Genome-wide identification and expression analysis of GL2-interacting-repressor (GIR) genes during cotton fiber and fuzz development

Xiaoxu Feng1,2 · Hailiang Cheng1 · Dongyun Zuo1 · Youping Zhang1 · Qiaolian Wang1 · Limin Lv1 · Shuyan Li1 · John Z. Yu3 · Guoli Song1

Received: 30 March 2021 / Accepted: 20 September 2021 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Main conclusion GL2-interacting-repressor (GIR) family members may contribute to fiber/fuzz formation via a newly discovered unique pathway in Gossypium arboreum.

Abstract

There are similarities between cotton fiber development and the formation of trichomes and root hairs. The GL2-interacting-repressors (GIRs) are crucial regulators of root hair and trichome formation. The GaFzl gene, annotated as GaGIR1, is negatively associated with trichome development and fuzz initiation. However, there is relatively little available information regarding the other GIR genes in cotton, especially regarding their effects on cotton fiber development. In this study, 21 GIR family genes were identified in the diploid cotton species Gossypium arboreum; these genes were divided into three groups. The GIR genes were characterized in terms of their phylogenetic relationships, structures, chromosomal distribution and evolutionary dynamics. These GIR genes were revealed to be unequally distributed on 12 chromosomes in the diploid cotton genome, with no GIR gene detected on Ga06. The cis-acting elements in the promoter regions were predicted to be responsive to light, phytohormones, defense activities and stress. The transcriptomic data and qRT-PCR results revealed that most GIR genes were not differentially expressed between the wild-type control and the fuzzless mutant line. Moreover, 14 of 21 family genes were expressed at high levels, indicating these genes may play important roles during fiber development and fuzz formation. Furthermore, Ga01G0231 was predominantly expressed in root samples, suggestive of a role in root hair formation rather than in fuzz initiation and development. The results of this study have enhanced our understanding of the GIR genes and their potential utility for improving cotton fiber through breeding.

Keywords Cis-acting elements · Expression patterns · Family genes · Gossypium arboreum · GaGIR

Abbreviations

DPA  Day post anthesis
GEB  Green fluorescence protein
GL2  GLABRA2
GIR  GL2-interacting-repressor
SCW  Secondary cell wall
SD  Synthetic dropout
TF  Transcription factort

Introduction

Cotton is the most important natural fiber crop because it provides the textile industry with raw materials (Fang et al. 2017). Cotton fibers, which differentiate from the epidermal cells of the ovule, undergo the following four distinct,
but overlapping, stages to reach final maturity: initiation, elongation, secondary cell wall (SCW) synthesis, and dehydrated maturation (Padmalatha et al. 2012; Wang et al. 2015; Hu et al. 2016; Sun et al. 2017). The initiation and early elongation stages are essential and highly correlated with fiber characteristics related to yield and quality, including fiber density, length, and uniformity (Rong et al. 2005; Kim et al. 2015; Zhu et al. 2018). Environmental conditions and genetic factors significantly influence these developmental stages, thereby affecting the final fiber quality (Hinchliffe et al. 2011; Gilbert et al. 2014; Liang et al. 2015; Chen et al. 2019). Therefore, the genetic mechanisms underlying fiber initiation should be investigated so that the associated genes may be used to improve cotton fiber quality via molecular breeding.

