



# Analysis of the *MIR160* gene family and the role of *MIR160a\_A05* in regulating fiber length in cotton

Guoyuan Liu<sup>1</sup> · Ji Liu<sup>1</sup> · Wenfeng Pei<sup>2</sup> · Xihua Li<sup>1</sup> · Nuohan Wang<sup>1</sup> · Jianjiang Ma<sup>1</sup> · Xinshan Zang<sup>1</sup> · Jinfa Zhang<sup>3</sup> · Shuxun Yu<sup>1</sup> · Man Wu<sup>1</sup> · Jiwen Yu<sup>1</sup>

Received: 23 February 2019 / Accepted: 5 September 2019 / Published online: 16 October 2019  
© The Author(s) 2019

## Abstract

**Main conclusion** The *MIR160* family in *Gossypium hirsutum* and *G. barbadense* was characterized, and *miR160a\_A05* was found to increase cotton-fiber length by downregulating its target gene (*ARF17*) and several *GH3* genes.

**Abstract** Cotton fiber is the most important raw material for the textile industry. MicroRNAs are involved in regulating cotton-fiber development, but a role in fiber elongation has not been demonstrated. In this study, *miR160a* was found to be differentially expressed in elongating fibers between two interspecific (between *Gossypium hirsutum* and *G. barbadense*) backcross inbred lines (BILs) with different fiber lengths. The gene *MIR160* colocalized with a previously mapped fiber-length quantitative trait locus. Its target gene *ARF17* was differentially expressed between the two BILs during fiber elongation, but in the inverse fashion. Bioinformatics was used to analyze the *MIR160* family in both *G. hirsutum* and *G. barbadense*. Moreover, qRT-PCR analysis identified *MIR160a* as the functional *MIR160* gene encoding the *miR160a* precursor during fiber elongation. Using virus-induced gene silencing and overexpression, overexpressed *MIR160a\_A05* resulted in significantly longer fibers compared with wild type, whereas suppression of *miR160* resulted in significantly shorter fibers. Expression levels of the target gene *auxin-response factor 17* (*ARF17*) and related genes *GH3* in the two BILs and/or the virus-infected plants demonstrated similar changes in response to modulation of *miR160a* level. Finally, overexpression or suppression of *miR160* increased or decreased, respectively, the cellular level of indole-3-acetic acid, which is involved in fiber elongation. These results describe a specific regulatory mechanism for fiber elongation in cotton that can be utilized for future crop improvement.

**Keywords** Backcross inbred lines (BILs) · Fiber elongation · *Gossypium* · *MIR160* · *ARF17* · Virus-induced gene silencing (VIGS)

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00425-019-03271-7>) contains supplementary material, which is available to authorized users.

- ✉ Shuxun Yu  
yu@cricaas.com.cn; ysx195311@163.com
- ✉ Man Wu  
wuman@cricaas.com.cn; wuman2004@163.com
- ✉ Jiwen Yu  
yujw@cricaas.com.cn; yujw666@hotmail.com

<sup>1</sup> Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Key Laboratory for Cotton Genetic Improvement, Ministry of Agriculture, Zhengzhou University, Anyang 455000, Henan, China

## Abbreviations

- ARF Auxin-response factor  
BIL Backcross inbred line  
CLCrV Cotton leaf crumple virus

- <sup>2</sup> Xinjiang Research Base, State Key Laboratory of Cotton Biology, Xinjiang Agricultural University, Urumqi 830001, China
- <sup>3</sup> Department of Plant and Environmental Sciences, New Mexico State University, Box 30003, Las Cruces, NM 88003, USA

DPA	Days post-anthesis
FL	Fiber length
STTMs	Short tandem target mimics
VIGS	Virus-induced gene silencing

## Introduction

Cotton fiber is the most important source material for the textile industry (Chen et al. 2007) and is an ideal model system for cell elongation studies (Kim and Triplett 2001). Fiber length (FL), an important index of fiber quality, is determined from fiber-cell initialization to 21–26 days post-anthesis (DPA) (Lee et al. 2007). Several genes have been shown to affect fiber-cell development, including K<sup>+</sup> transporters (Ruan et al. 2004), sucrose-encoding genes (Ruan et al. 2001), and reactive oxygen species-regulated Ca<sup>2+</sup> signaling and the vacuolar (H<sup>+</sup>)-ATPase (Qin and Zhu 2011). In addition, phytohormones such as auxin (Beasley and Ting 1974; Shindy and Smith 1975; Zhang et al. 2011), gibberellin (Beasley and Ting 1974; Gokani and Thaker 2002; Zhang et al. 2011), ethylene (Shi et al. 2006), brassinosteroid (Sun et al. 2004, 2005; Sun and Allen 2005; Shi et al. 2006), and cytokinins (Beasley and Ting 1974; Shindy and Smith 1975) play important roles in fiber development. Indole-3-acetic acid (IAA), which is considered the most important auxin, is associated with fiber initiation (Zhang et al. 2011) and elongation (Shindy and Smith 1975). IAA affects fiber-cell elongation in cotton ovules in vitro (Beasley and Ting 1974), and IAA and related auxin metabolites are found in eight DPA fibers (Guinn and Brummett 1988).

