Genome-wide identification, phylogenetic and expressional analysis of the UXS gene family in cotton

Yuxin Pan
North China University of Science and Technology
Jinpeng Wang
North China University of Science and Technology
Zhenyi Wang
North China University of Science and Technology
Hengwei Liu
Suzhou University of Science and Technology
Lan Zhang
North China University of Science and Technology
Li Wang
North China University of Science and Technology
Xiaoming Song
North China University of Science and Technology
Weina Ge
North China University of Science and Technology
Xiyin Wang (wang.xiyin@gmail.com)
https://orcid.org/0000-0003-3454-0374

Research article

Keywords: UDP-glucuronate decarboxylase, genomics, phylogeny, evolution, cotton

Posted Date: January 29th, 2020

DOI: https://doi.org/10.21203/rs.2.22104/v1

Abstract

Background: UDP-glucuronate decarboxylase (UXS) is an enzyme in plants and participates in cell wall noncellulose. Previous research suggested that cotton GhUXS gene regulated the conversion of non-cellulosic polysaccharides and modulates their composition in plant cell walls, showing its possible cellular function determining the quality of cotton fibers. Here, we performed evolutionary, phylogenetic, and expressional analysis of UXS genes from cottons and other selected plants.

Results: By exploring the sequenced cotton genomes, we identified 10, 10, 18, and 20 UXSs genes in Gossypium raimondii , Gossypium arboretum , Gossypium hirsutum and Gossypium barbadense , and retrieved their homologs from other representative plants, including 5 dicots, 1 monocot, 5 green alga, 1 moss, and 1 lycophyte. Phylogenetic analysis suggested that UXS genes could be divided into four subgroups and members within each subgroup shared similar exon-intron structures, motif and subcellular location. Notably, gene colinearity information indicates 100% constructed trees to have aberrant topology, and helps determine and use corrected phylogeny. In spite of conservative nature of UXS, during the evolution of Gossypium , UXS genes were subjected to significant positive selection on key evolutionary nodes. Expression profiles derived from RNA-seq data showed distinct expression patterns of GhUXS genes in various tissues and different development. Most of GhUXS gene expressed highly at 10, 20 and 25 DPA (day post anthesis) of fibers. Real-time quantitative PCR analysis GhUXS genes expressed highly at 20 DPA or 25 DPA.

Conclusions: UXS is relatively conserved in plants and significant positive selection affects cotton UXS evolution. The comparative genome-wide identification and expression profiling would lay an important foundation to understanding the biological functions of UXS gene family in cotton species and other plants.

Background

Plant cells walls are composed of cellulose microfibrils, pectins, hemicelluloses, and lignin[1, 2]. In dicots cell walls, Xyl (xylose) is a key component of hemicellulosic and pectic polysaccharides [1, 3]). UDP-Xyl is biosynthesized via decarboxylation of UDP-glucuronic acid by UDP-glucuronate decarboxylase (EC:4.1.1.35; UXS,), which has a central role in sugar nucleotide inter-conversion[4]. The UXS family bound with NAD^+, belongs to the NDP-sugar epimerases/dehydratases subclass [5]. UXS gene has an N-terminal GxxGxxG pattern that is characteristic of the ADP-binding βαβαβ-fold (Rosmann fold), associated with NAD(P)-binding proteins. The UXS gene family also has YxxxK motif with highly conserved amino acids Ser and Lys to activate Tyr to yield sugar intermediate[6].

UXS activity has been reported in microorganisms, vertebrate species, and plants. Eight hundred and twenty-six in bacteria predicted bacteria UXS enzymes have been identified [7]. The crystal structure, molecular dynamics, and response pathways of UXS have also been analyzed in humans [5]. Six UXS protein isoforms have been isolated from Arabidopsis and three UXSs encode cytosol-localized enzymes

down-regulation/mutations of UXS3, UXS5 and UXS6, encoding cytosol-localized enzymes, led to a drastic reduction in secondary wall thickening [8, 9]. In tobacco, several UXSs have been purified and antisense UXS could modification of hemicellulose content [10]. Also, rice and populous genes were cloned [11, 12]. In G. hirsutum, three GhUXS genes expression suggests UXS may play important roles in cotton fiber development [13, 14].

During plant genome evolution, tandem duplications, single-gene duplications, and large-scale or whole genome duplications (WGD) are the motor to promote duplicated genes derived [15]. To adjust to the changing environment, copies duplicated genes may be lost, nonfunctionalized, or develop novel functions through acquiring various mutations [16].

Gossypium is the main fiber plant in the world-wide. Until now, two diploid cottons (G. raimondii and G. arboreum) [17–19] and tetraploid cottons (upland cotton, G. hirsutum, and sea island cotton, G. barbadense,) [20–25] had their genomes deciphered, and an ancestral decaploidy occurred to have resulted in a complicated genome structure of cotton plants [26]. These efforts provided valuable opportunities to investigate the functional evolution and phylogenetic characteristics of cotton UXS genes. In this article we analyzed UXS genes in G. raimondii (Gr), G. arboretum (Ga), G. hirsutum (Gh), and G. barbadense (Gb). Detailed information on their genomic structures, chromosomal locations, phylogenetic trees, and expressional pattern is provided. The results obtained here provide a further characterization of the evolution and function of UXS genes within Gossypium and a framework for UXS evolutionary in plants.

Results

UXS family members in cotton and other plants

Six Arabidopsis UXS protein sequences were used in BLAST search to identify cotton UXS proteins. Finally, 112 UXS genes involved plants were identified (Table S1). A total of 10, 10, 18 and 20 UXS genes were found in Gr, Ga, Gh, and Gb, respectively. This shows a 1–1 orthologous correspondence between two diploid cottons, and between two tetraploids, respectively, and near 1–2 correspondence between the diploids and tetraploid, showing likely loss of copies after the tetraploidization. Using the same approach, a total of 54 UXS homologs were identified from other 12 plants (Table S1). Except Vv08G0862 and GhA05G097200, the core catalytic domain of the UXS, GXXGXXG and YXXXK, was conserved.

Seven primitive plant species have fewer UXS genes than higher plants. For example, there is only one UXS gene in Ch. Reinhardti, C. subellipsoidea, V. carteri and M.pusilla (Table S1).

Phylogeny, Gene Structural and Conserved Motifs

A total of 112 UXSs were split into 8 groups representing 4 subfamilies (I, II, III and IV) by the phylogenetic tree (Fig. 1). Group I, II, and III were like those in A. thaliana identified previously [11, 27]. Seven UXS genes from 5 different green alga species were included in group IV, indicating that they might be orthologs

originated from a single ancestral gene. Each of the dicotyledon, monocot, moss, and lycophyte species contributed at least one UXS gene to group II and group III, and these two clades were further divided into three different subgroups. Group II contained 3 subgroups, named group IIa, IIb and IIc. Group III also included three subgroups, named group IIIa, IIIb and IIIc. Group I only possessed member from dicotyledon and rice species, including 2 subgroups (group Ia and Ib).

UXS members from phylogenetically closer species were clustered together within four groups. For example, group Ia contained members from dicots, group Ib from monocot-specific UXS genes. Multiple homologs in subtrees showed duplication in each genome of cotton and durian. Most of UXS members were clustered reflecting the species evolutionary relationship. However, there were UXS members have subtree topologies not consistent to species relationship. The cotton genes were often split into subgroups by cacao and durian homologs, which should be their outgroups. This suggests that UXS genes expanded after species formation, possibly related to polyploidization, and the expansion might involve unbalanced evolutionary changes among duplicated genes. These findings will be further discussed below.

Notably, Gossypium UXS genes have longer evolutionary branches, which illustrates Gossypium have faster evolutionary rates than homologs from the other plants. This is evidence of adaptive evolution of UXS genes in cotton.

