960) between mutant and wild type pairs (Additional file 1 Figure 1). The results indicated high uniformity between biological repeats. Finally, 5516 DEGs including 1840 upregulated and 3676 downregulated genes were identified in the mutant compared with the wild type.

Table 1. Overview of the data quality and genome mapping of the transcriptome sequencing of *Xu142* and *Xu142fl*.

Ovule Sample	Total Raw Reads (M)	Total Clean Reads (M)	Clean Reads Q20(%)	Clean Reads Ratio (%)	Total Mapped (%)	Uniquely Mapped (%)
Xu142-1	70.82	66.39	98.56	93.74	95.89	76.20
Xu142-2	75.39	70.42	98.59	93.41	95.55	75.11
Xu142-3	73.31	68.7	98.59	93.71	95.95	75.99
<i>Xu142fl-1</i>	70.61	66.12	98.53	93.64	95.75	74.58
<i>Xu142fl-2</i>	75.72	70.78	98.55	93.48	95.85	75.43
<i>Xu142fl-3</i>	78.04	72.19	98.6	92.51	95.87	75.52

Note: M means megabase.

Complicated upstream transcription and downstream biosynthesis and metabolism events occurred during early lint fiber initiation

To investigate the biological processes and functions attributed to the DEGs, GO enrichment analysis of the downregulated and upregulated DEGs were conducted respectively. The results showed that the downregulated genes were enriched in 15 level 3 biological process (BP) terms including transcription (262 genes), RNA (288 genes) and nucleic acid metabolism (339 genes), nucleobase-containing compound biosynthesis (267 genes) and metabolism (358 genes), heterocycle biosynthesis (275 genes) and metabolism (377 genes), aromatic compound biosynthesis (273 genes) and metabolism (383 genes) and organic cyclic compound biosynthesis (276 genes) and metabolism (380 genes), photosynthesis (42 genes) and light harvesting (15 genes), respond to chitin (7 genes), and plant cell wall organization (20 genes) (Fig. 1a; Additional file 2 Table S1), with molecular functions (MF) including DNA binding transcription factor activity (205 genes), transcription regulator activity (208 genes) and dioxygenase activity (32 genes), and DNA (404 genes), nucleic acid (585 genes), organic cyclic compound and heterocyclic compound binding activities (1008 genes) (Fig. 1b; Additional file 3 Table S2). These results demonstrated that the essential transcriptional regulations in *Xu142fl* were impaired and led to the aborted lint fiber initiation.

Compared with the significantly enriched GO terms for the downregulated DEGs, the enriched GO terms for the upregulated DEGs were fewer on the BP level, and no GO terms were found on the MF level. The significantly enriched level 3 terms was DNA replication (21 genes), DNA-dependent DNA replication (11 genes), DNA replication initiation (8 genes) and phospholipid biosynthesis (19 genes) (Fig. 1c; Additional file 4 Table S3). Overall, these findings suggest that DNA replication and phospholipid biosynthesis in the process of cell mitosis division, were inhibited in the fiber initials of wild type during lint fiber initiation.

Transcription factor expression dynamics during early fiber initiation

Next, to characterize the complicated transcription regulations during lint fiber initiation, the TF coding genes were firstly predicted and then filtered to obtain the differentially expressed TF genes closely related to lint fiber initiation (only TF family containing more than two genes were considered), and finally 490 genes belonging to 26 TF families were identified, consisting of 424 downregulated genes and 66 upregulated genes. Moreover, most TF families consisted more downregulated genes (57.14%-100%), except for LOB domain-containing and MADS domain-containing families which contained more up-regulated genes (76.92% and 71.43% respectively) (Table 2).

As shown in Table 2, among the down-regulated TF families containing members more than 20, the biggest TF family was AP2-EREBP, which contained 123 DEGs, followed by MYB (54 DEGs), WRKY (41 DEGs), NAC(33 DEGs), C2C2-Dof (33 DEGs), bHLH (29 DEGs) and GRAS (23 DEGs). Subsequently, we chose six TF families including MYB, bHLH, NAC, C2C2 Dof, GRAS, and WRKY containing members ranging from 23-54 to plot expression heat maps. The results showed that 6 *Malvaceae*-specific *MML* homologs (Wu et al., 2018) including two *MML3*, one *MML4*, two *MML8* and one *MML9* were all downregulated. The *MML3* on chromosome D12 chromosome (Gh_D12G1628) had higher expression level in the wild type and were much more downregulated in the mutant than the one on A12 chromosome (Gh_A12G1503) (Fig. 2a). Although most members were downregulated in the identified TF families, subclasses of members were still upregulated for 5 TF families, including the MYB family which contains one *RL6*, 3 *MYB44* and 4 *GAM1* (Fig. 3a), bHLH family which contains 5 members including *BHLH82*, *BHLH130*, *BEE3* and two novel genes (Fig. 2b), NAC family which contains 6 *NAC* genes, three of which annotated as *NAC100* had dominant expression in both varieties (Fig. 2c), the GRAS family which contains 2 members (Fig. 2d), and WRKY family which contains 7 members, with the genes annotated *WRKY48* and *WRKY65* had higher expression levels while expression of the other five genes were lower in both varieties (Fig. 2e), except for the C2C2 Dof family genes which were all down-regulated (Fig. 2f).

Table 2 Statistics of the differentially expressed transcription factor genes.

TF Family	Total DEGs	Upreg	Downreg	Upreg (%)	Downreg (%)
ABI3VP1	7	1	6	14.29	85.71
AP2-EREBP	123	4	119	3.25	96.75
bHLH	29	9	20	31.03	68.97
C2C2-Dof	33	0	33	0	100
C2C2-GATA	6	0	6	0	100
C2H2	17	2	15	11.76	88.24
C3H	11	0	11	0	100
FAR1	5	1	4	20	80
G2-like	7	2	5	28.57	71.43
GRAS	23	2	21	8.7	91.3
GRF	4	0	4	0	100
HSF	12	1	11	8.33	91.67
LIM	3	1	2	33.33	66.67
LOB	13	10	3	76.92	23.08
MADS	7	5	2	71.43	28.57
mTERF	7	1	6	14.29	85.71
MYB	54	8	46	14.81	85.19
NAC	33	6	27	18.18	81.82
OFP	11	0	11	0	100
PLATZ	4	0	4	0	100
SBP	4	1	3	25	75
TCP	5	0	5	0	100
Tify	14	0	14	0	100
Trihelix	7	3	4	42.86	57.14
WRKY	41	7	34	17.07	82.93
zf-HD	10	2	8	20	80
Total	490	66	424		

Note: Upreg means the upregulated genes; Downreg means the downregulated genes.