The recent development and application of high-throughput DNA sequencing technology has resulted in the publication of increasing amounts of cotton genome data, which have been compiled in databases that used as resources for predicting and screening functional genes (Wan et al. 2016; Cheng et al. 2016; Zhu et al. 2017; Wu et al. 2018; Fang et al. 2020). Many genes involved in fiber development have been identified, cloned and functionally characterized (Jiang et al. 2015; Wan et al. 2016; Hu et al. 2016; Thyssen et al. 2017; Wu et al. 2018; Patel et al. 2020; Sun et al. 2020a). For example, GhbHLH18 is negatively associated with fiber quality because it encodes a protein that strongly binds to the E-box element of the GhPER8 promoter to activate expression and modulate peroxidase-mediated lignin metabolism during the fiber elongating stage (Gao et al. 2019). Additionally, GhFSN5, which is a NAC domain transcription factor (TF) gene, is reportedly preferentially expressed during the SCW synthesis stage to negatively regulate the expression patterns of SCW-associated genes related to cellulose, xylan, lignin and several TFs mediating SCW formation (Sun et al. 2020b). In earlier studies, GhPIN3a expression is down-regulated by cytokinin, and the resulting change to the polar distribution of GhPIN3a disrupts the asymmetric accumulation of auxin in the ovule epidermis to inhibit cotton fiber initiation (Mei and Zhang 2019; Zeng et al. 2019). Some loci or candidate genomic regions associated with fuzz fiber initiation and formation in tetraploid cotton lines were detected by multiple combined analyses, including GhMML3_A12 for the N1 locus (Wan et al. 2016), GhMML3_D12 for the n2 locus (Zhu et al. 2018; Chen et al. 2020), the GhMML3_A12 allele for the n1 locus (Chen et al. 2020), the 411-kb genomic interval on chromosome D04 for the n1 locus (Naoumkina et al. 2021a, 2021b), and the 250-kb candidate region on chromosome D13 for the N2 locus (Zhu et al. 2021).

Cotton fibers share many similarities with Arabidopsis thaliana trichomes and root hairs (Lee et al. 2007). Studies on the regulatory network involved in determining Arabidopsis cell fates have provided researchers with a framework for investigating cotton fiber initiation and elongation (Balkunde et al. 2010; Yang and Ye 2013). The multimeric complex that activates trichome initiation and development includes the R2R3 MYB protein GLABROUS1 (GL1), the WD40 repeat-containing protein TRANSPARENT TESTA GLABRA 1 (TTG1), and the bHLH proteins GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) (Pattanaik et al. 2014; Dai et al. 2016; Matías-Hernández et al. 2016). Researchers have identified and verified the homolog of the complex-encoding genes as well as other regulatory genes in cotton (e.g., GhMYB25-like, GhTTG1/4, GhHOX3 and GhHD-1) and subsequently confirmed their effects on fiber development (Walford et al. 2012; Huang et al. 2013; Shan et al. 2014; Liu et al. 2015; Ioannidi et al. 2016; Wan et al. 2016; Ding et al. 2020).

Most genes contributing to trichome development are also involved in root hair patterning, suggesting the underlying models are similar, but the resulting phenotypes are different (Matías-Hernández et al. 2016; Wei and Li 2018). The overexpression of GLABRA2 (GL2) in the shoot eventually activates trichome formation. Root epidermal cells expressing GL2 are prevented from forming root hairs (Rerie et al. 1994; Di Cristina et al. 1996; Ohashi et al. 2002, 2003; Hülskamp 2004). The GL2-interacting-repressors (GIRs) negatively regulate root hair development by interacting with GL2 and controlling the interaction network (Wu and Citovsky 2017a, b). Several recent studies indicated that GaFzl, which is annotated as GIR1 in diploid cotton, adversely affects fuzz development (Du et al. 2018; Feng et al. 2019; Liu et al. 2020; Wang et al. 2020c). However, the characteristics and underlying mechanisms of the GIRs functions and regulatory network in cotton remain unknown.

In this study, we revealed the basic features of GaFzl by determining the subcellular localization and analyzing its transcriptional activation of the encoded protein. Furthermore, on the basis of the published cotton genomes, we screened and identified 21, 21 and 40 GIR gene family members in Gossypium arboreum, Gossypium raimondii and Gossypium hirsutum, respectively, and then characterized their phylogenetic relationships, gene structures, chromosomal distribution and evolutionary dynamics. We also investigated the expression patterns of GaGIR genes and other fiber-related genes in cotton fiber. The results presented herein will be useful for clarifying the molecular mechanisms associated with the GIR genes in diploid cotton.

### Materials and methods

#### Plant materials and bacterial strains

The plant materials used in this study included G. arboreum lines DPL971 (wild-type), DPL972 (near isogenic fuzzless
mutant), *G. hirsutum* lines XZ142 (wild-type), fuzzless mutant N1, and fuzzless-lintless mutant XZ142FLM, which were obtained from Germplasm Repository of Institute of Cotton Research, Chinese Academy of Agricultural Sciences (CRI of CAAS, Anyang, Henan province, China) only for scientific research purpose. Plants were grown and self-pollinated for conservation annually in accordance with standard agronomic practices at the farm of CRI. *Nicotiana benthamiana* was grown in a 26 °C incubator with 16 h light/ 8 h dark.