The above classification is supported by analysis of exon-intron structures and conserved motifs as shown below. To investigate the possible structural evolution of UXSs, we characterized the exon-intron organizations of UXS genes (Fig. 2). Generally, the exon-intron structure is highly conserved within a certain group. Number of introns (0–13) and the length of introns (22–1970 bp) in the UXS family genes are significantly diverse. All genes in group I have a relatively small number (1–8) of exons. In group II, there were 4 to 7 introns. Groupe III genes have more (12–14) exons. Ol09G33309, Ol14G43376, Ol21G29801 and Mp07G109504 and Cs03G64859 in group IV have none intron. As shown in Fig. 2, the UXS genes clustered in the same subfamily shared similar exon-intron patterns. As an older species, Vitis vinifera had distinctively longer introns.

To better characterize the UXS gene family, conserved motifs were identified in UXS proteins (Table S2) by MEME. In total, 20 distinct motifs were identified. UXS typical domain GXXGXXG and YXXXK were located in the motif 6 and motif 1. Although the core catalytic domain of the UXS is conserved, except Vv08G0862and GhA05G097200, variable regions are mainly focused on the N termini. Overall, the same subgroup shares similar motif compositions, indicating a highly functional conservation. Group I and III have more motifs than group II and IV. Group I and III had the particular motif 20 and motif15 at N-terminal. In group III, in most of genes motif 18 and 6 replace motif 10 and 12 of group I. In group I, several cotton genes have a special motif 19. Group II and IV genes do not have the N-motif, as compared to genes in group I and III. Motif 15 is at the N-terminal and motif 11 is its special motif in group II. Motif 11 is partially exemplified the distribution of the N-terminal domains. All these specific motifs may have contributed to the functional divergence of UXS genes.

In addition, analysis of the UXS using PSORT and TargetP1.1 program indicated that UXS were predicted in cytoplasm, chloroplast, endoplasmic reticulum (E.R.), and mitochondrion (Table S1). Similar subcellular localization genes are distributed in the same branch, for example, At02G28760, At03G46440, At05G59290, Pt01G237200, GhA05G097200, GbA05G009370, Dz06G0335 and Tc10G032030 are all located in chloroplast. This illustrates that UXS gene function is related to its cell location [9, 27]

Chromosomal synteny and duplication in Gossypium

The chromosomal location of each UXS genes was established in cotton. As shown in Fig. 3, UXS genes are distributed on 9 chromosomes in Ga, 9 in Gr, 18 in Gh, and 18 in Gb. Most chromosomes have a single UXS gene and two genes are present in chromosome Gr09, Ga04, GbA05, and GbD04.

Gene duplications in genomes could provide important information for gene evolution analysis. UXS gene duplication analysis in genome A, genome D and genome AD were performed respectively. A total of 40 pairs of syntenic paralogs are identified among the 4 cotton species (Fig. 3, Table S3). UXS genes form 7, 12, 6 and 16 synteny-supported paralogous pairs in Ga, Gr, Gh, and Gb, respectively. Checking gene collinearity, gene structure and phylogeny, we found that syntenic paralogs had similar gene structure or motif and were in the same branch in the phylogenetic tree, such as, Ga11G2843, Ga08G0887, Ga11G2843 and Ga03G0563. This shows UXS gene expansion through the ancestral decaploidy and more ancient genome duplication [28, 29].

Ks was estimated of duplicated UXS genes (Table S3). Two paralogs with Ks < 0.03 between allotetraploid cotton (Gh and Gb), consistent with the time of the divergence of Gb and Gh [24]. There were also 18 paralogs with Ks to be 0.40–0.80, associated with the decaploidy occurring approximately 16.6 (13.3–20.0) million years ago in the Gossypium. In addition, around one fourth of pairs occurred within the Ks at 1.3-2.0, corresponding to the paleo-hexaploidization (ancient hexaploidization) event shared among the dicots. In view of species collinearity, many UXS are whole genome duplication (Table S3). Therefore, expansion of UXS genes in cotton might have occurred due to the large-scale duplication events in evolution.

The micro conservative collinear indicated that UXS gene may be lost much in cotton plants after decaploidization. For example, though affected by a decaploidy event, the number of UXS genes in G. raimondii is much fewer than five times of those homologs in grape (Table S4). Based on gene colinearity analysis, we found that each of four cacao UXS genes had one to three orthologous copies in cotton A or D subgenomes, 2 to 6 copies in the tetraploid cotton (Table S4), suggesting extensive gene losses after polyploidization.

Selective pressure analysis

The ratio of Ka/Ks can be used to show selective pressure acting on coding sequences [30]. The Ka/Ks of cotton UXS paralogous pairs were estimated. The Ka/Ks ratios of UXS pairs were all < 1 (Table S3), suggesting that most of them were subjected to negative selection.

According to the homeologous UXS gene colinearity (Table S4), we found that the tree of UXS genes was often not well reconstructed. Paralogs produced by the cotton-specific decaploidy are expected to group together on the tree, with their Dz and Tc orthologs to form outgroups. However, we found five branches of Tc and Dz orthologs were grouped together, which were not in accordance with the species relationship (Fig. 1). These aberrant groupings account 100% of homologous subgroups of Tc, Dz, and Gossypium genes. Actually, at least four times the Tc and Dz orthologs came between their cotton orthologs, splitting the cotton orthologs into two subgroups. For examples, unexpectedly each of Tc02G030560, Tc06G005310, Tc0910g032030, and Tc11G003610 was grouped with a subset of their cotton orthologs, with the other cotton orthologs being the outgroup. These aberrant groupings account 80% of 5 cases of homologous subgroups of Tc, Dz, and Gossypium genes. This intervening phenomenon of Tc and Dz gene into its cotton orthologous cluster on the phylogenetic tree can be explained by increasing substitution accumulation in partial cotton paralogs.

In order to speculate natural selection effect on the evolution, we must have a correct phylogenetic tree. Here, based on gene colinearity, for each subgroup we restricted the Tc and Dz UXS genes, as outgroups, to be clustered with their corresponding collinear cotton orthologs. Therefore, we made trees with possibly corrected topology. Based on the corrected trees of UXS genes (Fig. 4), we inferred positive selective pressure along each lineage. We found that different subgroups of cotton UXS genes have been under divergent evolutionary pressure. The subgroup displayed in Fig. 4a, with a common Tc ortholog Tc02G030580, has two branches under positive selection, with one decaploid-produced lineages of gene (Ga08G0887) was likely subjected to positive selection after the decaploidization event. As to the subgroup displayed in Fig. 4b, one decaploid-produced lineages of gene and one D subgenome gene were likely subjected to positive selection. In subgroup displayed in Fig. 4c, none selection was detected. In subgroup displayed in Fig. 4d, with a Tc ortholog Tc10G032030, three A subgenome gene (GhA05G097200, GbA05G009370 and GhA03G040100) of them were likely under positive selection after the formation of the tetraploid cotton. In the subgroup displayed in Fig. 4e, with a GbA05G035510 and Ga04G1413 genes were likely under positive selection after the origination of A genome.

Expression of UXS among various tissues and development

To better understand the tissue-specific expression profile of cotton UXS genes, FPKM (Fragments per Kilobase of transcript per Million mapped reads) values were used to assess their expression levels across different organs and developmental stages (Table S5). As shown in Fig. 5, of these 18 GhUXS genes, GhD11G119700, GhD10G225110, GhA11G114600 and GhA10G216100 were ubiquitously abundant in most tissues, suggesting that they might execute some universal roles in plant growth and development processes. GhD11G208200 and GhD10G225110 is highestly expressed in petal. GhA11G114600 is preferentially expressed in fiber at 10DPA. GhA10G216100 was highestly expressed in root. On the contrary, expression of GhA02G050000, and GhD02G055100 were relatively low in all tissues. With respect to fiber, all transcripts are abundant in 10 and 20 DPA and GhA11G114600 is the highest. During fiber development, all the transcripts are preferentially expressed at 10 or 20DPA.