Main *GhMMLs* contributed to lint fiber initiation

The 9th subfamily R2R3-MYB transcription factors MMLs is considered as *Malvaceae*-specific through evolutionary analysis [12] and among the 10 *GhMMLs* from *GhMML1* to *GhMML10*, *GhMML3* and *GhMML4* had been demonstrated responsible for fuzz and lint fiber initiation respectively [11, 12]. Our TF classification had revealed that *GhMML3*, *GhMML4*, *GhMML8*, and *GhMML9* were involved in lint fiber initiation, to confirm that, we conducted RT-PCR of all ten *GhMMLs* in *n2NSM* and *Xu142fl* during early lint fiber initiation period from -1 DPA to 1 DPA, given the fact that *n2NSM* and *Xu142fl* are all naked seed mutants and the only difference is whether the lint fiber initiates or not [12]. The results showed that 8 *MMLs* can be detected except *GhMML8* and *GhMML9* (Fig. 3), and 4 were down-regulated including *GhMML1*, *GhMML3*, *GhMML4* and *GhMML7* in *Xu142fl* compared with *n2NSM*. Further investigation of their expression patterns by qRT-PCR confirmed that *GhMML3*, *GhMML4* and *GhMML7* were significantly down-regulated in *Xu142fl* compared with *n2NSM* at all three time points while no obvious expression differences were observed for *GhMML1* (Fig. 4). We also found that expression levels of *GhMML3* and *GhMML4* in *n2NSM* were decreased, while that of *GhMML7* were increased from -1 DPA to 1 DPA, implying different mechanisms between *GhMML7*, and *GhMML3* and *GhMML4* (Fig. 4).

Expansins enriched in cell wall organization may contributed to early lint fiber initiation

Cell wall reorganization is an essential event during fiber development involving multiple enzymes and wall proteins [5]. Here, we had a detailed investigation of the GO term-plant cell wall organization which contains 20 genes, and the results showed that 17 were EXPA encoding genes, including 4 *EXPA1*, 1 *EXPA2*, 6 *EXPA4*, 3 *EXPA8* and 3 *EXPA15*, one was Pectinesterase (PE) encoding gene, and 2 were COBL10 (protein transport protein SEC61 subunit alpha) encoding genes, and they were all downregulated in *Xu142fl* comparing with that in the wild type (Table 3). It was worth noting that the *EXPA2* gene, one *EXPA4* encoding gene (*Gh_A10G2323*) and the two COBL10 encoding genes had very low expression levels in the wild type and hardly detectable in the mutant, implying that they might not be the dominant genes for early lint fiber initiation. Of special note, data indicated that EXPAs might be the most important cell wall proteins for the early fiber cell initiation.

Table 3. Expression differences of the cell wall organization related genes.

Gene ID	Gene Annotation	Xu142_RPKM	<i>Xu142fl</i> _RPKM	log2FoldChange (<i>Xu142fl</i> /Xu142)
Gh_A06G0018	EXPA1	8.39	3.13	-1.41
Gh_D05G1754	EXPA1	35.18	13.74	-1.36
Gh_A05G1576	EXPA1	27.35	11.60	-1.25
Gh_D12G1759	EXPA1	19.72	9.50	-1.06
Gh_A13G0672	EXPA15	16.72	4.63	-1.86
Gh_D13G0786	EXPA15	20.71	6.30	-1.71
Gh_A03G0885	EXPA15	48.83	21.23	-1.21
Gh_D10G1145	EXPA2	0.69	0.00	-6.20
Gh_A10G2323	EXPA4	0.32	0.00	-5.12
Gh_A07G0902	EXPA4	30.35	9.97	-1.59
Gh_D09G1463	EXPA4	125.90	43.37	-1.54
Gh_A04G0707	EXPA4	27.78	11.04	-1.35
Gh_A09G1454	EXPA4	63.47	27.08	-1.23
Gh_D07G0974	EXPA4	28.64	13.56	-1.01
Gh_A05G3493	EXPA8	17.43	2.33	-2.93
Gh_D04G1924	EXPA8	3.52	1.60	-1.17
Gh_D10G1861	EXPA8	11.23	5.25	-1.11
Gh_A10G1502	Pectinesterase	26.62	10.09	-1.40
Gh_D06G1606	COBL10	0.27	0.02	-3.70
Gh_A06G1281	COBL10	0.78	0.09	-3.11

Note: *COBL10* encodes a protein transport protein SEC61 subunit alpha

Discussion

Lint fiber initiation is a complicated **morphogenesis** process involving complex metabolite biosynthesis and metabolism

In this study, through GO analysis of the downregulated and upregulated DEGs, multiple enriched biological processes were identified which were associated lint fiber initiation (Fig. 1). Based on our findings, we speculated that the downregulated DEGs mainly function as positive regulatory factors for lint fiber initiation, while the upregulated DEGs mainly function to inhibit lint fiber initiation.

For the downregulated DEGs, more than 200 genes were associated DNA-templated transcription (Fig. 1a), which have transcription associated molecular functions as DNA binding transcription factor activity (Fig. 1b). These biological processes related genes may construct the upstream regulation networks during lint fiber initiation, and the downstream regulation involves complicated metabolic cascades associated with organic compound biosynthesis and metabolism including nucleobase-containing compound, organic cyclic compound, heterocyclic compound and aromatic compound such as flavonoid, sugar and Phytohormones biosynthesis and metabolism (Fig. 1a,b;). Several *NCEDs* (data not shown) which encode the rate-limiting dioxygenases controlling ABA biosynthesis [26], were found in the GO term of dioxygenase (Fig. 1b), implying that ABA might be a positive regulator during lint fiber initiation. This can be supported by the gradually accumulation of endogenous ABA content during the fiber cell initiation and elongation stages [27] although ABA was considered as a negative regulator of fiber initiation [28]. Chitin oligosaccharides can induce various defense responses in a wide range of plant cells including both monocots and dicots [29]. However, do defense responses related to fiber initiation is still not clear. Immature fiber (im) mutant with thinner fiber cell wall compared to the isogenic wild type (TM-1) with fiber of normal thickness revealed that the mutant had lower net photosynthesis, because of the lower chlorophyll content per unit leaf area due to less chlorophyll a levels than the wild type [30], however how the photosynthesis related genes were changed in *Xu142fl* need further investigation (Fig. 1a).