*Escherichia coli* DH5α competent cells were used for gene cloning. *Agrobacterium tumefaciens* strain GV3101 was used for the transformation of tobacco (*Nicotiana benthamiana*).

### Subcellular localization

Subcellular localization was performed using tobacco mesophyll cells. The CDS of *GaFzl* was cloned into binary vector pBI121-GFP to construct the GaFzl-GFP fusion protein downstream of CaMV 35S promoter. The recombinant constructs, 35S: GaFzl-GFP and the control vector pBI121- GFP, were respectively transformed into the *A. tumefaciens* strain GV3101 and inoculated into young leaves of tobacco. About 48 h later, the injected tobacco leaves were stained with 1-[3-(triethylamino)propyl]-4-[6-4-(diethylamino)phenyl]-1,3,5-hexatrienyl]pyridinium (FM4-64, a membrane-specific dye) and 4′6-diamidino-2-phenylindole (DAPI, a nucleus-specific dye) for 5 min, respectively, and afterward the fluorescence signals of fusion proteins in tobacco mesophyll cells were observed using a confocal fluorescence microscope (Leica). Primers used for gene cloning and vector construction are listed in Supplemental Table S1.

### The transcriptional activation analysis

The transcriptional activation assay of GaFzl was performed in yeast. The CDS of *GaFzl* from DPL971 and DPL972 were cloned into the bait vector pGBK7 to construct the GaFzl-BD fusion protein, respectively. Both the recombinant constructs were transformed into yeast strain Y2HGold. The transformants were further cultivated on Synthetic dropout (SD) medium (SD/-Trp, SD/-Trp/X-α-Gal, SD/-Trp/X-α-Gal/AbA and SD/-Trp/-Ade/-His/X-α-Gal/AbA).

### Identification and characterization of GaGIR genes

The protein sequence of AtGIR1 was downloaded from NCBI and subsequently used as a query to identify the GIR members in cotton genomes and search the hidden Markov model (HMM) profile from the Phytozome v12.1 database ([https://phytozome.jgi.doe.gov/pz/portal.html#](https://phytozome.jgi.doe.gov/pz/portal.html#)) and the PANTHER Classification System ([http://pantherdb.org/](http://pantherdb.org/)). We downloaded the latest versions of genome annotation data of Arabidopsis from the Arabidopsis Information Resource (TAIR; available online: [https://www.arabidopsis.org/index.jsp](https://www.arabidopsis.org/index.jsp)). The genome data of *G. hirsutum* (version: AD1_ZJU v2.1), *G. aboreum* (version: A2_CRI) and *G. raimondii* (version: D5_JGI) were downloaded from Cotton Functional Genomic Database (CottonFGD; [https://cottonfgd.org/](https://cottonfgd.org/)). The HMM profile was used to screen and identify the GIR genes in *Gossypium* by the HMMSEARCH program from HMMER 3 software. Furthermore, BLAST was also used to confirm the family members.

### Phylogenetic, structure and conserved motif analysis

The conserved region of the 21 GIR protein sequences from *G. aboreum* and 6 from *A. thaliana* were used for phylogenetic tree construction. Multiple sequence alignments were conducted using ClustalX program with default parameters. The phylogenetic tree was constructed in MEGA10 using the neighbor-joining algorithm with the P-distance model and the pairwise deletion option with 1000 bootstrap replicates.

The Gene Structure Display Server 2.0 ([http://gsds.cbi.pku.edu.cn/](http://gsds.cbi.pku.edu.cn/)) was employed to graphically visualize the exon–intron structure of GIR genes. The MEME (Multiple Expectation Maximization for Motif Elicitation) ([http://meme-suite.org/tools/meme](http://meme-suite.org/tools/meme)) program was used to predict both conserved and potential motifs of GaGIR protein sequences using the parameter settings: the minimum motif width = 6, the maximum motif width = 50, and the maximum number of motifs = 10.

### Chromosomal locations and synteny analysis

MapChart software ([http://www.earthatlas.mapchart.com/](http://www.earthatlas.mapchart.com/)) was used to visualize the distribution of GIR genes on cotton chromosomes. MCScan was utilized to analyze the synteny pairs of GIR genes of different cotton species. The synonymous substitution (Ks) and nonsynonymous substitution (Ka) rate of GIR gene pairs were calculated using the DnaSP software.