Certain microRNAs (miRNAs), a type of small RNA, have also been shown to regulate fiber development (Liu et al. 2019). Such miRNAs have been identified based on sequencing from developing fibers of wild-type and mutant strains, including upland cotton Xuzhou 142 and its fuzzless/lintless mutant (Kwak et al. 2009; Li et al. 2012), TM-1 and its lintless mutant (Pang et al. 2009), CRI35 (Xue et al. 2013), the short-fiber mutants *Li1* and *Li2* (Naoumkina et al. 2016), and Pima 3-79 (Liu et al. 2014; Hu et al. 2015). However, only a few studies have focused on the specific function of these miRNAs in cotton. For example, the miRNAs miR828 and miR858 have been shown to regulate *MYB2* and affect fiber initiation (Guan et al. 2014), and miR157 affects FL by regulating *SQUAMOSA promoter-binding protein-like (SPL)* genes (Liu et al. 2017). A high level of miR160 leads to defects in embryonic, root, vegetative, and floral development in *Arabidopsis* by regulating *auxin response factor (ARF)* genes (Mallory et al. 2005; Wang et al. 2005; Wu et al. 2006) and results in narrow leaflet blades, sepals, and petals and abnormally shaped fruit in *Solanum lycopersicum*

(Hendelman et al. 2012). In cotton, miR160 regulates anther indehiscence under high-temperature conditions (Ding et al. 2017), but a relationship between miR160 and fiber development has not been reported.

This study analyzed the *MIR160* family in *G. hirsutum* and *G. barbadense*. In our previous studies, a backcross inbred line (BIL) population was generated from *G. hirsutum* and *G. barbadense* and used to detect FL quantitative trait loci (QTL) (Yu et al. 2013). Two BILs with significantly altered FL were selected here for comparative small-RNA transcriptome analysis, and *MIR160a\_A05* was found to colocalize with an FL QTL on chromosome A05 in the BIL population (Liu et al. 2019). Based on bioinformatics and quantitative reverse transcription-PCR (qRT-PCR), a comparative analysis was conducted of miR160a and its target *ARF17* in elongating fibers at 10 DPA. Virus-induced gene silencing (VIGS) was then used to validate the effect of miR160a on FL. The miR160a was identified as a candidate miRNA for determining FL in two interspecific BILs by regulating differential expression of its target gene during the fiber-elongation stage.

## Materials and methods

### Plant materials

Two BILs (“Long” and “Short”) with significantly altered FL were chosen from a previously constructed population of 146 progeny lines (Yu et al. 2013) to identify FL-related genes (Li et al. 2017) and miRNAs (Liu et al. 2019). The two BILs and their parent lines (*G. hirsutum* SG747 and *G. barbadense* Giza 75) were grown on an experimental farm at the Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan, China, in 2014. Each line was planted in three biological replicates with 20 plants per replicate. For the two BILs, developing ovules at 0 and 3 DPA and fibers at 5, 10, 15, 20, and 25 DPA were collected from each replicate during the early morning. For the parent lines, developing ovules at 0 DPA and fibers at 5, 10, 15, and 20 DPA were collected from each replicate during the early morning. All collected ovules and fibers were stored at –80 °C.

### Identification of *MIR160* family members

The mature miR160 sequences identified in our previous study (Liu et al. 2019) were first mapped onto the genome of Upland cotton TM-1 and *G. barbadense* 3–79 with no more than two mismatches (Zhang et al. 2015; Wang et al. 2019).

Mireap\_0.2 (<http://sourceforge.net/projects/mireap>) was then used to predict whether the miRNA precursor sequences could form hairpin structures as described by Meyers et al. (2008). Alignment and phylogenetic tree construction were performed using the ClustalW and neighbor-joining methods, respectively, with Mega X software.

Based on sequences of simple-sequence repeats and miRNA precursors, 4 FL QTL identified in our previous study (Yu et al. 2013) and 13 FL QTL hotspots from the Cotton QTL Database (CottonQTLdb version 2.3, [www.cottonqtl.org](http://www.cottonqtl.org)) reported by Said et al. (2015) were first mapped onto the sequenced TM-1 genome by performing a BLASTN search. The identified *MIR160* genes were then localized on the TM-1 genome to identify the colocalized *MIR160* genes.

### qRT-PCR of mature miRNAs, pri-miRNAs, and target gene

Total RNA and miRNAs from 0 and 3 DPA ovules and 5–25 DPA fibers were extracted using the Spectrum Plant Total RNA kit (Sigma) and mirPremier microRNA Isolation kit (Sigma), respectively. Expression levels of the miRNA and its targets in each biological replicate were measured by qRT-PCR using three technological replicates. To measure the level of mature miR160, the TaqMan<sup>®</sup> MicroRNA Reverse Transcription kit (Applied Biosystems) was used for stem-loop RT-PCR. The reaction was performed at 16 °C for 30 min, followed by 42 °C for 30 min and 85 °C for 5 min. The PrimeScript<sup>™</sup> RT Reagent kit (Takara) was used to perform RT-PCR of the target gene and the primary miR160 (pri-miR160). Reverse transcription of the target gene was performed at 37 °C for 30 min and 85 °C for 5 s. The One-Step SYBR Primer Script PLUS RT-PCR kit (Takara) was used for qRT-PCR of mature miR160, pri-miR160, and the target gene. The reaction was performed at 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 s, 56 °C for 30 s, and 72 °C for 20 s.