The upregulated DEGs were mainly enriched in DNA replication and phospholipid biosynthetic processes, which were closely related to cell division. This is reasonable because fiber cells are unable to undergo cell division during fiber development. Before fiber cell differentiation, the ovular epidermal cells are closely packed, cuboidal, and rich in cytoplasm containing a large nucleus, which represents a status of rapid cell division [31]. The early development of fibers consists of two intergrading steps-spherical expansion above the ovular epidermis and elongation [32]. The morphological differentiation of a fiber occurs when an epidermal cell balloons above the epidermal surface, followed by transition to elongation phase and stopping division [33]. Collectively, cell division should be stopped to initiate fiber development in the wild type while rapid cell division continues which inhibit cell differentiation into fiber cells in *Xu142fl* mutant.

Complicated transcription regulation during lint fiber initiation

In this study, through TF annotation and expression pattern profiling, many down-regulated TF families were identified, including known fiber development related MYBs , bHLHs and TCPs [6]. Other TF families such as NAC, WRKY, GRAS and Dof identified here were also detected in another study exploring fiber elongation related pathways by transcriptome analysis of a short fiber mutant and a wild type [34], suggesting that some TF family genes might have dual role both in fiber initiation and elongation. For example, fiber cell expansion and elongation can be mediated by a homeodomain leucine zipper gene, *GhHD-1*, through a WRKY transcription factor by regulating the levels of ethylene and reactive oxidation species (ROS) [35]. Because some lint fiber initiation related TFs which were genetically characterized as

positive regulators for fiber development were firstly found downregulated in fiber related mutants, including R2R3 MYB proteins [8, 9, 36, 37], HD-ZIP proteins [15, 16] and bHLH proteins [14], so the other downregulated TF family genes found in this study deserve further investigation of their roles in regulating lint fiber initiation.

Besides this, two class of TF families involved more upregulated DEGs than the downregulate DEGs (Table 2), including LOB (Lateral Organ Boundaries) family which are essential in the regulation of plant lateral organ development [38] and MADS which regulate floral organ differentiation and development [39], and small group of genes in other TF families were also up-regulated (Fig. 2, Table 2), which might be negative regulators of fiber initiation and contribute to the lintless ovules of *Xu142fl* due to their upregulation.

Dynamic expression of *Malvaceae*-specific *MMLs* during early lint fiber initiation

Among the 10 pair of *Malvaceae*-specific *MYB MIXTA-like* (*GhMML*) homoeologs at least one copy on one sub-genome in allotetraploid cotton were predominantly expressed during fiber initiation in the wild type TM-1 [10], and *MML3-A12* and *MML4-D12* were demonstrated contributing to the fuzz and lint fiber initiation respectively [11, 12], implying the subfunctionalization of *MML* genes. Combined our transcriptome analysis and RT-PCR analysis, *MML3*, *MML4* and *MML7* should be the *MMLs* most related to lint fiber initiation (Fig. 2a, Fig. 3, 4). However, it was contrary in that *GhMML3-D12* (Gh_D12G1628) and *GhMML4-A12* (Gh_A12G1504) might be more important for lint fiber initiation, because *GhMML3-D12* had higher expression level in the wild type and was more downregulated in the mutant than *GhMML3-A12*, while *GhMML4-D12* was not detected as DEG in our study (Fig. 2a).

Different from *GhMML3* and *GhMML4* which were downregulated in the fuzzless-linted mutant *n2NSM* from -1 DPA to 1 DPA, *GhMML7* was upregulated and had highest expression level in ovules at 1 DPA (Fig. 4). This phenomenon indicated that, except for the role in fiber initiation, *GhMML7* may also involve in fiber elongation, which can be assisted by the evidences that *GhMML7/GhMYB25* expressed in the epidermis of ovules, developing fiber initials and fibers, and *GhMML7/GhMYB25*-silenced cotton produced shorter fiber, while overexpression of *GhMYB25* promoted fiber initiation [8].

Cell wall organization initiates from the very beginning during lint fiber initiation

Previously, Stage-specific developmental markers such as EXPAs, xyloglucan endo-transglycosylases (XETs) and PEs have been reported to regulate fiber cell expansion in cotton [4, 5, 40]. Here, we also identified many EXPA encoding genes and a PE encoding gene, indicated that EXPAs are more important during fiber initiation and elongation. Taken together, upstream regulators of EXPAs will further facilitate our understanding underlying mechanisms of lint fiber initiation and elongation.

Conclusion

In this study, we compared the whole genome transcription changes between Xu142 and *Xu142fl* at early fiber initiation stages and identified 3676 downregulated and 1840 upregulated genes. Gene Ontology (GO) enrichment analysis revealed that the downregulated genes were mainly involved in the biological processes related to transcription, organic cyclic compound biosynthesis and metabolism, photosynthesis, response to chitin and plant cell wall organization, and had molecular functions such as transcription related binding, organic cyclic compound binding and dioxygenase activity, while the upregulated DEGs were enriched in DNA replication and phospholipid biosynthesis related processes. Classification of the transcription factor genes revealed that various transcription factor families involved in lint fiber initiation, including the members belonging to the MBW complex. Examination of the cell wall organization related genes revealed 17 EXPA coding genes, one pectinesterase (PE) coding gene, and two COBL10 coding genes, which may contribute to cell loosening and reorganization during lint fiber initiation. Overall, these findings provide preliminary data which suggest importance of introducing the essential transcription factor genes or *EXPAs* for improving lint fiber initiation rate.

In conclusion, this study provides new information for lint fiber initiation mechanism, which might involve dynamic expression of multiple types of TF family genes, which mediate the complicated downstream organic compound biosynthesis and metabolism resulting into the termination of cell division and cell wall reorganization of the expanding fiber cells during lint fiber initiation (Fig. 5).

Methods

Plant materials

One cotton variety *Gossypium hirsutum* cv. Xu142 and two fibreless cotton mutants *Xu142fl* and *n2NSM* were selected in this study. The seeds of these lines were retrieved from National Medium-term Gene Bank of Cotton in China and National cotton germplasm resources platform. The seeds were grown in the experimental field of Linyi University in April 14, 2018. Ovules at -3, -1, 0 and 1 DPA were collected between July and August during the flowering peak, frozen immediately in liquid nitrogen and stored at -80°C.

Transcriptome sequencing and bioinformatics analysis

High-quality RNA extraction was performed from the frozen ovules tissues collected at -3 DPA and -1 DPA of Xu142 and *Xu142fl* respectively as previously reported [41]. Subsequent cDNA libraries were constructed and sequenced with BGI-SEQ500 at the Beijing Genomics Institute (BGI, Shenzhen, China).