### Cis-acting element analysis and identification of transcription factor (TF) binding sites

The 2000 bp sequence upstream of the start codons of *GaGIR* genes were downloaded from Cotton Functional Genomics Database ([https://cottonfgd.org/](https://cottonfgd.org/)). The cis-acting elements were screened and predicted using the Plant Cis-Acting Regulatory Element website ([http://bioinformatics.psb.ugent.be/webtools/plantcare/html/](http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)) database, and the diagram was generated by TBtools. The TF binding sites in the promoter regions of *GaGIR* genes were analyzed.
and identified using Plant Transcription Factor Database (http://planttfdb.gao-lab.org/).

**RNA isolation and qRT-PCR analysis**

For samples collection, cotton lines DPL971, DPL972, XZ142, XZ142FLM and N1 were grown in the farm of CRI of CAAS with standard field management in Anyang, Henan province (China). Fiber-bearing ovules were harvested at −1, 0, 1, 3 and 5 days post-anthesis (DPA). The ovule samples were wrapped in foil and frozen directly in liquid nitrogen. Total RNA was extracted using RNA Prep Pure Plant kit (Tiangen) according to the kit manuals and cDNA was reverse-transcribed from 1 μg total RNA using TransScript all-in-one first-strand cDNA Synthesis SuperMix (TransGen Biotech). The qRT-PCR amplifications were conducted using TransStart TOP Green qPCR SuperMix (TransGen Biotech) on ABI Prism7500 Fast Real-time PCR System (Applied Biosystems). Gene expression levels were normalized and calculated using GaHis3 as a reference gene. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). Primers used in this study were designed using NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and qPrimerDB (https://biodb.swu.edu.cn/qprimerdb/) and synthesized by Sangon Biotech company. Primer sequences are provided in Suppl. Table S1.

**Results**

**Subcellular localization analysis of GaFzl**

In our previous study, GaFzl was fine-mapped and identified as the candidate gene controlling fuzz development in
Fig. 2 Transcriptional autoactivation of GaFzl was detected in yeast. GaFzl amplified from DPL971 and DPL972 was inserted into pGBK7 to construct expression vectors indicated on the left; Y2Hgold stains containing the expression constructs were diluted and inoculated on the SD medium indicated on the right. The number of 1, 1/10 and 1/100 indicate yeast stain without dilution, tenfold dilution and 100-fold dilution.
G. arboreum. To further explore the GaFzl function related to cotton fuzz initiation, its subcellular localization was analyzed using tobacco leaves (Nicotiana benthamiana). The GaFzl-GFP fusion protein was transiently expressed in young tobacco leaves, which were then stained using a membrane-specific dye (FM4-64) and a nucleus-specific dye (DAPI) to ascertain the localization of the GaFzl-GFP signal by confocal microscopy. Consistent with the findings of an earlier investigation by Wang et al. (2020c), the strong fluorescence of the GaFzl-GFP fusion protein was detected in the membrane and nucleus, which were stained with FM4-64 (red fluorescence) and DAPI (blue fluorescence), respectively (Fig. 1). Accordingly, GaFzl appears to be co-localized in the nuclear and membrane of cells.

Transcriptional activation analysis of GaFzl

Because AtGIR1 is a TF that regulates root hair formation as, we assessed whether GaFzl has transcriptional activation activity by conducting an autoactivation analysis of GaFzl in yeast. The GaFzl- BD (GAL4 binding domain) fusion protein in yeast activated the expression of the reporter genes MEL1(X-α-Gal) and AUR1-C (AbA), suggesting that GaFzl has transcriptional activation activity and may function as a transcriptional regulator (Fig. 2). To examine whether the transcriptional activation activity of the fusion protein could be repressed, SD agar medium (SD-Trp/-Ade/-His/X-α-Gal/AbA) was inoculated with Y2HGold cells expressing the fusion proteins. In yeast cells, the fusion protein with GaFzl from DPL972 (fuzzless isogenic line) had a higher transcriptional activation activity level than the fusion protein with GaFzl from DPL971 (wild-type), and it activated the expression of all four reporter genes. The observed difference in the transcriptional activation activity may be related to the GaFzl sequence variation between DPL971 and DPL972.