In the qRT-PCR reactions, the cotton U6 snRNA was used as the internal control for miRNA and pri-miRNAs, whereas *His3* (i.e., *histone 3*) served as the internal control for the target. Relative expression was measured using the  $2^{-\Delta\Delta C_t}$  method; the expression level in “Long” or Giza 75 was selected as the reference sample (with a relative expression of 1.0). Primers are listed in Table S1.

### Analysis of sequencing data

Sequencing reads from small RNAs, the degradome, and the transcriptome were produced using an Illumina HiSeq 2000 sequencer at BGI (Shenzhen, China). Small-RNA sequencing data from 10 DPA fibers from “Long” and “Short”

were deposited in the National Center for Biotechnology Information (NCBI) with accession number SRX743622. Using SOAP2, the clean reads were mapped onto the mature sequence of miR160 family members with no mismatch and no reverse complementation (Li et al. 2009). Expression was normalized by the tags per million method. Similarly, transcriptome sequencing data for the 10 DPA fibers from “Long” and “Short” were deposited in the NCBI with accession number SRP039385, and SOAP2 was again used to align clean reads to the TM-1 genome (<http://cgp.genomics.org.cn/page/species/index.jsp>). The expression of each gene was calculated as fragments per kilobase transcriptome per million mapped reads (FPKM), and the absolute value of the log<sub>2</sub>-fold change  $\geq 1$  and threshold FDR  $\leq 0.001$  were chosen as criteria to select differentially expressed genes. To identify the group II glycoside hydrolase family 3 (*GH3*) members in Upland cotton, sequences from *GH3.2*, *GH3.3*, *GH3.5*, and *GH3.6* of *Arabidopsis* (<https://www.arabidopsis.org>) were aligned to Upland cotton genes using a BLASTN search with a threshold *E* value of  $10^{-10}$ . The Hidden Markov Model profile of the GH3 auxin-responsive promoter (PF03321) was downloaded from the Pfam database (<http://pfam.xfam.org/>) as the query. HMMER3.1 was used to search whether proteins contained PF03321. Expression levels of these genes were normalized by FPKM, and the pheatmap package was used to generate the heatmap.

### Use of VIGS and overexpression of *miR160a\_A05*

To perform VIGS and overexpression experiments, CLCrV-based vectors (containing pCLCrV-A and pCLCrV-B) were received from Dr. Xue-Ping Zhou’s group (Gu et al. 2014). It had already been verified that the vector could affect gene expression in developing bolls (Gu et al. 2014) and fibers (Liu et al. 2015). Moreover, it was shown that a vector constructed with a miRNA precursor could result in high expression of the mature miRNA and that a vector with an artificial small tandem target mimic (STTM) could lead to low expression of the mature miRNA (Yan et al. 2012; Gu et al. 2014). Vector pCLCrV-A was cut by restriction enzymes *SpeI* and *AscI*. To create the miR160 overexpression vector pCLCrV-OX-miR160a, the precursor of *MIR160a\_A05* was amplified from total genomic DNA by PCR, verified by sequencing, and then introduced into the vector using the One-Step Cloning kit (Vazyme, Nanjing, China). For the miR160 suppression vector pCLCrV-STTM-miR160a, the STTM sequence of miR160a was designed and synthesized based on Yan et al. (2012) and then introduced into the vector using the same kit. A diagram of the vectors is shown in Fig. S1, and the primers are listed in Table S1.

Vectors were individually introduced into *Agrobacterium tumefaciens* strain LBA4404 by electroporation. Transformed *Agrobacterium* cells were cultured to an  $OD_{600}$  of 1.5–2.0, then pelleted, resuspended in transformation buffer to an  $OD_{600}$  of 1, and allowed to stand for no less than 3 h at room temperature. The transformation buffer contained 10 mM  $MgCl_2$ , 10 mM MES, and 200  $\mu$ M acetosyringone.

Upland cotton CRI 24 was selected as the virus receptor and was infiltrated based on Gu et al. (2014) and Liu et al. (2015). Prepared *Agrobacterium* solutions were infiltrated through needleless syringes into cotyledons of 2-week-old cotton seedlings. A month after infection, total DNA was extracted from newly emerged true leaves of infected cotton plants, and PCR was performed to determine whether vector CLCrV was successfully inoculated into the plants to confirm that no less than 15 cotton plants were infiltrated by each vector. PCR was performed at 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 40 s. Primers are listed in Table S1 and were based on Gu et al. (2014). Confirmed CLCrV-inoculated plants were transplanted in the greenhouse in Sanya, China, infiltrated in September 2015, and harvested in March 2016. Developing fibers of 2–3 bolls were collected from each plant at 10 DPA in the early morning and stored at –80 °C.