The raw reads were filtered first to get clean data, and then aligned to the genome of *G. hirsutum* L. (<https://cottonfgd.org/about/download/assembly/genome.Ghir.NAU.fa.gz>) to reconstruct transcripts

through String Tie (<http://ccb.jhu.edu/software/stringtie>, v1.0.4), identify known genes by Bowtie2 (<http://bowtie-bio.sourceforge.net/Bowtie2>, v2.2.5) [42], and predict new transcripts using HISAT2 (<http://www.ccb.jhu.edu/software/hisat>, v2.0.4) [43].

Expression levels of genes and transcripts were calculated using RSEM [44], and expression corrections were calculated by Cor package of R (v3.6.2). DEGs (filtered by fold change ≥ 2 , Q-value ≤ 0.001) between Xu142 and *Xu142fl* were identified through DEGseq [45], after normalizing raw reads of each gene as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). GO enriched terms were determined by Q-value ≤ 0.001 using the phyper package of R.

The transcription factor (TF) coding genes were predicted by using Getorf (<http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html>) to get the ORF of all unigenes first, and then aligned to TF protein structure domain using hmmsearch (<http://hmmer.org>, v3.0) to annotate TF through the property of specific TF family described in PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>, v5.0). Heatmaps of TF proteins were generated by Genesis v1.7.6 [46].

RT-PCR and RT-qPCR analyses

RT-PCR and qRT-PCR were used to evaluate expression levels of *GhMMLs*. Total RNA from ovules at -1, 0 and 1 DPA of *n2NSM* and *Xu142fl* was extracted as previously reported [41]. Subsequent cDNA was synthesized using a Thermo Scientific RevertAid First Strand cDNA Synthesis Kit according to the manufacturer's instruction. Total 20- μ l-reaction volume was applied for RT-PCR analysis to evaluate expression levels of *GhMMLs*. After this, PCR reaction mixture was subjected to 95°C denaturation for 3 min, then 29 cycles of amplification of the endogenous reference gene *GhUbq7* or 38 cycles for *GhMMLs* at 95°C for 30 sec, 55°C annealing for 30 sec and 72°C extension for 30 sec, plus a final extension at 72°C for 5 min. Quantitative real-time PCR (qRT-PCR) was carried out using Hieff qPCR SYBR Green Master Mix (No Rox) (Yeason, Shanghai, China). The calculation of the relative expression levels of each gene and statistical analysis were determined as previously reported [47]. Primers were designed by Primer 5.0 and synthesized commercially (Genscript Bioscience, Nanjing, China). The sequences of all the primers sequences are listed in Additional file 5 Table 4.

Declarations

Acknowledgements

We thank Kashif Rafiq for critiquing the manuscript. We thank National Medium-term Gene Bank of Cotton in China and National cotton germplasm resources platform for providing the cotton seeds used in this study.

Authors' contributions

L. W.: project design, transcriptome data mining, writing the manuscript. Y. L., X. L. and Z. F.: collecting the ovules and extracting the total RNA. W. L: experiments shown in Figure 3 and Figure 4. All authors read and approved the final manuscript.

Funding

The authors are sponsored by State Key Laboratory of Cotton Biology Open Fund (Grant No. CB2018A20), and Linyi University PhD Fund for Scientific Start-up (LYDX2018BS029).

Availability of data and materials

Raw data for the transcriptomes are available on the GEO platform as series GSE176384.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

No competing interests declared.

References

1. Hu Q, Xiao S, Guan Q, Tu L, Sheng F, Du X, Zhang X: **The laccase gene GhLac1 modulates fiber initiation and elongation by coordinating jasmonic acid and flavonoid metabolism.** *The Crop Journal* 2020, **8**(4):522–533.
2. Qin Y-M, Zhu Y-X: **How cotton fibers elongate: a tale of linear cell-growth mode.** *Current opinion in plant biology* 2011, **14**(1):106–111.

3. Graves D, Stewart J: **Chronology of the differentiation of cotton (*Gossypium hirsutum* L.) fiber cells.** *Planta* 1988, **175**(2):254–258.
4. Lee JJ, Woodward AW, Chen ZJ: **Gene expression changes and early events in cotton fibre development.** *Annals of botany* 2007, **100**(7):1391–1401.
5. Ruan YL: **Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: insights from the single-celled cotton fibre.** *Functional plant biology: FPB* 2007, **34**(1):1–10.
6. Wang L, Wang G, Long L, Altunok S, Feng Z, Wang D, Khawar KM, Mujtaba M: **Understanding the role of phytohormones in cotton fiber development through omic approaches; recent advances and future directions.** *International journal of biological macromolecules* 2020, **163**:1301–1313.
7. Szymanski DB, Lloyd AM, Marks MD: **Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*.** *Trends in plant science* 2000, **5**(5):214–219.
8. Machado A, Wu Y, Yang Y, Llewellyn DJ, Dennis ES: **The MYB transcription factor GhMYB25 regulates early fibre and trichome development.** *The Plant journal: for cell and molecular biology* 2009, **59**(1):52–62.
9. Walford SA, Wu Y, Llewellyn DJ, Dennis ES: **GhMYB25-like: a key factor in early cotton fibre development.** *The Plant journal: for cell and molecular biology* 2011, **65**(5):785–797.
10. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM *et al*: **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement.** *Nature biotechnology* 2015, **33**(5):531–537.
11. Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W *et al*: **Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development.** *The New phytologist* 2016, **210**(4):1298–1310.
12. Wu H, Tian Y, Wan Q, Fang L, Guan X, Chen J, Hu Y, Ye W, Zhang H, Guo W *et al*: **Genetics and evolution of MIXTA genes regulating cotton lint fiber development.** *The New phytologist* 2018, **217**(2):883–895.
13. Stracke R, Werber M, Weisshaar B: **The R2R3-MYB gene family in *Arabidopsis thaliana*.** *Current opinion in plant biology* 2001, **4**(5):447–456.
14. Shangguan XX, Yang CQ, Zhang XF, Wang LJ: **Functional characterization of a basic helix-loop-helix (bHLH) transcription factor GhDEL65 from cotton (*Gossypium hirsutum*).** *Physiologia plantarum* 2016, **158**(2):200–212.
15. Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY: **The HD-Zip IV gene GaHOX1 from cotton is a functional homologue of the *Arabidopsis* GLABRA2.** *Physiologia plantarum* 2008, **134**(1):174–182.
16. Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ *et al*: **Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3.** *Nature communications* 2014, **5**:5519.
17. Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, Zhu YX: **Transcriptome profiling, molecular biological, and physiological studies reveal a major role for**