GhA11G114600 is highly expressed in fiber. Therefore, UXSs play a role in cotton growth. Further analysis indicated that all these highly expressed genes belonged to clade I of the phylogenetic tree. And also, these genes mainly located in cytoplasm, except GhD10G225110.

We carried out quantitative real-time RT-PCR analysis of GhUXS genes in fiber development (Table S6, Fig. 6). The data shows that all of the 18 genes are expressed during fiber development stage and have distinct but partially overlapping expression profiles. The expressions of most of genes increase significantly from 10 DPA to 25 DPA, with peak values about 20 DPA or 25 DPA. This implied that UXS genes expressed highly in the overlapping stage of fiber primary and secondary cell wall synthesis. Contrary to RNA-seq, some of genes are highest expressed at 20DPA. For example, GhD11G119700 is in high expression at 20 DPA, whereas RNA-seq was lower. The inconsistency probably was the consequence of genotype-dependent expressions. In addition, GhA11G114600, GhD11G119700, GhD05G097300 and GhD03G127900 located in cytoplasm have relative high expression. All of this illustrated that UXS gene subcellular location may affect their function.

Discussion

The number of UXS family genes indicated the UXS family is not a large family. The number of UXS family genes in Arabidopsis, poplar, and rice was same to previous analysis [4, 12]. Compared to the grape, larger number of UXS arises in other dicotyledons. Gene duplication, including tandem duplication, segmental duplication, transposition events, and whole-genome duplication, acts the major role in plants evolution process [31–34]. Here, we found that many UXS genes are in colinearity in each cotton genome, and can be related to the cotton-specific decaploidy. This shows that ancient polyploidies contributed to their expansion. Besides, gene losses following cotton polyploidies occurred to reduce increased gene redundancy [26], but still retain appreciable number of duplicated genes. Furthermore, some truncated genes were detected in grape and G. hirsutum, indicating that some UXS genes might become nonfunctional through pseudogenization [35].

Evolutionary analysis always starts from reconstruction of gene phylogeny, and in animals often use orthologous genes as subjects of study. This is not sufficient in plant research in that enormous duplicated genes exist in plant genomes. Recursive rounds of polyploidies or whole-genome duplications (and other gene duplications) resulted in thousands of duplicated genes in extant plant genomes [36]. These duplicated genes provide enormous opportunities for biological innovation [37–39]. It has been more and more evident that whole-genome duplications have contributed to the origination, fast divergence [40], establishment of major plant taxa[41, 42], such as seed and flowering plants [43], and furthermore the subsets of the latter monocots and dicots [44–47].

Up to date, duplicated genes have been a hot spot in plant biological research. Theoretically, the existence of abundant copies of a gene may lead to different evolutionary avenues: they may divide their ancestral gene's functions (subfunctionalization), one of them may develop new functions (neofunctionalization), or one may lose its functional (pseudo-functionalization) [48]. A meaningful discussion of gene

functional changes depends on a reconstruction of gene phylogeny, reflecting their evolutionary history. However, often divergent evolutionary rates among genes may result in difficulty to reconstruct the phylogeny, and a wrong phylogenetic tree would lead to problematic explanation of their functional changes about adaption and natural selection [45].

Here, according the gene colinearity produced by the cotton-specific decaploidy, we found that accelerated evolution of some subgroups of orthologs may lead to topological changes in phylogenetic trees. Further exploration of functional changes must be based on corrected tree topology. Fortunately, the gene colinearity information could help us perform tree construction, and by using the correct trees we explored positive pressure along each gene lineage. Notably, with a small sample, we found that 80% of cases could have aberrant trees. This is a percentage that could not be neglected in phylogenetic and evolutionary analysis. We propose that similar analyses from the plant community take a serious attitude in their research of gene evolution and use gene colinearity as an indispensible check of their reconstructed trees.

UXS is a crucial enzyme for the non-cellulosic polysaccharides biosynthetic pathway. We showed that the majority of the UXS gene exhibited similar expression trend, increased significantly from 20 DPA to 35 DPA.UXS family had extremely expressed in elongation and secondary cell wall synthesis stage [13]. Most of the paralogous UXS gene pairs exhibited different expression profile in the fiber development stage, indicating that duplicated genes could have much divergent expression profiles. These duplicated genes might have evolved new functions.

UXS subcellular location may affect its function. Antisense-expressed one cytoplasm UXS gene could reduce xylose content in stem [49]. T-DNA mutant of UXS3, UXS5 and UXS6, located in the cytoplasm, had lower xylans than that in the wild type. The three UXS genes located in the cytoplasm play a more important role than the UXS genes located in the Golgi apparatus [9]. Due to the different prediction method, there were considerable discrepancy, but the AtUXSs in cytoplasm were ties specific function. Similar to AtUXS genes, antisense-expressed GhUXS3 gene, which was located in cytoplasm, could reduce xylose content in stem in Arabidopsis [14]. GhD08G085500 and GhD12G134200, predicted in cytoplasm, had higher expression. Further research into gene subcelluar location and function is still needed to explore to understand their biological functions.

Conclusions

Here, 112 UXS genes are identified in 17 plants. We found that (1) UXS genes in plants have highly conservative sequence; (2) UXS genes mainly expand through segmental duplication and have significant positive selection during evolution, especially in cotton; (3) UXS genes lost much following cotton decaploidy; (4) Most of the UXS genes expressed higher at secondary cell wall synthesis stage. The present research will contribute to understanding biological synthesis of fibers in Gossypium species.

Methods

Identification of UXS Genes and subcellular location analysis

The Gr (JGI, D subgenome) Ga (A subgenome), Gh (JGI, AADD) and Gb (HAU, AADD) genome sequence, and annotations were downloaded from CottonFGD database (http://www.cottonfgd.org/). The genome information of other five dicotyledons (Arabidopsis thaliana, Populus trichocarpa, Vitis vinifera, Therobroma cacao, Durio zibethinus), a monocot (Oryza sativa), five green alga species (Chlamydomonas reinhardti, Ostreococcus lucimarinus, Coccomyxa subellipsoidea, Volvox carteri, Micromonas pusilla), amoss (Physcomitrella patens) and a lycophyte (Selaginella moellendorffii) was also retrieved from Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) to analyze UXS.

The amino-acid sequence of A. thaliana UXS was used to search for potential UXS homolog hits in the whole genome sequence with local BLASTP at the Gr, Ga, Gh, and Gb genome database with E-value less than 1e-5. At the same criteria, UXS homologs were identified from other 13 plant genomes. The annotated protein sequences were then checked for the conserved UXS domain using SMART (http://smart.emblheidelberg.de)[50] and performing Pfam (http://pfam.sanger.ac.uk) searches [16]. Systematic names were assigned based on the chromosome distribution for plant.

The subcellular location of UXS was predicted using PSORT (https://psort.hgc.jp/)[51] and TargetP (http://www.cbs.dtu.dk/services/TargetP/) [52].

Phylogenetic, structural and motif analyses of the UXS proteins

To gain the insight into the evolutionary relationships of cotton UXS and their homologs in other plant species. The UXS sequences of the candidate proteins were aligned using the ClustalW program with the default parameters [53]. Sequence alignment of multiple UXS was performed using MEGA 6.0 [54] and a phylogenetic tree was constructed using the Maximum Likelihood method with a bootstrap analysis of 1000 replicates.

The exon/intron structures of individual UXS genes were analyzed using the GSDS (Gene Structure Display Server) software through the cDNA sequences and their corresponding genomic DNA sequences position information [55].

Conserved motifs in different UXS proteins were analyzed using MEME (http://meme-suite.org/tools/meme), and the optimized parameters were: maximum number of motifs, 20; and, the optimum width of each motif, between six and 100 residues.