### Measurement of IAA by high-performance liquid chromatography (HPLC)

Free IAA was extracted from 10 DPA fibers using the modified method of Gou et al. (2010). A total of 0.5 g fibers were homogenized in liquid nitrogen, followed by extraction of IAA with 80% methanol (v/v). The extract was evaporated, redissolved into the aqueous phase, and then purified on a Sep-Pak Plus tC18 cartridge (Waters) (Zhang et al. 2011) using 10 ng  $^{13}C_6$ -IAA as the internal standard. Samples were then dried and stored at –20 °C. Each sample was redissolved in 500  $\mu$ L methanol and subjected to HPLC to determine the IAA content in three technological replicates. For each sample, 20  $\mu$ L was loaded onto a 5- $\mu$ m C18 column (150 $\times$ 4.6 mm; Welch, Ultimate). A linear gradient with increasing methanol (solvent A) against acetonitrile (solvent B) at a flow rate of 1 mL min<sup>-1</sup> over 30 min was used to separate IAA at 35 °C. Both solutions contained 0.02 M  $H_3PO_4$ . Each peak on the chromatogram was monitored at 210 nm.

### Measurement of FL

To measure FL of VIGS plants, bolls having no less than 2 g lint were chosen from the middle fruiting branches of each plant, and more than ten plants infiltrated by each vector were measured. The Advanced Fiber Information System Indexes (Kelly et al. 2012) were taken to measure FL for each boll at the Test Centre of Cotton Fiber Quality affiliated with the Agriculture Ministry of China, Anyang, China.

## Results

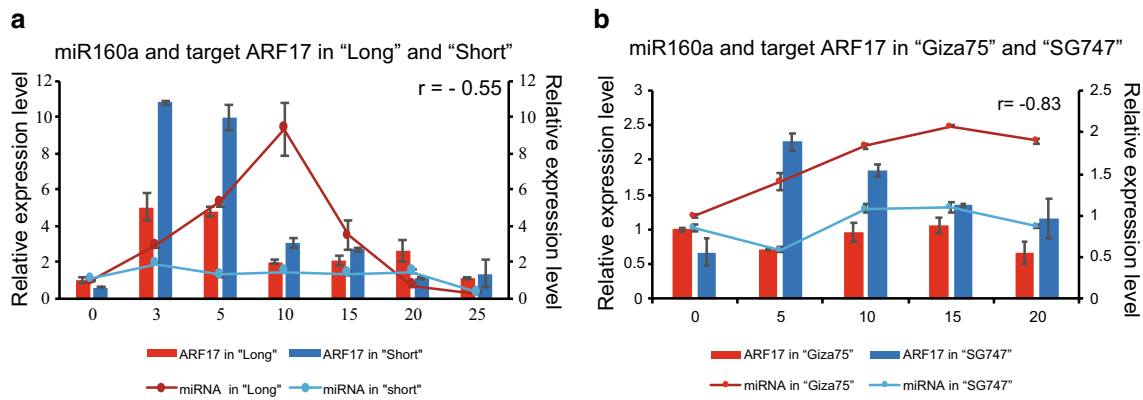
### Identification and verification of miR160a and its target ARF17 in elongating fiber cells in two BILs

We previously sequenced miRNAs from 10 DPA fibers of “Long” and “Short” BILs (Liu et al. 2019). A total of 463 miRNAs were identified (380 known and 83 novel), including 47 that were differentially expressed between “Long” and “Short”. Among these, miR160a colocalized with an FL QTL and a QTL hotspot that mapped in the BIL population, including the “Long” and “Short” lines, onto chromosome A05 of the sequenced TM-1 genome (Yu et al. 2013). In addition, a target gene (*ARF17*, *CotAD\_42076*) responsive to auxin was identified and verified using degradome analysis and RNA ligase-mediated rapid amplification of the cDNA ends (RLM-RACE), to be cleaved by miR160a between nucleotides 10 and 11.

Here, a dynamic analysis of the expression of miR160a and its target *ARF17* was performed in the two BILs and their parent lines (*G. hirsutum* SG747 and *G. barbadense* Giza 75) using developing cotton ovules at 0 and 3 DPA (fiber initiation) and in developing fibers at 5, 10, 15, 20, and 25 DPA. As shown in Fig. 1a, expression of miR160a remained at a low and stable level in “Short” from 0 DPA ovules to 25 DPA fibers. In contrast, its expression in “Long” was greatly increased in 5 DPA fibers, reaching its highest level in 10 DPA fibers, and then decreasing. At 15 DPA, expression was significantly higher in fibers from “Long” than in fibers from “Short”. Expression of the target gene *ARF17* showed the opposite trend (as indicated by the negative correlation coefficient of –0.55). In Fig. 1b, as expected, miR160a was expressed at a higher level in the longer-fibered *G. barbadense* Giza 75 than in *G. hirsutum* SG747, whereas the target gene *ARF17* was expressed at a lower level in Giza 75 than in SG747 (negative correlation coefficient –0.83). Given that the fiber-elongation stage lasts longer in *G. barbadense* than in *G. hirsutum* (Hawkins and Serviss 1930), it is understandable that the expression profiles of miR160a and *ARF17* in “Long” and its *G. barbadense* parent were similar. Based on these observations, miR160a was analyzed further.