- ethylene in cotton fiber cell elongation.** *The Plant cell* 2006, **18**(3):651–664.
18. Lv L, Zuo D, Wang X, Cheng H, Ma Z: **Genome-wide identification of expansin gene family reveals expansin genes are involved in fibre cells growth in cotton.** 2020.
 19. Mukhtar A, Ali SA, Sidra A, Ayesha L, Ud DS, Ma F, Rao AQ, Bilal SM, Tayyab H, Wang X: **Sucrose synthase genes: a way forward for cotton fiber improvement.** *Biologia* 2018, **73**:1–11.
 20. McQueen-Mason, S.: **Two endogenous proteins that induce cell wall extension in plants.** *The Plant cell* 1992, **4**(11):1425–1433.
 21. Sampedro J, Cosgrove DJ: **The expansin superfamily.** *Genome biology* 2005, **6**(12):242.
 22. Harmer S, Orford S, Timmis J: **Characterisation of six alpha-expansin genes in Gossypium hirsutum (upland cotton).** *Molecular Genetics & Genomics* 2002, **268**(1):1–9.
 23. Li Y, Tu L, Pettolino FA, Ji S, Hao J, Yuan D, Deng F, Tan J, Hu H, Wang Q *et al.* **GbEXPATR, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring.** *Plant biotechnology journal* 2016, **14**(3):951–963.
 24. Bajwa KS, Shahid AA, Rao AQ, Bashir A, Aftab A, Husnain T: **Stable transformation and expression of GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton.** *Frontiers in plant science* 2015, **6**:838.
 25. Xu B, Gou JY, Li FG, Shangguan XX, Zhao B, Yang CQ, Wang LJ, Yuan S, Liu CJ, Chen XY: **A cotton BURP domain protein interacts with alpha-expansin and their co-expression promotes plant growth and fruit production.** *Molecular plant* 2013, **6**(3):945–958.
 26. Matilla AJ, Carrillo-Barral N, Rodríguez-Gacio M: **An Update on the Role of NCED and CYP707A ABA Metabolism Genes in Seed Dormancy Induction and the Response to After-Ripening and Nitrate.** *Journal of Plant Growth Regulation* 2015, **34**(2):274–293.
 27. Davis LA, Addicott FT: **Abscisic Acid: Correlations with Abscission and with Development in the Cotton Fruit.** *Plant physiology* 1972, **49**(4):644–648.
 28. Xiao G, Zhao P, Zhang Y: **A Pivotal Role of Hormones in Regulating Cotton Fiber Development.** *Frontiers in plant science* 2019, **10**:87.
 29. Shibuya N: **Oligosaccharide signalling for defence responses in plant.** *Physiol Mol Plant Pathol* 2001, **59**.
 30. Gordon M, Arasah A, Stern K, Kim HJ, Ratnayaka H: **Photosynthesis and growth characteristics of cotton immature fiber mutant (im) in comparison with isogenic wild type (TM-1).** In: *Plant Biology 2016, American Society of Plant Biologists: 2016.*
 31. Joshi PC, Wadhvani AM, Johri BM: **Morphological and embryological studies of Gossypium L.** 1967.
 32. Stewart JM: **FIBER INITIATION ON THE COTTON OVULE (GOSSYPIUM HIRSUTUM).** *American Journal of Botany* 1975, **62**(7).
 33. Basra AS, Malik CP: **Development of the Cotton Fiber.** *International review of cytology* 1984, **89**(6):65–113.

34. Kang LA, Jing SA, Ly A, Yy B: **Transcriptome analysis reveals critical genes and key pathways for early cotton fiber elongation in Ligon lintless-1 mutant.** *Genomics* 2012, **100**(1):42–50.
35. Walford SA, Wu Y, Llewellyn DJ, Dennis ES: **Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1.** *The Plant journal: for cell and molecular biology* 2012, **71**(3):464–478.
36. Huang Y, Liu X, Tang K, Zuo K: **Functional analysis of the seed coat-specific gene GbMYB2 from cotton.** *Plant physiology and biochemistry: PPB* 2013, **73**:16–22.
37. Pu L, Li Q, Fan X, Yang W, Xue Y: **The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development.** *Genetics* 2008, **180**(2):811–820.
38. Majer C, Hochholdinger F: **Defining the boundaries: structure and function of LOB domain proteins.** *Trends in plant science* 2011, **16**(1):47–52.
39. Theissen G: **Development of floral organ identity: stories from the MADS house.** *Current opinion in plant biology* 2001, **4**(1):75–85.
40. Al-Ghazi Y, Bourot S, Arioli T, Dennis ES, Llewellyn DJ: **Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality.** *Plant & cell physiology* 2009, **50**(7):1364–1381.
41. Wang L, Liu N, Wang T, Li J, Wen T, Yang X, Lindsey K, Zhang X: **The GhmiR157a-GhSPL10 regulatory module controls initial cellular dedifferentiation and callus proliferation in cotton by modulating ethylene-mediated flavonoid biosynthesis.** *Journal of experimental botany* 2018, **69**(5):1081–1093.
42. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nature methods* 2012, **9**(4):357–359.
43. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *Nature methods* 2015, **12**(4):357–360.
44. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC bioinformatics* 2011, **12**:323.
45. Wang L, Feng Z, Wang X, Zhang X: **DEGseq: an R package for identifying differentially expressed genes from RNA-seq data.** *Bioinformatics* 2010, **26**(1):136–138.
46. Sturn A, Quackenbush J, Trajanoski Z: **Genesis: cluster analysis of microarray data.** *Bioinformatics* 2002, **18**(1):207–208.
47. Yang X, Wang L, Yuan D, Lindsey K, Zhang X: **Small RNA and degradome sequencing reveal complex miRNA regulation during cotton somatic embryogenesis.** *Journal of experimental botany* 2013, **64**(6):1521–1536.