Chromosomal Location and estimation of duplication events

Chromosomal distribution of cotton UXS genes was displayed using MapInspect. Conserved collinear blocks in cotton were evaluated using MCSCAN program [56], and alignments with E-value 1e-5 or lower were considered as significant matches. UXS genes were defined as segmentally duplicated paralogous if four or more colinear genes were located within 15 adjacent predicted genes upstream and downstream. Tandem duplications were characterized as two or more genes of UXS located in 1 Mb chromosomal locations. At the same time, WGD effect for cotton was also estimated using the colinearity information at the genome level. Multiple alignment of cacao, durian, and cotton was downloaded from the supplemental materials of the previous publication and provided information of decaploid-produced cotton paralogous genes, preserving gene colineartiy, and their cacao and durian orthologs[26, 45].

Ks and Ka calculation and frequent drive of positive selection

Synonymous [57] and non-synonymous (Ka) substitution per site between UXS genes were estimated using DnaSp5.0 [58].

To more rigorously evaluate the positive selection on the evolution of UXS family in Gossypuim plants, Codeml from PAML package was used to calculate ω (the ratio of nonsynonymous to synonymous distances) UXS family [59, 60].

Expression pattern analysis of UXS genes in cotton different tissues

To provide further insight into cotton UXS function, transcriptome analyses were performed using RNA-seq data information from the CottonFGD (https://cottonfgd.org/) of Gh. The corresponding FPKM (fragments per kilobase per million reads) values were log2 transformed. Finally, UXS genes was found in multiple tissue types and various development stages, including ovule (-5-35DPA), fiber (5-25DPA), root, stem, leaf, flower (torus, petal, stamen, pistil and calycle). The normalized expression data was used to generate heatmap using the TBtools [61] to visualize the similarities and differences in the GhUXS family.

RNA extraction and quantitative real-time RT-PCR of cotton fiber

Plants of G.hirsutum ‘CCRI 8’ were grown in an experimental field in Baoding, China. Ovules and epidermal fibers were collected at 0 DPA and 5, 10, 15, 20, 25, 30, and 35 DPA, respectively, and were immediately immersed in liquid nitrogen.

Total RNA was extracted using EASYspin Plus plant RNA kit (Aidlab, China) according to the manufacturers’ instruction. RNA integrity was verified by 1.2% agar gel electrophoresis and the RNA concentration was measured using NanoDrop1000 (Thermo, USA). The PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRA, JAPAN) was used to remove genomic DNA contamination and synthesis the first cDNA strand.

Quantitative real-time polymerase chain reaction (qRT-PCR) was conducted using THUNDERBIRD qPCR Mix (TOYOBO, China) on a ABI7300 Real-Time PCR Detection System (ABI, USA). Cotton EF1A2 (GenBank Accession No. DQ174251.1) was amplified as the reference gene. No-template controls were also used for each primer pair. Each sample had three replicates. Data were examined by the $^{2-\Delta\Delta CT}$ value method.

Conditions for PCR procedures included 40 cycles of 10 s at 95 °C, 10 s at 58 °C, and 20 s at 72 °C.

All the 18 primers for qRT-PCR were designed based on the reference sequence obtained from the G. hirsutum UXS genome database.

Abbreviations

UXS UDP-glucuronate decarboxylase

DPA day post anthesis

FPKM Fragments per kilobase of transcript per million mapped reads

qRT-PCR Quantitative real-time polymerase chain reaction

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data sets supporting the results of this article are included within the article and its additional file.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Hebei Province Science and Technology Support Program, grant number 14962905D, Department of Education of Hebei Province, grant number Y2012025 and the National Natural Science Foundation of China, grant number 31371282.

Author Contributions

XW conceived and led the project, supervised the writing. YP implemented, coordinated the analysis, and writing; JW, ZW, and LZ perform the gene colineartiy analysis; HL and ZW performed the real-time quantification PCR experiments; XS, LW and WG provided the assistance for selective pressure analysis. All authors contributed to revising the manuscript. All authors had read and approved the final manuscript.

Acknowledgments

Thanks the professor Zhiying Ma for providing the cotton materials.

References

1. Rennie EA, Scheller HV: **Xylan biosynthesis**. *Curr Opin Biotechnol* 2014, **26**:100-107.
2. Liepman AH, Wightman R, Geshi N, Turner SR, Scheller HV: **Arabidopsis - a powerful model system for plant cell wall research**. *Plant J* 2010, **61**(6):1107-1121.
3. Ebert B, Rautengarten C, Guo X, Xiong G, Stonebloom S, Smith-Moritz AM, Herter T, Chan LJ, Adams PD, Petzold CJ *et al*: **Identification and Characterization of a Golgi-Localized UDP-Xylose Transporter Family from Arabidopsis**. *The Plant cell* 2015, **27**(4):1218-1227.
4. Harper AD, Bar-Peled M: **Biosynthesis of UDP-xylose. Cloning and characterization of a novel Arabidopsis gene family, UXS, encoding soluble and putative membrane-bound UDP-glucuronic acid decarboxylase isoforms**. *Plant physiology* 2002, **130**(4):2188-2198.
5. Eixelsberger T, Sykora S, Egger S, Brunsteiner M, Kavanagh KL, Oppermann U, Brecker L, Nidetzky B: **Structure and mechanism of human UDP-xylose synthase: evidence for a promoting role of sugar ring distortion in a three-step catalytic conversion of UDP-glucuronic acid**. *The Journal of biological chemistry* 2012, **287**(37):31349-31358.
6. Wierenga RK, Terpstra P, Hol WG: **Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint**. *Journal of molecular biology* 1986, **187**(1):101-107.
7. Coyne MJ, Fletcher CM, Reinap B, Comstock LE: **UDP-glucuronic acid decarboxylases of Bacteroides fragilis and their prevalence in bacteria**. *Journal of bacteriology* 2011, **193**(19):5252-5259.
8. Kuang B, Zhao X, Zhou C, Zeng W, Ren J, Ebert B, Beahan CT, Deng X, Zeng Q, Zhou G *et al*: **The role of UDP-glucuronic acid decarboxylase in xylan biosynthesis in Arabidopsis**. *Molecular plant* 2016, **9**(8):1119-1131.
9. Zhong R, Teng Q, Haghighat M, Yuan Y, Furey ST, Dasher RL, Ye ZH: **Cytosol-Localized UDP-Xylose Synthases Provide the Major Source of UDP-Xylose for the Biosynthesis of Xylan and Xyloglucan**. *Plant & cell physiology* 2017, **58**(1):156-174.
10. Bindschedler LV, Tuerck J, Maunders M, Ruel K, Petit-Conil M, Danoun S, Boudet AM, Joseleau JP, Bolwell GP: **Modification of hemicellulose content by antisense down-regulation of UDP-glucuronate**

decarboxylase in tobacco and its consequences for cellulose extractability. *Phytochemistry* 2007, **68**(21):2635-2648.