### Genome-wide analysis of the MIR160 family in *G. hirsutum* and *G. barbadense*

We then analyzed the distribution of members of the *MIR160* family in the sequenced genomes of *G. hirsutum* and *G. barbadense*. A total of 30 and 29 loci were identified, respectively, containing mature miR160 sequences that each had no more than two mismatches. Among



**Fig. 1** Cellular levels of miR160a and expression of OR Expression levels of miR160a and its target gene *ARF17* in developing ovules and fibers of **a** BILs "Long" and "Short" and **b** parent strains Giza 75 and SG747. Error bars denote standard error (SE,  $n = 3$ ). The X axis

indicates the different stages during fiber elongation (DPA, days post-anthesis); the Y axis shows the expression levels of the targets (left) and miRNAs (right)

these, 4 loci containing no mismatch and 19 loci containing one mismatch were identified in both *G. hirsutum* and *G. barbadense*, and 7 and 6 loci with two mismatches were identified, respectively. The stem-loop structures of the precursors were then predicted to verify whether the mature miRNA sequences located on the stem contained no more than four mismatches (Meyers et al. 2008). From this analysis, 18 and 17 *MIR160* genes belonging to the miR160 family were identified in *G. hirsutum* and *G. barbadense*, respectively. In *G. hirsutum*, both the At and Dt subgenomes contained 9 *MIR160* family members. In *G. barbadense*, 8 and 9 *MIR160* family members were identified in the At and Dt subgenomes, respectively.

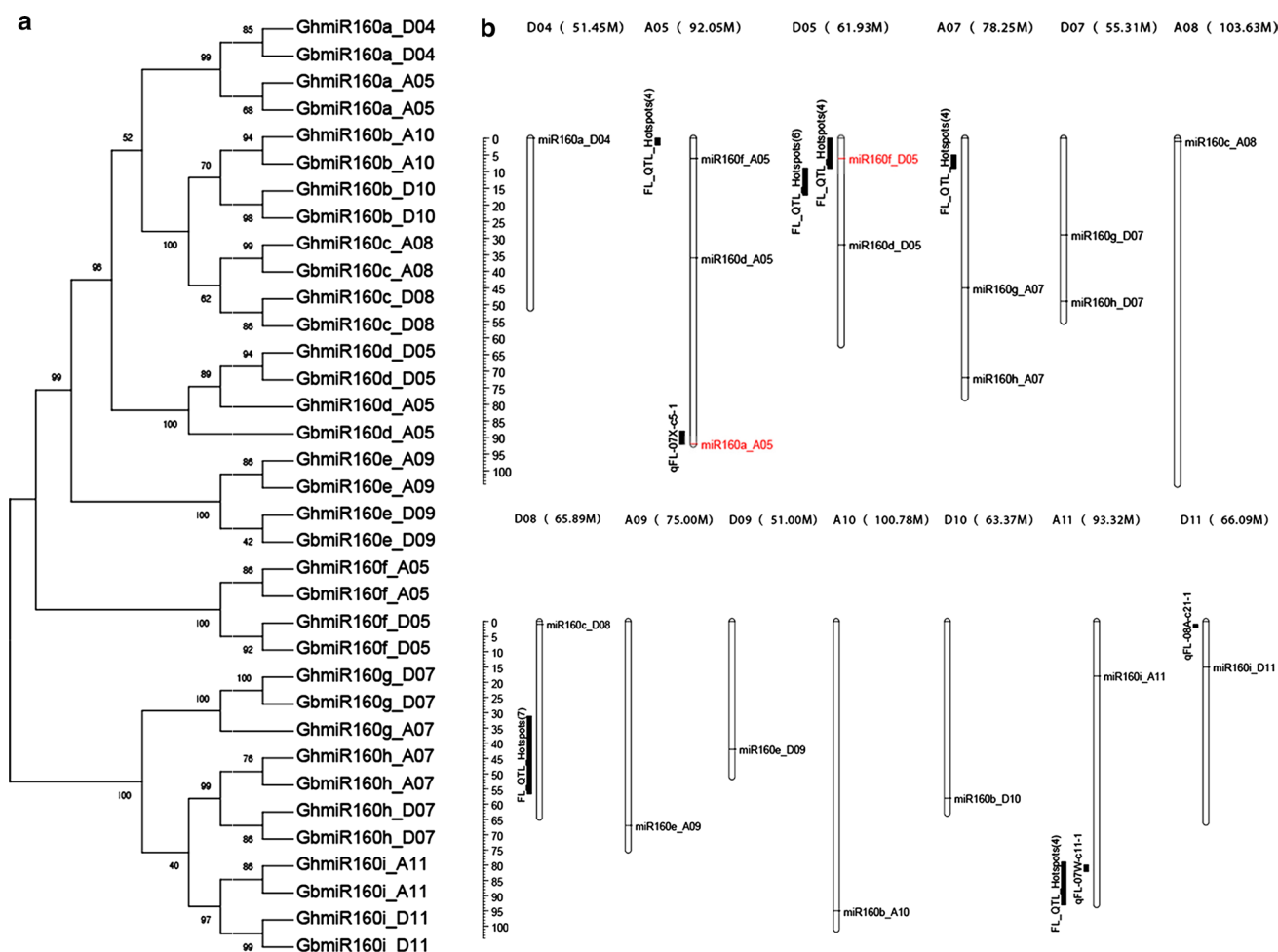
We then compared the 35 precursor sequences of *MIR160* genes in *G. hirsutum* and *G. barbadense* and divided them into nine subfamilies (Fig. 2a). The mature miRNA at the 5' terminus of each of miR160a, miR160b, miR160c, miR160d, and miR160e contained the same sequence (Fig. S2). A single-nucleotide polymorphism was found at position 15 of the mature sequences in miR160f, miR160g, miR160h, and miR160i. *MIR160f* and *MIR160h* also contained the same mature miRNA sequence. Four members (corresponding to the At and Dt subgenomes in *G. hirsutum* and *G. barbadense*) were identified in each *MIR160* subfamily, with the exception of the *MIR160g* subfamily. *MIR160g* of the At subgenome was missing in *G. barbadense* due to the lack of an *MIR160g* gene and mature miR160g sequence (Fig. 2a). These results revealed that the precursors of *MIR160* in each subfamily are conserved between both species and subgenomes (Fig. S2).