Figures

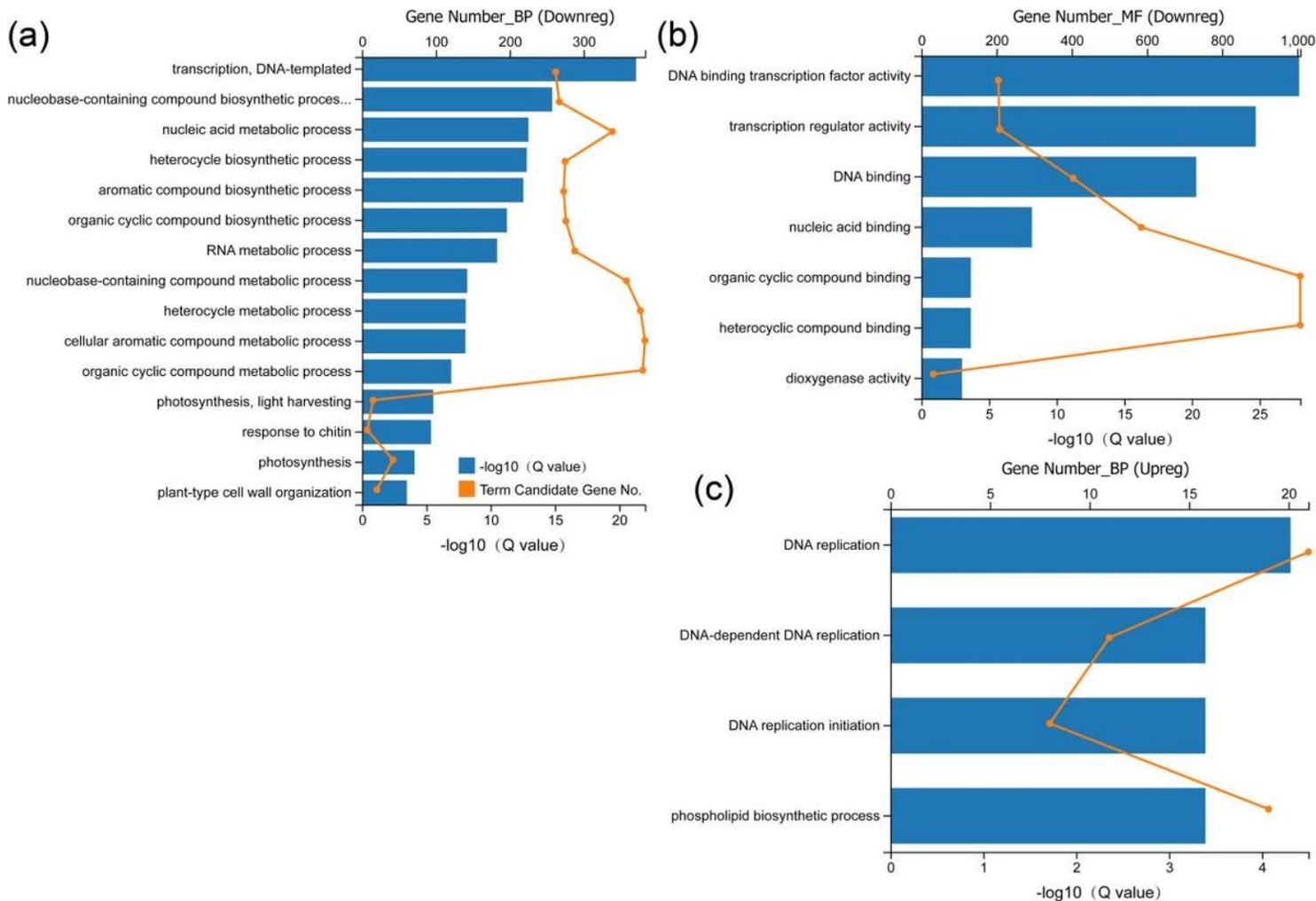


Figure 1

Significantly enriched GO terms on level 3 of the down- and upregulated genes respectively. Graphs showing functional annotation GO terms and gene number for the downregulated genes on the level of biological process (a) and molecular function (b), and for the upregulated genes on the level of biological process (c). The blue bars showing $-\log_{10}(Q \text{ value})$ of the enrichment analysis of each term, and orange dots showing the gene number contained by each GO term. Upreg: upregulated genes; Downreg: downregulated genes; BP: biological process; MF: molecular function.