11. Suzuki K, Watanabe K, Masumura T, Kitamura S: **Characterization of soluble and putative membrane-bound UDP-glucuronic acid decarboxylase (OsUXS) isoforms in rice**. *Archives of biochemistry and biophysics* 2004, **431**(2):169-177.
12. Du Q, Pan W, Tian J, Li B, Zhang D: **The UDP-glucuronate decarboxylase gene family in Populus: structure, expression, and association genetics**. *PloS one* 2013, **8**(4):e60880.
13. Pan Y, Wang X, Liu H, Zhang G, Ma Z: **Molecular cloning of three UDP-glucuronate decarboxylase genes that are preferentially expressed in Gossypium fibers from elongation to secondary cell wall synthesis**. *Journal of Plant Biology* 2010, **53**(5):363-373.
14. Zhang DM, Pan YX, Zhang Y, Li ZK, Wu LQ, Liu HW, Zhang GY, Wang XF, Ma ZY: **Antisense expression of Gossypium hirsutum UDP-glucuronate decarboxylase in Arabidopsis leads to changes in cell wall components**. *Genetics and molecular research : GMR* 2016, **15**(1).
15. Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH: **Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses**. *Genome biology* 2009, **10**(6):R68.
16. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: towards a more sustainable future**. *Nucleic acids research* 2016, **44**(D1):D279-285.
17. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S *et al*: **The draft genome of a diploid cotton Gossypium raimondii**. *Nature genetics* 2012, **44**(10):1098-1103.
18. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J *et al*: **Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres**. *Nature* 2012, **492**(7429):423-427.
19. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C *et al*: **Genome sequence of the cultivated cotton Gossypium arboreum**. *Nature genetics* 2014, **46**(6):567-572.
20. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J *et al*: **Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution**. *Nature biotechnology* 2015, **33**(5):524-530.
21. 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.
22. Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L *et al*: **The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres**. *Scientific reports* 2015, **5**:17662.
23. Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L *et al*: **Gossypium barbadense genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites**. *Scientific reports* 2015, **5**:14139.

24. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J *et al*: **Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton**. *Nature genetics* 2019, **51**(4):739-748.
25. Wang M, Tu L, Yuan D, Zhu, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G *et al*: **Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense**. *Nature genetics* 2019, **51**(2):224-229.
26. Wang X, Guo H, Wang J, Lei T, Liu T, Wang Z, Li Y, Lee TH, Li J, Tang H *et al*: **Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation**. *The New phytologist* 2016, **209**(3):1252-1263.
27. Kuang B, Zhao X, Zhou C, Zeng W, Ren J, Ebert B, Beahan CT, Deng X, Zeng Q, Zhou G *et al*: **Role of UDP-Glucuronic Acid Decarboxylase in Xylan Biosynthesis in Arabidopsis**. *Molecular plant* 2016, **9**(8):1119-1131.
28. Wang J, Sun N, Deng T, Zhang L, Zuo K: **Genome-wide cloning, identification, classification and functional analysis of cotton heat shock transcription factors in cotton (Gossypium hirsutum)**. *BMC genomics* 2014, **15**:961.
29. Schauser L, Wieloch W, Stougaard J: **Evolution of NIN-like proteins in Arabidopsis, rice, and Lotus japonicus**. *Journal of molecular evolution* 2005, **60**(2):229-237.
30. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions**. *Molecular biology and evolution* 1986, **3**(5):418-426.
31. Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K: **The flowering world: a tale of duplications**. *Trends in plant science* 2009, **14**(12):680-688.
32. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R: **Gene duplication as a major force in evolution**. *Journal of genetics* 2013, **92**(1):155-161.
33. Kaessmann H: **Origins, evolution, and phenotypic impact of new genes**. *Genome research* 2010, **20**(10):1313-1326.
34. Flagel LE, Wendel JF: **Gene duplication and evolutionary novelty in plants**. *The New phytologist* 2009, **183**(3):557-564.
35. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(15):7799-7806.
36. Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC *et al*: **Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses**. *Proc Natl Acad Sci U S A* 2005, **102**(37):13206-13211.
37. Liu SL, Pan AQ, Adams KL: **Protein subcellular relocalization of duplicated genes in Arabidopsis**. *Genome Biol Evol* 2014, **6**(9):2501-2515.
38. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B *et al*: **Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome**. *Science* 2014, **345**(6199):950-953.

39. Coate JE, Doyle JJ: **Divergent evolutionary fates of major photosynthetic gene networks following gene and whole genome duplications**. *Plant Signal Behav* 2011, **6**(4):594-597.
40. Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C, Mikenas J *et al*: **Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family Expansion Associated with Adaptation to Harsh Environments**. *Mol Biol Evol* 2019, **36**(1):112-126.
41. Charon C, Bruggeman Q, Thareau V, Henry Y: **Gene duplication within the Green Lineage: the case of TEL genes**. *J Exp Bot* 2012, **63**(14):5061-5077.
42. Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, Sun S, Lei T, Meng F, Wei C *et al*: **Polyploidy Index and Its Implications for the Evolution of Polyploids**. *Front Genet* 2019, **10**:807.
43. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS *et al*: **Ancestral polyploidy in seed plants and angiosperms**. *Nature* 2011, **473**(7345):97-100.
44. Wang JP, Yu JG, Li J, Sun PC, Wang L, Yuan JQ, Meng FB, Sun SR, Li YX, Lei TY *et al*: **Two Likely Auto-Tetraploidization Events Shaped Kiwifruit Genome and Contributed to Establishment of the Actinidiaceae Family**. *iScience* 2018, **7**:230-240.
45. Wang J, Yuan J, Yu J, Meng F, Sun P, Li Y, Yang N, Wang Z, Pan Y, Ge W *et al*: **Recursive Paleohexaploidization Shaped the Durian Genome**. *Plant Physiol* 2019, **179**(1):209-219.
46. Wang J, Sun P, Li Y, Liu Y, Yang N, Yu J, Ma X, Sun S, Xia R, Liu X *et al*: **An Overlooked Paleotetraploidization in Cucurbitaceae**. *Mol Biol Evol* 2018, **35**(1):16-26.
47. Sun S, Wang J, Yu J, Meng F, Xia R, Wang L, Wang Z, Ge W, Liu X, Li Y *et al*: **Alignment of Common Wheat and Other Grass Genomes Establishes a Comparative Genomics Research Platform**. *Front Plant Sci* 2017, **8**:1480.
48. Abascal F, Corpet A, Gurard-Levin ZA, Juan D, Ochsenbein F, Rico D, Valencia A, Almouzni G: **Subfunctionalization via adaptive evolution influenced by genomic context: the case of histone chaperones ASF1a and ASF1b**. *Mol Biol Evol* 2013, **30**(8):1853-1866.
49. Kuang B, Zhao X, Zhou C, Zeng W, Ren J, Ebert B, Beahan CT, Deng X, Zeng Q, Zhou G *et al*: **The role of UDP-glucuronic acid decarboxylase (UXS) in xylan biosynthesis in Arabidopsis**. *Molecular plant* 2016.
50. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: identification of signaling domains**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(11):5857-5864.
51. Gardy JL, Spencer C, Wang K, Ester M, Tusnady GE, Simon I, Hua S, deFays K, Lambert C, Nakai K *et al*: **PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria**. *Nucleic acids research* 2003, **31**(13):3613-3617.
52. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites**. *Protein engineering* 1997, **10**(1):1-6.
53. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*: **Clustal W and Clustal X version 2.0**. *Bioinformatics* 2007, **23**(21):2947-2948.

54. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0**. *Molecular biology and evolution* 2013, **30**(12):2725-2729.
55. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server**. *Bioinformatics* 2015, **31**(8):1296-1297.
56. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. *Nucleic acids research* 2012, **40**(7):e49.
57. Jackson S, Chen ZJ: **Genomic and expression plasticity of polyploidy**. *Current opinion in plant biology* 2010, **13**(2):153-159.
58. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data**. *Bioinformatics* 2009, **25**(11):1451-1452.
59. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood**. *Computer applications in the biosciences : CABIOS* 1997, **13**(5):555-556.
60. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites**. *Genetics* 2000, **155**(1):431-449.
61. Chen CJ, Chen H, He YH, Xia R: **TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface**. *New Results* 2018.

Additional Files

Additional file 1: Table S1 Features of all the UXS gene family in cotton and other plants.

Additional file 2: Table S2: Plant UXS protein sequences used for phylogenetic tree construction.

Additional file 3: Table S3: Colinear gene pair, duplication and selective pressure value.

Additional file 4: Table S4: The homeologous UXS genes in *Gossypium*, *Durio dulcis* and *Theobroma*.

Additional file 5: Table S5: FPKM of UXS genes in different tissues.