All 18 *MIR160* family members in Upland cotton were further mapped onto the 26 chromosomes of the sequenced TM-1 Upland cotton genome. After colocalizing all the *MIR160* genes with FL QTLs and FL QTL hotspots on the TM-1 genome, two *MIR160* genes colocalized with an FL

QTL or FL QTL hotspot on two chromosomes, A05 and D05 (Fig. 2b), and were dubbed *MIR160a\_A05* and *MIR160f\_D05*. As shown in Fig. 2b, *MIR160a\_A05* colocalized with an FL QTL on chromosome A05 identified in our previous study using the BIL population, and *MIR160f\_D05* colocalized with an FL QTL hotspot on chromosome D05.

### Identification of *MIR160* genes involved in fiber elongation

To identify which *MIR160* gene affects mature miRNA expression in the two BILs, we attempted to measure the expression of all precursors in 10 DPA fibers using RT-PCR. However, *MIR160* homologs in different subgenomes were too similar to distinguish. Thus, precursor expressions of only different subfamilies were measured. As shown in Fig. 3a, only pri-mir160a and pri-mir160h were detected in "Long" and "Short". When quantified (Fig. 3b), the expression of pri-mir160a was much higher in "Long" than in "Short", whereas no significant difference was found for pri-mir160h. We then calculated the counts and tags per million of different mature miR160s in the small-RNA sequencing data from 10 DPA fibers. Four types of miR160 were detected (Fig. 3c). Type I included mature miR160a to miR160e, which have the same mature miRNA sequence, and type II included miR160f and miR160h, which also share a mature miRNA sequence. Types III and IV represented mature miR160g and miR160i, respectively. The miR160 in type I was expressed at a significantly higher level than the other types. Furthermore, it was differentially expressed between "Long" and "Short", whereas the other types were not. Although the sequencing analysis could not determine which precursor yielded the type I miR160, combining these results with the RT-PCR and qRT-PCR results (Fig. 3a, b) indicated that pri-mir160a is likely the most



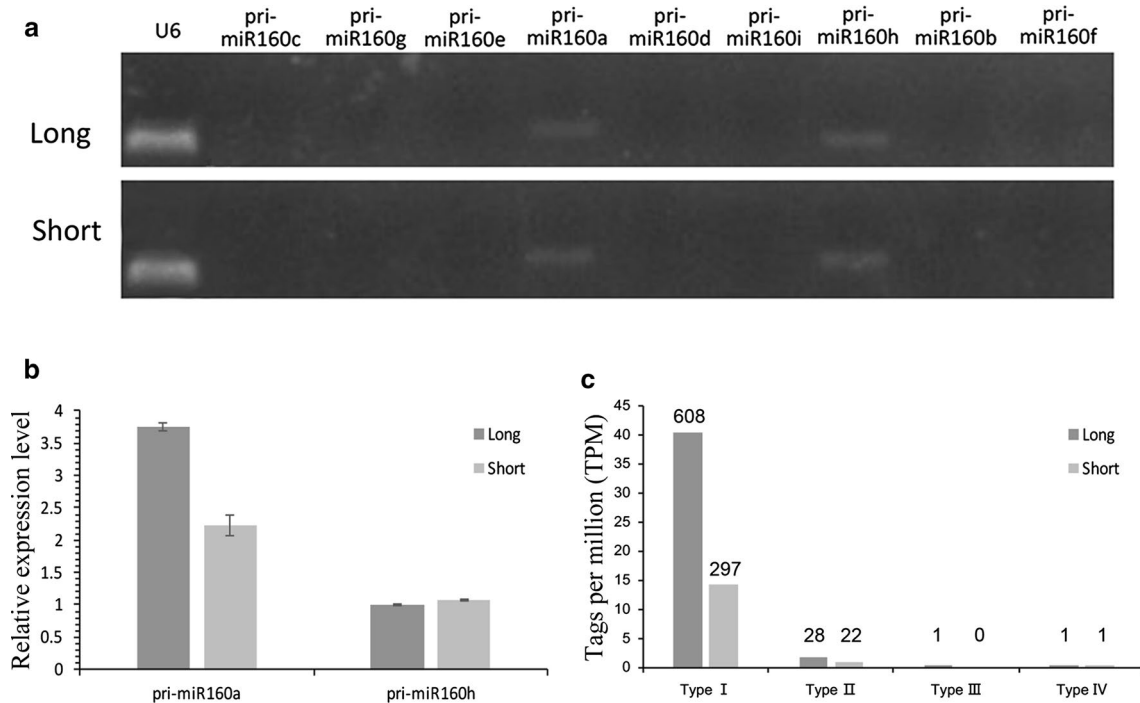
**Fig. 2** Genome-wide analysis and distribution of the *MIR160* family. **a** Phylogenetic tree of *MIR160* genes showing nine subfamilies. **b** Chromosomal location of the *MIR160* FL QTL and FL QTL hotspots in the Upland cotton genome. Colocalized miRNAs are marked in red