Genome-wide identification of GIR genes in G. arboreum

To thoroughly characterize the GIR gene family members in cotton, the whole-genome sequences of cotton species (G. arboreum (version: A2_CRI), G. hirsutum (version: AD1_ZJU v2.1) and G. raimondii (version: D5_JGI)) were analyzed to identify GIR genes. The Arabidopsis GIR1 protein sequence was used as the query to screen three cotton reference genomes for candidate GIR proteins, with an E-value threshold of 0.01. On the basis of the hidden Markov model (HMM) of PTHR33177, the HMMER 3 program was used to identify GIR proteins in G. arboreum. Additionally, 21, 40 and 6 GIR proteins were detected in G. raimondii, G. hirsutum, and A. thaliana, respectively (Suppl. Table S2, Table 1).
Specific information of GaGIR genes, such as gene ID, chromosomal location, protein size (AA), molecular weight (kDa) were listed in Table 1. The encoded proteins comprised 87 amino acids (Ga08G0121) to 243 amino acids (Ga07G1259), suggesting these are small proteins that may function as adapters or small molecular peptides that interact with other proteins to perform their functions.

**Phylogenetic, gene structure, and motif analyses of GIR proteins in G. arboreum**

To classify the GIR proteins and clarify their evolutionary relationships, a multiple sequence alignment analysis (Suppl. Fig. S1) was performed using 21 and 6 homologous protein sequences from *G. arboreum* and *A. thaliana*, respectively. Additionally, an unrooted phylogenetic tree was constructed according to the neighbor-joining algorithm of the MEGA10.0 software, with 1000 bootstraps replicates. In the constructed phylogenetic tree (Fig. 3), the 27 proteins were divided into three groups. Within the GaGIR family, 7 members were clustered in Group 1, 4 members belonged to Group 2 and 10 members were included in Group 3. Group 1 included GaGIR1 and two AtGIR proteins (AtGIR1 and AtGIR2), implying that these members may be associated with similar functions.

Gene structures and organization may provide additional evidence of the evolutionary relationships among GIR family members. A comparative analysis of gene structures and motifs detected 0–2 introns in the GaGIR genes. More specifically, seven genes had no intron, whereas one gene had two introns, and the remaining genes comprised only one intron (Fig. 4). Ten conserved motifs in the GaGIR proteins were identified using the MEME online software (i.e., motifs 1 to 10). Almost all GaGIR proteins had motifs 1, 2 and 3, implying these motifs may function as a highly conserved domain. Most of the Group 1 GaGIR proteins contained motif 4. In contrast, most of the Group 2 proteins had motifs 5, 6, 8, 9, and 10, whereas the Group 3 proteins had motifs 3, 5, and 7.

**Chromosomal distribution of GaGIR genes**

The GaGIR genes were mapped on cotton chromosomes to gain new insights into the organization of GaGIR genes in
the cotton genome. Using the available *G. arboreum* genome as a reference, we determined that the 21 GaGIR genes were unequally distributed among the chromosomes (Fig. 5). Seven chromosomes (chr. 2, 4, 8, 9, 10 and 12) contained only one GaGIR gene, whereas four chromosomes (chr. 1, 3, 5 and 7) had two GaGIR genes. Chromosome 13 included six GaGIR genes, four of which were tandemly distributed. However, the chromosomal distribution of the GIR genes varied among *G. arboreum*, *G. raimondii*, and *G. hirsutum* (At and Dt-subgenomes), indicative of gain and loss events during whole-genome duplications (Suppl. Fig. S2, S3, S4).

To clarify the collinearity of the GIR gene families between...
two diploid cotton ancestors and *G. hirsutum*, linked gene pairs were identified using MCScan. A total of 34 collinear gene pairs were identified between *G. hirsutum* and *G. arboreum*, 18 of which involved the A-subgenome of *G. hirsutum* (Fig. 6). Additionally, 39 collinear gene pairs were detected between *G. hirsutum* and *G. raimondii*, with 20 genes from the D-subgenome. To elucidate the evolutionary dynamics and selection pressures, we calculated the nonsynonymous (Ka) and synonymous (Ks) substitution rates of the *GIR* gene pairs. Most of the Ka/Ks ratios were less than 1.0 in the intergenomic and intragenomic analyses, implying that most *GIR* genes underwent a purifying selection in the diploid and allotetraploid cotton species (Suppl. Table S4, S5, S6).