Additional file 6: Table S6: qRT-PCR primers of GhUXSs

Figures

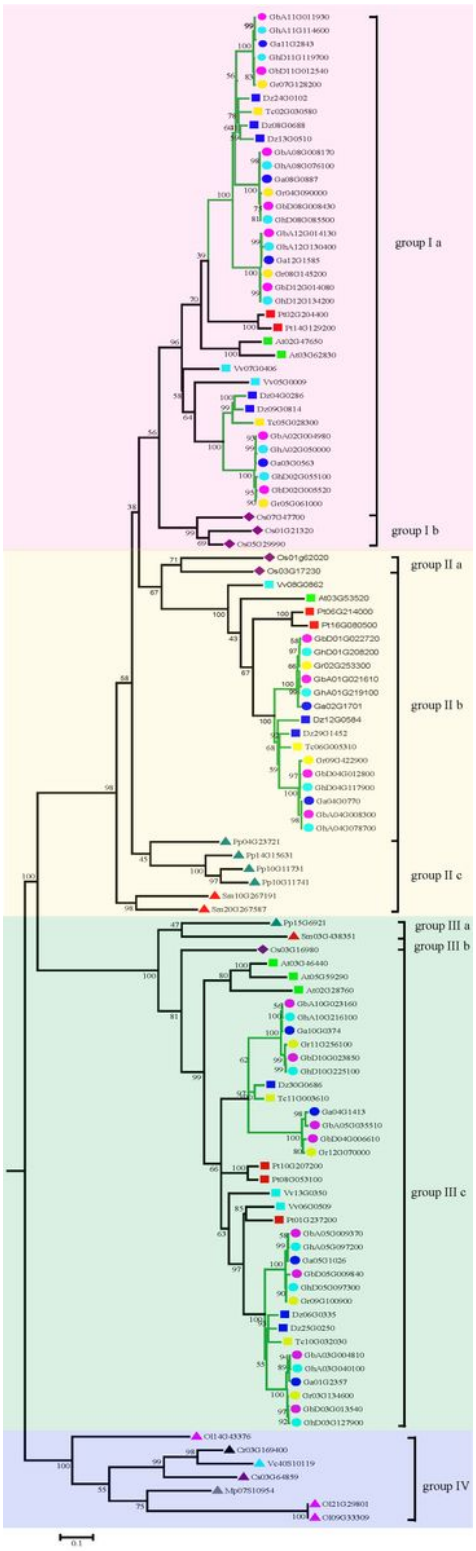

Figure 1

Phylogenetic tree representing relationships of UXS proteins using MEGA (v6) software. The different-colored rectangular background indicates different groups of UXS. Numbers on branches are ML bootstrap percentages. The green line showed the special branch of cotton, cacao and durian.

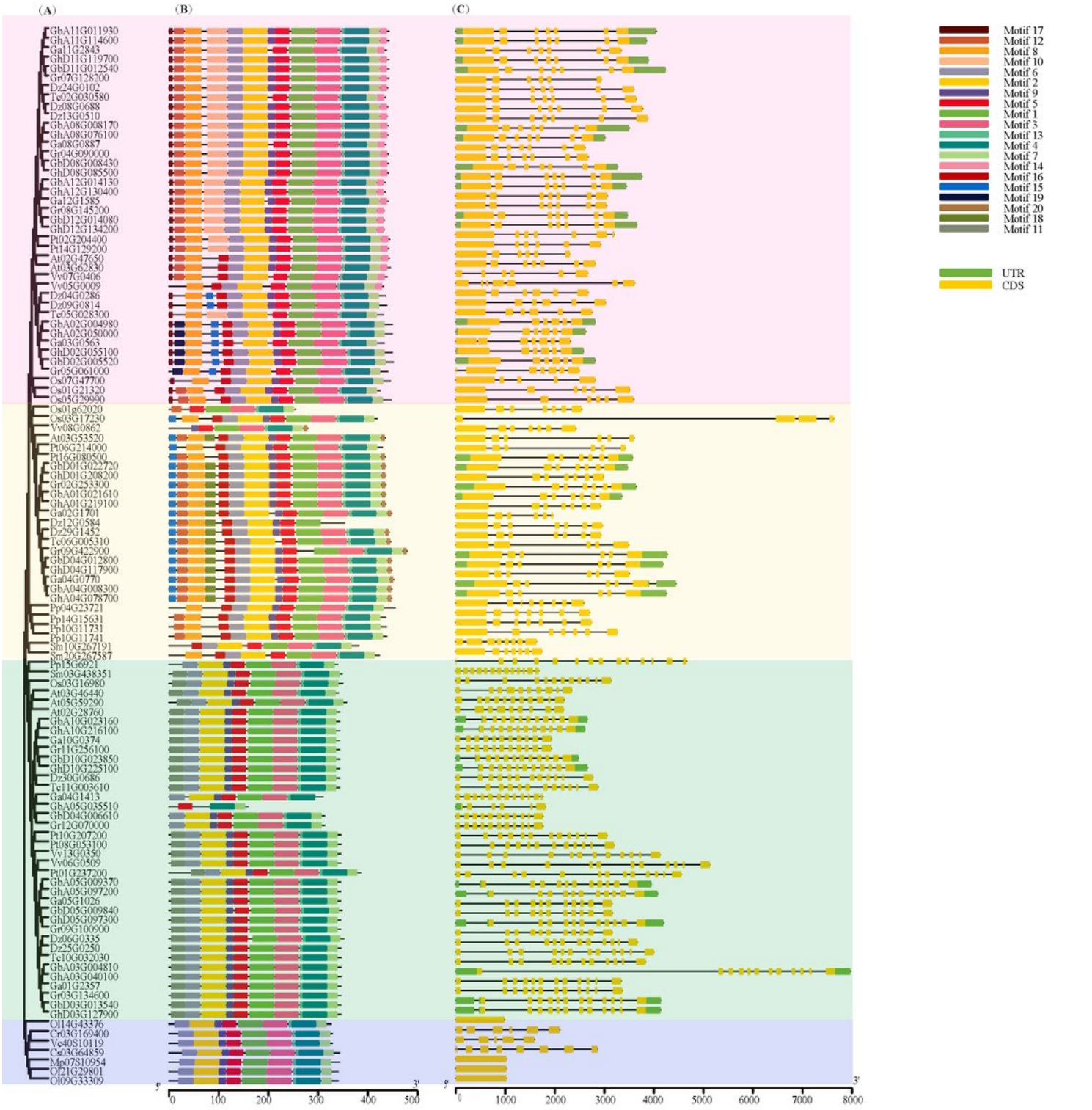

Figure 2

Phylogenetic relationships, gene structure and architecture of conserved protein motifs in UXS genes. (a) The phylogenetic tree. (b) The motif composition of UXS proteins. The motifs, numbers 1-20, are displayed in different colored boxes. (c) Exon-intron structure of UXS genes. Green boxes indicate untranslated 5'- and 3'-regions; yellow boxes indicate exons; black lines indicate introns.