functional precursor. This may partly explain the differential expression of miR160a (i.e., miR160 in type I) between the two BILs. There are two members in the *MIR160a* subfamily in both *G. hirsutum* and *G. barbadense* located on chromosomes D04 and A05 (Fig. 2a, b). Given that *MIR160a\_A05* colocalized with a previously identified FL QTL, *MIR160a\_A05* was selected for further analysis.

### Overexpression of *MIR160a\_A05* promotes fiber elongation, and suppression of miR160a expression inhibits fiber elongation

We constructed the overexpression vector pCLCrV-OX-miR160a and suppression vector pCLCrV-STTM-miR160a to test whether the precursor *MIR160a\_A05* affects fiber elongation in “Long”. As shown in Fig. 4a, the photobleaching phenotype was seen in new leaves and bolls in pCLCrV:*GhPDS*-infected plants, indicating that the vector was capable of affecting gene expression at

least until boll maturity. The height of infected plants was similar to that of vector control plants (Fig. 4a), but FL differed significantly. Overexpression of *MIR160a\_A05* in pCLCrV-OX-miR160a-infected plants resulted in fibers that were on average 1.8 mm longer than in vector control plants. On the other hand, suppression of the miRNA in pCLCrV-STTM-miR160a-infected plants resulted in fibers that were on average 1.0 mm shorter than in vector control plants (Fig. 4a). Expression analysis of mature miR160a and the target gene *ARF17* (*CotAD\_42076*) using qRT-PCR revealed an approximately 40% increase of mature miR160a in 10 DPA fibers of plants infected with pCLCrV-OX-miR160a and a 55% reduction in pCLCrV-STTM-miR160a-infected plants (Fig. 4b). As expected, the expression of *ARF17* (*CotAD\_42076*) was significantly reduced in pCLCrV-OX-miR160a-infected plants and significantly elevated in pCLCrV-STTM-miR160a plants compared with vector control plants (Fig. 4c).



**Fig. 3** Expression analysis of mature miR160 and its precursors in 10 DPA Upland cotton fibers “Long” and “Short”. **a** RT-PCR analysis of pri-miR160 precursor expression. **b** qRT-PCR analysis of pri-miR160a and pri-miR160h precursor expression. **c** Expression of mature miR160 based on sequencing data. Type I, mature miR160a–