**Cis-acting element analysis**

Because most *GIRs* remain unannotated, the putative functions of *GIRs* were first assessed by identifying the cis-acting elements in the 2-kb promoter region of the corresponding genes. In addition to the general elements (i.e., TATA-box and CAAT-box), the detected cis-acting elements were mainly enhancer activity elements and elements responsive to light, phytohormones, defense activities and abiotic stress (i.e., drought, low-temperature) (Fig. 7). Interestingly, 20 of the 21 GaGIR genes had cis-acting elements responsive to 3–6 phytohormones (ABA, followed by ethylene, MeJA, SA, GA and auxin). An analysis of the transcription factor-binding site in the promoter regions indicated the binding sites of ERF, BBR-BPC, MYB, AP2/B3, DOF and C2H2 transcription factors were extensively distributed in the GaGIR promoters (Suppl. Table S7). These results suggest that GaGIR genes likely contribute to plant growth and development through various signal transductions pathways, especially those related to phytohormone, light, and abiotic stress responses.

**Expression patterns of GaGIR genes at different stages of fiber development**

To investigate the potential functions of *GIR* genes in *G. arboreum*, we analyzed their expression patterns using the transcriptome data for different stages of fiber development (1, 3 and 5 DPA). According to their expression levels, the GaGIR genes were roughly classified as strongly expressed or barely expressed. The heatmap representation of expression revealed that the GaGIRs genes were not significantly differentially expressed between DPL971 (wild-type) and DPL972 (fuzzless isogenic line), with the exception of *GaFzl* (Fig. 8). Consistent with previous studies, the *GaFzl* expression level was much higher at 1, 3 and 5 DPA in DPL972 than in DPL971, implying that the up-regulated expression of *GaFzl* likely suppresses fuzz development. This gene is reportedly significantly up-regulated by a 6.2 kb insertion in fuzzless mutant and thus leads to a fuzzless phenotype (Wang et al. 2020c). Hence, other genes which normally expressed at low levels in wild-type cotton might also alter fuzz development if they are expressed at higher levels. To verify the accuracy and reliability of the RNA-seq results, we also examined the relative expression levels of *GIR* family genes in ovules at different stages by qRT-PCR. The results were consistent with the RNA-seq data,
confirming the transcriptome data were reliable (Fig. 9). Previous studies proved AtGIRs are involved in Arabidopsis root hair formation (Wu and Citovsky 2017a, b). To explore the functions of GaGIR genes, we examined the root phenotypes of two diploid cotton species and then compared the GaGIR expression patterns between the two materials by qRT-PCR. There were no obvious differences in the root hair phenotypes (Suppl. Fig. S5) and gene expression levels (Fig. 9). Moreover, most of the GIR genes were expressed at much lower levels in the root than in the fiber. However, Ga01G0231 was highly expressed in the root (Fig. 9), indicative of an important role in root hair formation and development.

**Regulatory relationship between GaFzl and other fiber-associated genes**

To further investigate the potential regulatory network of GaGIR genes involved in fuzz development, we analyzed the expression level of GaFzl in various fiber mutants as well as the expression of other fiber-related genes in G. arboreum by qRT-PCR. Because gene expression
levels may be influenced by allele dosage effects in polyploids, we designed sequence-specific primers to examine the expression level of the \textit{GaFzl} homolog (\textit{GhFzl}) in tetraploid cotton lines. Unlike the expression pattern in diploid cotton, \textit{GhFzl\textsubscript{A}} and \textit{GhFzl\textsubscript{D}} were expressed at very low levels (Fig. 10), with no major differences among wild-type XZ142, fuzzless mutant N1, and fiberless mutant XZ142FLM. Additionally, the genes encoding well-known fiber-related regulatory factors were similarly expressed in DPL971 and DPL972 (Fig. 11). For example, the \textit{MML3}, \textit{TTG1}, and \textit{GL2} expression levels were unaffected by changes to \textit{GaFzl} expression, indicating that the altered expression of \textit{GaFzl} had no effect on the regulatory network of the MIXTA-WD40 complex and other fiber-related genes.