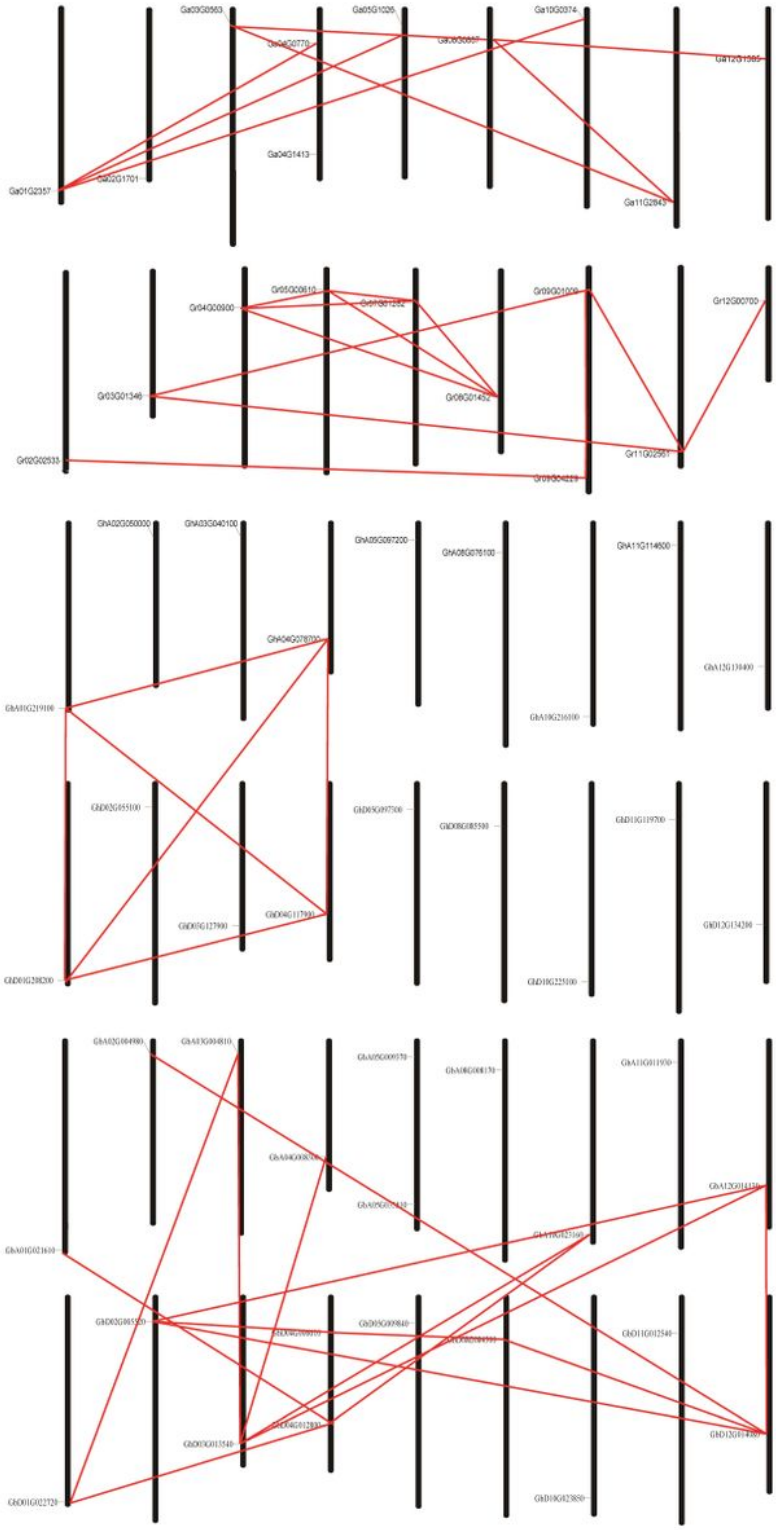

Figure 3

Genomic distribution of cotton UXS genes on chromosomes. The chromosome numbers and size are indicated at the top and bottom of each bar, respectively. The arrows next to gene names show the transcription directions. The number on the right side of the bars designated the approximate physical position. The red line that connects the two genes represents the colinear gene pair.

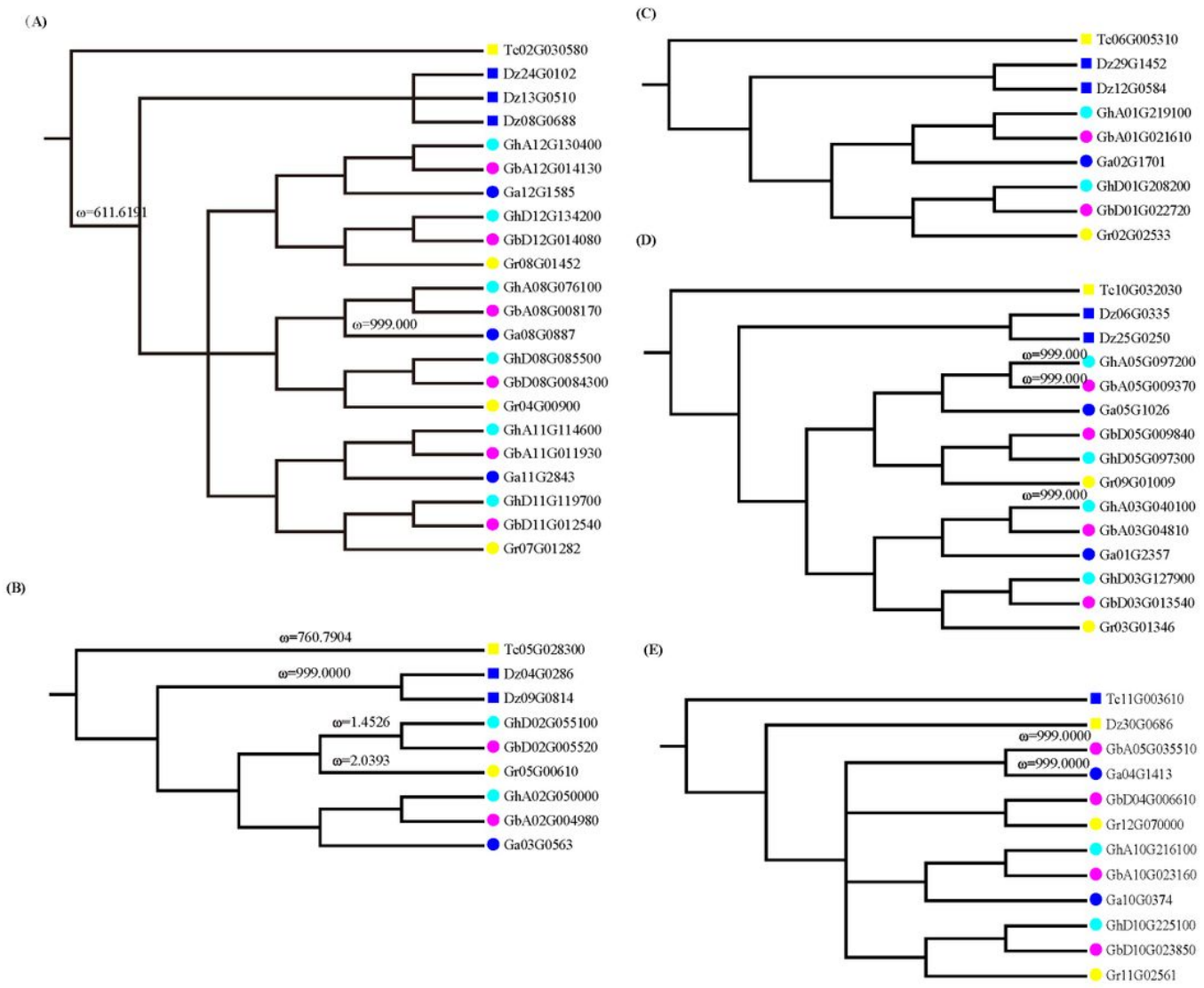

Figure 4

The corrected phylogenetic tree of cacao, durian and cotton. (a) The corrected phylogenetic tree corresponding to the Tc02G30580 branch. (b) The corrected phylogenetic tree corresponding to the Tc05G028380 branch. (c) The corrected phylogenetic tree corresponding to the Tc06G005310 branch. (d) The corrected phylogenetic tree corresponding to the Tc10G032030 branch. (e) The corrected phylogenetic tree corresponding to the Tc11G003610 branch.