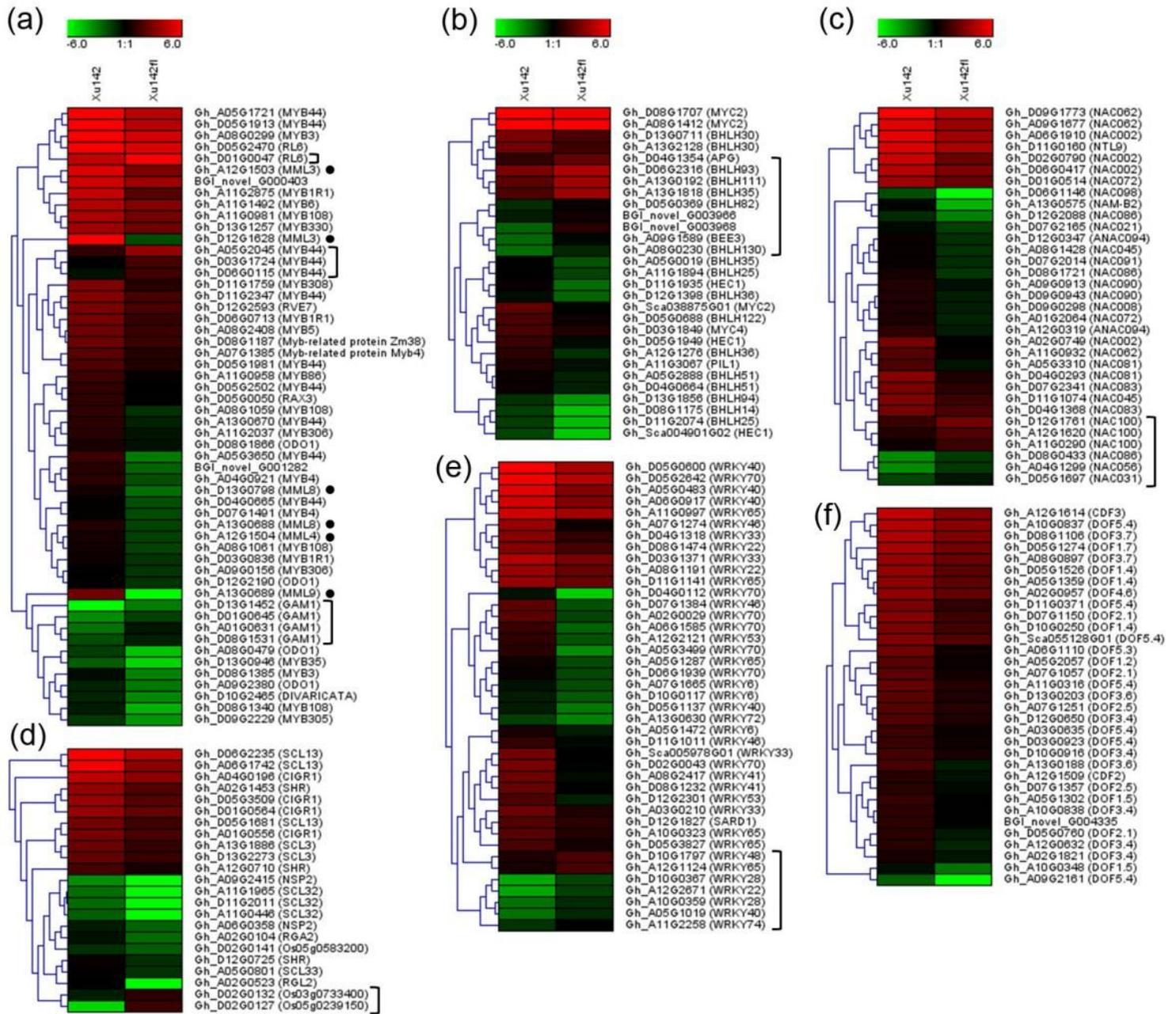


Figure 2

Expression patterns of 6 transcription factor family genes in ovules of Xu142 and Xu142fl Heat maps of MYB (a), bHLH (b), NAC (c), GRAS (d), WRKY (e), and DOF (f) domain containing transcription factor genes. The graphs were generated by Genesis which showing the hierarchical clusters of each type of genes. MMLs homologs in (a) were indicated with black dots, and the upregulated genes in (a)-(e) were marked using a bracket. The log2 values of the RPKM of each gene were used to plot the map, which were indicated by the gradient color bars (red to green reflecting the expression levels from high to low).

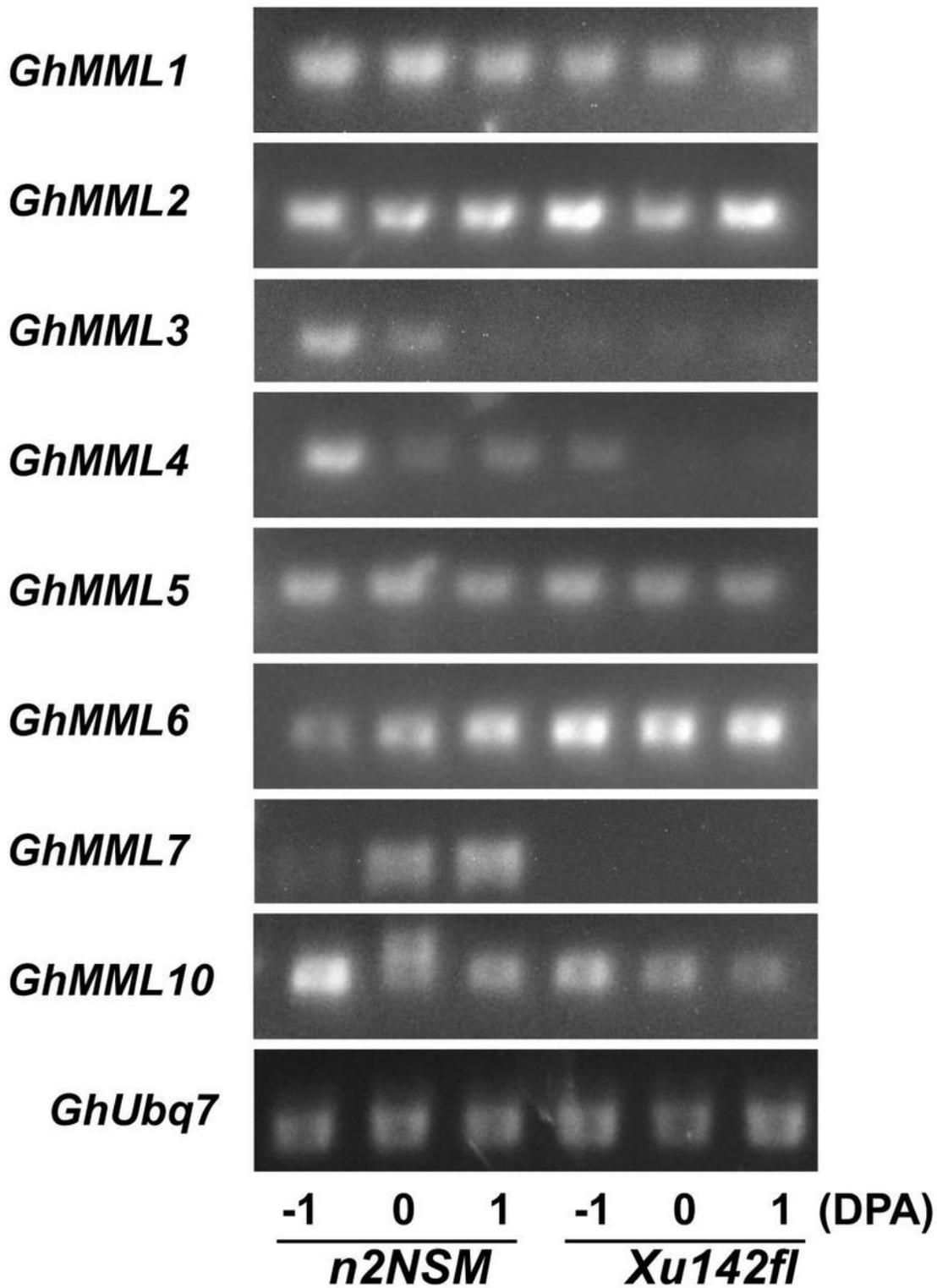


Figure 3

Expression patterns of 8 GhMMLs in ovules of n2NSM and Xu142fl during lint fiber initiation. The transcripts of GhMML1-7 and GhMML10 in -1, 0 and 1 DPA ovules of n2NSM and Xu142fl were detected by RT-PCR. GhUbq7 was used as an endogenous reference gene.