**Discussion**

\textit{Arabidopsis} is a model plant system for studying cell fate determination and differentiation (Balkunde et al. 2010; Yang and Ye 2013). Because of the considerable similarity in the underlying process, studies on \textit{Arabidopsis} trichome and root hair formation may provide useful insights into cotton fiber initiation and elongation (Lee et al. 2007; Ding et al. 2014; Wan et al. 2014; Wang et al. 2019). The \textit{GIR} genes have been identified in diverse monocotyledonous and dicotyledonous plant species. The \textit{GIR} genes in \textit{Arabidopsis} are involved in the GL2-mediated control of root hair development (Wu and Citovsky 2017a). In tomato, \textit{SiCycB2} encodes a GIR1/2, homolog to GL2, reportedly interacting with Wo, to regulate trichome formation and development (Yang et al. 2011). Several recent studies demonstrated that \textit{GaFzl}, annotated as \textit{GaGIR1}, is closely associated with the fuzzless phenotype of diploid cotton. It encodes a negative regulator of trichome formation and fuzz development, but it does not affect root hair formation (Du et al. 2018; Feng et al. 2019; Liu et al. 2020; Wang et al. 2020c). These studies suggest that the functions of these homologs may have diverged slightly as the different species evolved. However, very few of the \textit{GIR} genes in cotton have been identified and functionally annotated. Thus, it is important that these genes in cotton are identified so their expression patterns and functions related to cotton fiber development may be determined.

On the basis of a whole-genome analysis, we identified 21, 21, and 40 \textit{GIR} family genes in \textit{G. arboreum}, \textit{G. raimondii} and \textit{G. hirsutum}, respectively. The \textit{GaGIRs} genes, which were divided into three groups, were unevenly distributed on 12 of 13 chromosomes. Additionally, the encoded proteins varied substantially in terms of amino acid sequence length and molecular weight. Analyses of the phylogenetic relationships, structures, and motifs revealed that most of the \textit{GIR} proteins in the same subgroup were similar regarding the organization of exons and conserved motifs, whereas obvious differences were detected between subgroups. These differences may reflect some functional diversification during
Moreover, some conserved motifs were identified in all GIR proteins, whereas others were subgroup-specific, suggesting that motif diversity might be related to functional diversity. The gain or loss of key motifs as species evolved may have resulted in changes to protein functions that altered plant development. Phytohormones and abiotic stresses might affect fiber development through signal transduction pathways (Chen et al. 2019; He et al. 2019; Cheng et al. 2020; Wang et al. 2020a, b; Zhang et al. 2020; Tian and Zhang 2021; Wu et al. 2021). We identified numerous cis-acting elements responsive to light, phytohormones, defense processes, and stress as well as multiple TF-binding sites upstream of transcriptional start sites. Accordingly, GIR proteins may participate in signal transduction pathways that help regulate fiber and plant development.

We compared the GaGIRs expression patterns between wild-type and fuzzless mutant G. arboreum at different stages of fiber development. Some of the genes were highly expressed at multiple fiber developmental stages, indicating they may positively regulate fiber formation. In contrast, some GaGIR genes were expressed at relatively low levels, suggesting that they are not directly involved in fiber development or they may negatively regulate fiber formation.

There were no obvious differences in the expression of most of these genes between the wild-type and mutant cotton. Interestingly, GaFzl was significantly expressed in the fuzzless mutant but not in the wild-type control, during the fiber and fuzz initiation stages (0, 1, 3 DPA) (Fig. 10). This is in accordance with the results of previous studies (Feng et al. 2019; Liu et al. 2020). This gene was annotated as GaGIR1, and its upregulated expression is associated with the fuzzless phenotype in G. arboreum. The significant difference in the expression of this gene between the wild-type and mutant cotton plants may be caused by a ~ 6.2-kb insertion (Copy Number Variation) in the upstream region. Our analyses of phylogenetic relationships and gene expression patterns indicated that five Group 1 members that were clustered with GaFzl were expressed at low levels during the fiber and fuzz initiation stages. We hypothesize that overexpressing GaFzl could lead to a fuzzless phenotype in cotton.
the genes classified in the same group as GaFzl with the same expression patterns may similarly affect the fuzz initiation stage, ultimately resulting in the fuzzless phenotype. Because the functions of most of these genes are unknown, they will need to be more thoroughly investigated in future studies.