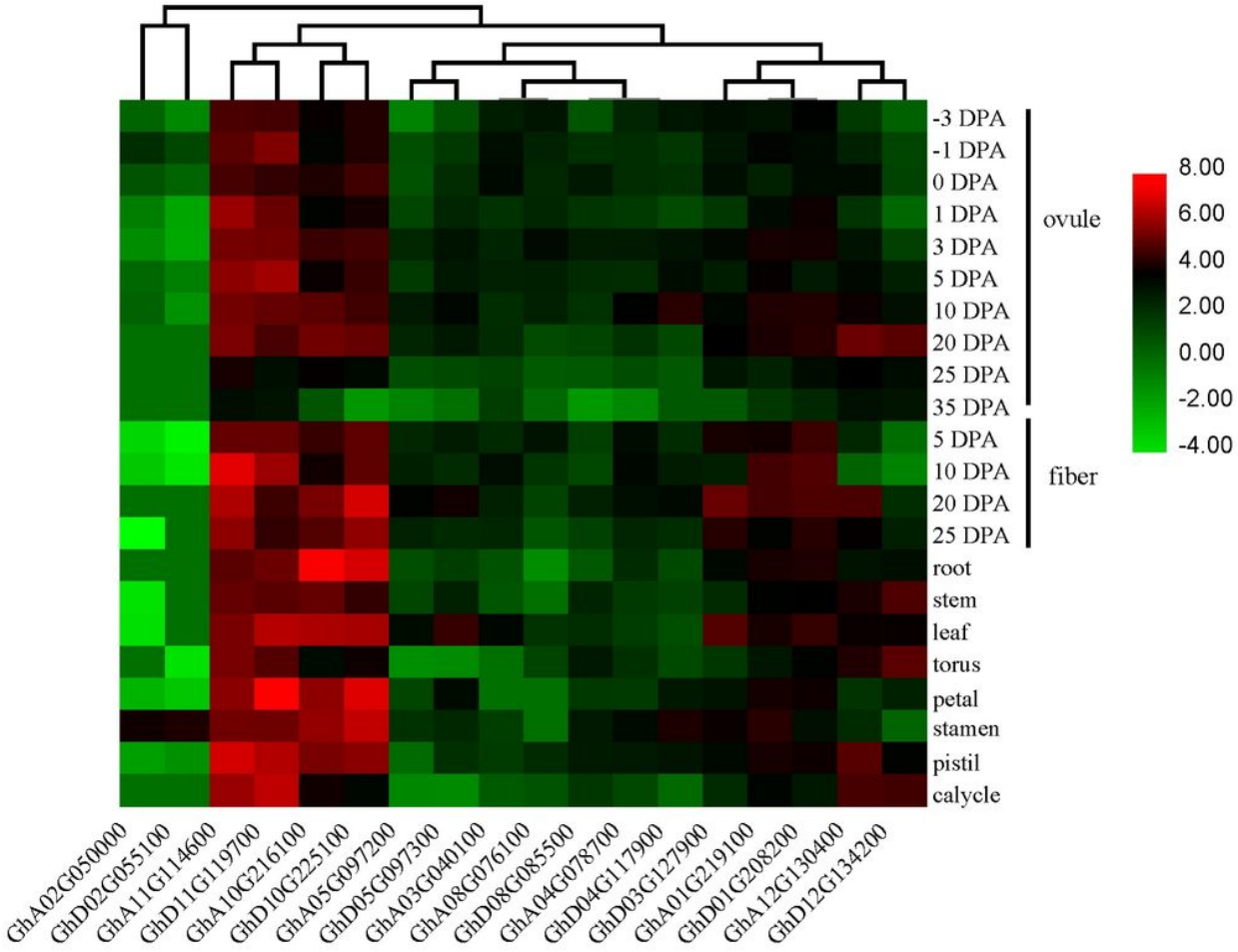

Figure 5

Heat map of expression profiles for GhUXS genes at different tissue and developmental stages.

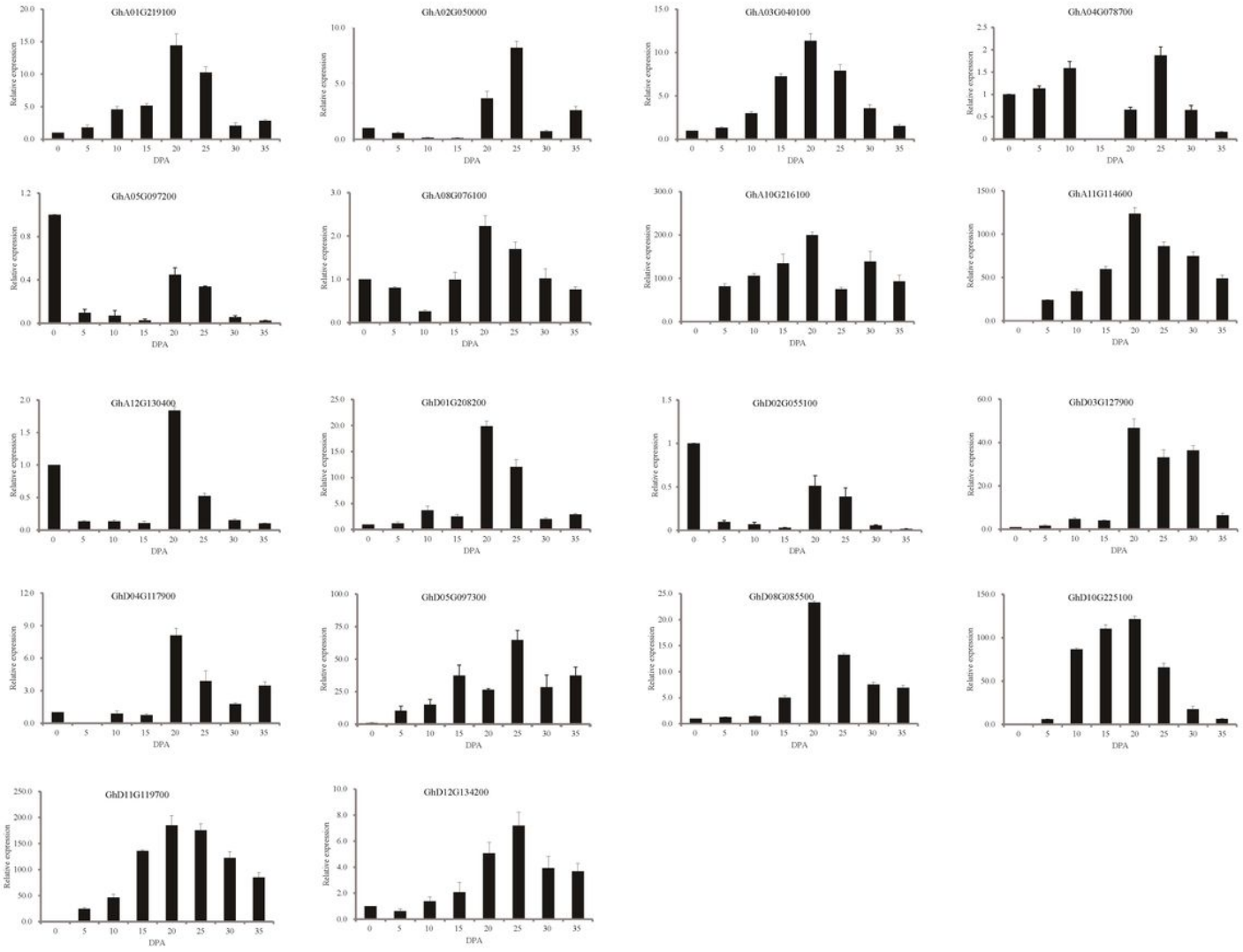

Figure 6

Expression profiles of 18 GhUXS genes in fiber development. Data were normalized to EF1A2 gene and vertical bars indicate standard deviation. Y-axis represents relative expression values and X-axis represents stages of fiber development.

Supplementary Files

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

- TableS4.xlsx
- TableS2.xlsx
- TableS6.xls
- TableS1.xlsx
- TableS5.xlsx

- TableS3.xlsx