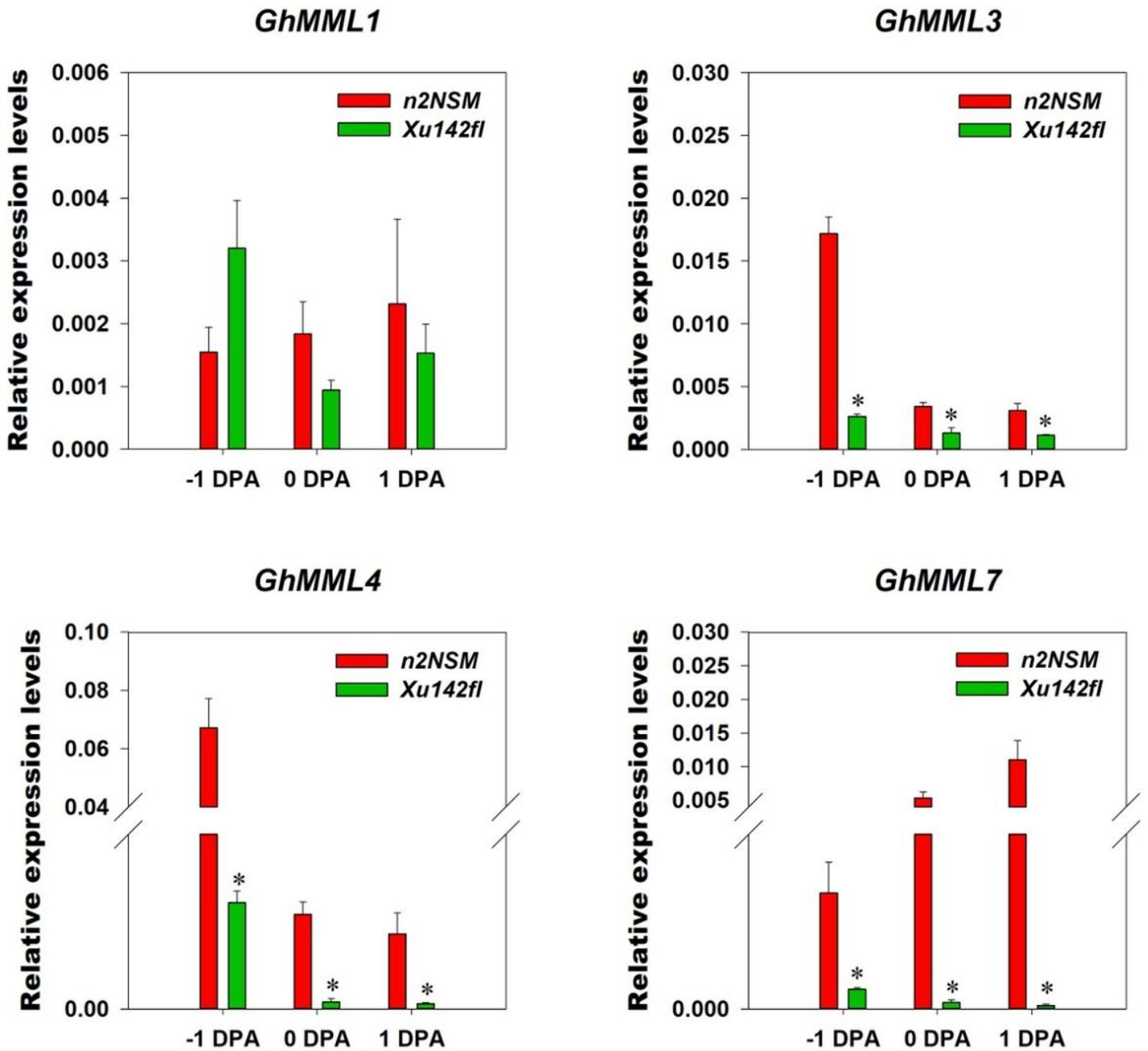


Figure 4

Expression patterns of GhMML1/3/4/7 in ovules of n2NSM and Xu142fl during lint fiber initiation qRT-PCR analysis showing relative expression levels of GhMML1, GhMML3, GhMML4 and GhMML7 in -1, 0 and 1 DPA ovules of n2NSM and Xu142fl. GhUbq7 was used as an endogenous reference gene, and the data represents the mean \pm SD of three biological replicates. "*" represent $p < 0.05$.

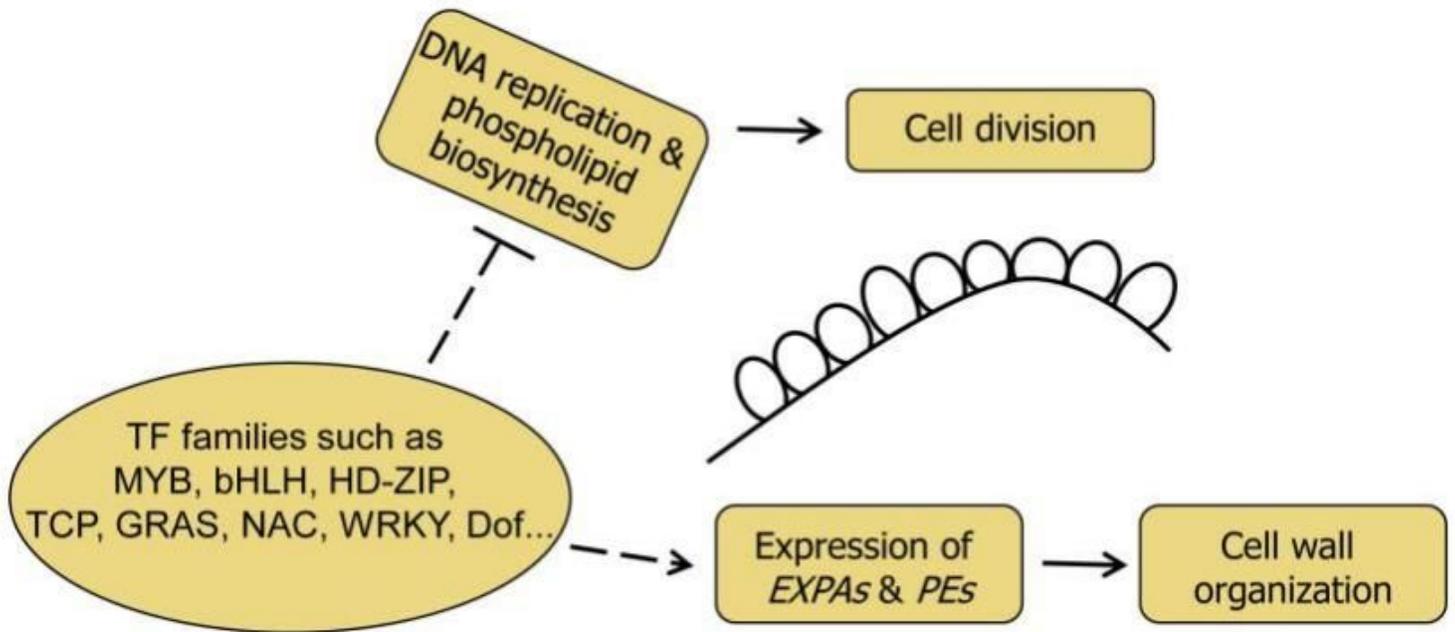


Figure 5

Transcription regulation model of lint fiber initiation in cotton. Multiple TF family proteins are implicated during lint fiber initiation to inhibit DNA replication and phospholipid biosynthesis to terminate cell division, and activate expression of EXPAs and PEs to cause cell wall reorganization and fiber cell expansion. EXPAs: Alpha-expansin encoding genes; PEs: Pectinesterase.

Supplementary Files

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

- [AdditionalFiles.docx](#)