There were no obvious differences in GhFzl expression among G. hirsutum XZ142, the fuzzless mutant N1, and the lintless–fuzzless mutant XZ142FLM, which is in contrast to the corresponding expression patterns in G. arboreum. Furthermore, there was no significant difference between G. arboreum DPL971 and DPL972 regarding the expression patterns of the core genes involved in fiber and fuzz development (i.e., MML3, TTG1, and CPC). Thus, the mechanism regulating GaFzl expression in G. arboreum may not be associated with the well-known MYB-bHLH-WD40 complex involved in the fiber/fuzz initiation and development of G. hirsutum. This finding may lead to the development of new ways to specifically regulate fiber/fuzz formation in the initial steps of this process in G. arboreum.

In previous studies, GaFzl was identified as a candidate gene controlling fuzz initiation and development in G. arboreum. Liu et al. (2020) suggested that the variability in the sequences of the GaGIR1 haplotypes in diverse fuzzless mutants may be related to the changes in expression patterns or gene functions, while Wang et al. (2020c) identified an enhancer as the essential element for controlling gene expression and fuzz development. To explore the potential function of GaFzl during fuzz initiation, we conducted a transcriptional activation analysis in yeast, which demonstrated that GaFzl has strong transcriptional activation activity and may act as a transcriptional activator. However, the fusion proteins constructed from two materials revealed minor differences in the transcriptional activation activity in yeast cells. Because of the differences in the GaFzl coding sequence, we speculated that sequence variations between two parental lines may be responsible for the differences in transcriptional activation activities or even gene functions. This possibility will need to be experimentally verified in future studies.
Author contribution statement  Research conception and design, GLS and JY; Data collection, XXF; Data analysis, XXF, HLC, DYZ, YPZ, QLW, LML and SYL; Research management, GLS; Writing—initial draft, XXF; Writing—revised manuscript, JY. All authors read and approved the final manuscript.

Supplementary Information  The online version contains supplementary material available at https://doi.org/10.1007/s00425-021-03737-7.

Fig. 11  Expression analysis of several core genes in *G. arboreum* DPL971 and DPL972 in upland cotton fiber development. *GhHis3* was applied as the internal control. The expression value of fiber samples in DPL971 at −1 DPA was set as 1. Data are presented as mean ± standard deviation (*n* = 3).

Acknowledgements  The authors thank Lima Soares Emanoella from Plant Genetics Laboratory of Gembloux Agro Bio-Tech, University of Liège for kindly revising the draft.

Funding  This work was funded by grants from the National Natural Science Foundation of China (No. 31621005), the National Natural Science Foundation of China (No. 31901581), the National Key R & D Plan of China (No. 2018YFD0100402) and the United States Department of Agriculture—Agricultural Research Service (USDA-ARS Project No. 3091-21000044-00D). The funders had no role in the experimental design, data collection and analysis, or in writing work.
Data availability statements  All related datasets supporting the results of this study are available within the manuscript and its supplementary files.

Declarations

Conflict of interest  The authors declare that they have no conflicts of interest.

References


Tian Y, Zhang T (2021) MIXTA and phytohormones orchestrate cot-
Thyssen GN, Fang DD, Li P, Florane CB (2021b) Elu-
Ohashi Y, Oka A, Ruberti I, Morelli G, Aoyama T (2020a) Mapping-by-
Wang Y, Yu Y, Chen Q, Bai G, Gao W, Qu Y, Ni Z (2019) Hetero-
GhTCP4, a Class II TCP transcription factor, regulates trichome formation and root hair development in Arabidopsis. Genes 10(9):726
Wu R, Citovsky V (2017a) Adaptor proteins GIR1 and GIR2. I. Interaction with the co-repressor GLABRA2 and regulation of root hair development. Biochim Biophys Res Commun 488(3):547–553
Wu R, Citovsky V (2017b) Adaptor proteins GIR1 and GIR2. II. Interaction with the co-repressor TOPESS and promotion of histone deacetylation of target chromatin. Biochim Biophys Res Commun 488(4):609–613


**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.