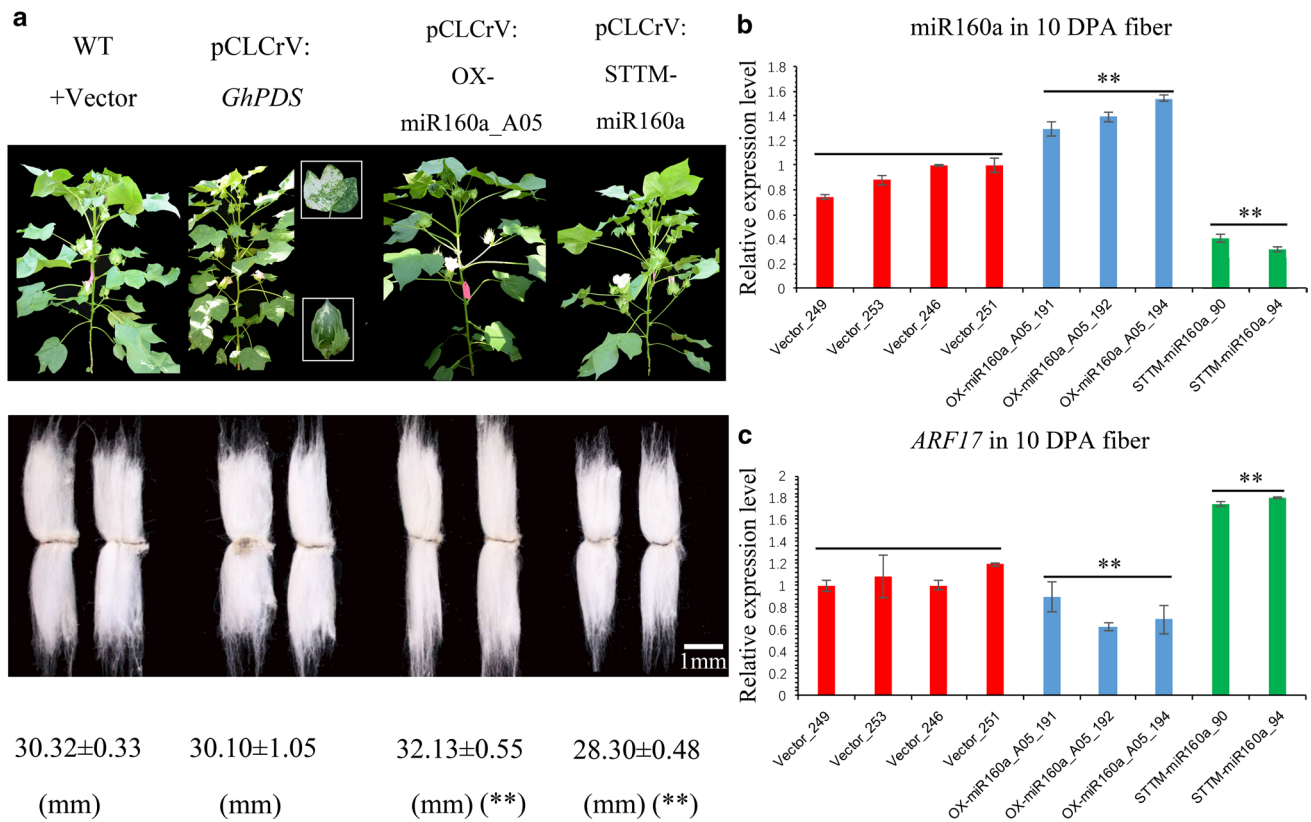
e; type II, mature miR160f and miR160h; type III, mature miR160g; type IV, and mature miR160i. The y axis represents normalized expression in tags per million (TPM). Count numbers are shown above the columns

Previous studies found that *ARFs* are transcription factors that can activate auxin-response genes including *GH3* (Mallory et al. 2005), which can decrease the cellular level of activated IAA. IAA is known to promote fiber elongation both in vivo and in vitro (Beasley and Ting 1974; Guinn and Brummett 1988; Liu et al. 2012). In addition, *ARF17* regulates several *GH3* genes in *Arabidopsis*, especially the group II members *GH3.2*, *GH3.3*, *GH3.5*, and *GH3.6* (Mallory et al. 2005; Staswick et al. 2005). To investigate any expression differences in these genes between “Long” and “Short”, we compared RNA-seq data for group II *GH3* members between the two BILs. A total of nine genes was identified, including one gene annotated as *GH3.2*, four as *GH3.3*, one as *GH3.5*, and three as *GH3.6* (Fig. 5a). Six *GH3* genes had expression trends similar to that of *ARF17*, with lower expression in “Long” than in “Short”, especially in 10 DPA fibers. The exceptions were one *GH3.3* gene (*CotAD\_41357*), which had the opposite expression of *ARF17*. The single *GH3.2* gene (*CotAD\_33008*) and another *GH3.3* gene (*CotAD\_70072*) were not differentially expressed between the two BILs (Fig. 5a).

Four of the differentially expressed *GH3* genes were chosen for qRT-PCR analysis to determine expression levels in cotton plants with overexpressed or suppressed

levels of miR160a. Three genes (*GH3.3*, *GH3.5*, and one *GH3.6* gene) were expressed at significantly lower levels in pCLCrV-OX-miR160a-infected plants and at significantly higher levels in pCLCrV-STTM-miR160a-infected plants compared with vector control plants (Fig. 5b–d). The expression of the fourth gene (another *GH3.6* gene) did not differ significantly among the pCLCrV-OX-miR160a-infected plants, pCLCrV-STTM-miR160a-infected plants, and vector control plants (Fig. 5e). This indicated that not all group II *GH3* genes are regulated by *ARF17*. Finally, we used HPLC to measure the accumulation of active IAA in 10 DPA fibers of infected cotton plants. Compared with vector control plants, pCLCrV-OX-miR160a-infected plants had higher levels of active IAA, and pCLCrV-STTM-miR160a-infected plants had reduced levels (Fig. 6), indicating that altering miR160a level affects the accumulation of active IAA.

Taken together, our results indicate that the higher expression of miR160a in “Long” compared with “Short” suppresses the expression of its target gene *ARF17*, which in turn downregulates the expression of certain group II *GH3* genes, leading to higher levels of active IAA and increased FL.



**Fig. 4** Altered expression of *MIR160a\_A05* by VIGS affects cotton-fiber elongation. **a** Phenotypes of control and gene-silenced cotton plants. Wild type (WT) transformed by empty vector was used as the control. Insets show the leaf and boll positions on *PDS*-silenced plants, as indicated. The bottom shows fiber lengths of correspond-

ing plants. **b, c** qRT-PCR analysis of miR160a and *ARF17* transcript levels in 10 DPA fibers of silenced and non-silenced cotton plants. Error bars denote standard error (SE,  $n=3$ ); \*\* $P<0.01$  versus empty vector control

## Discussion

### The *MIR160* family in tetraploid cotton

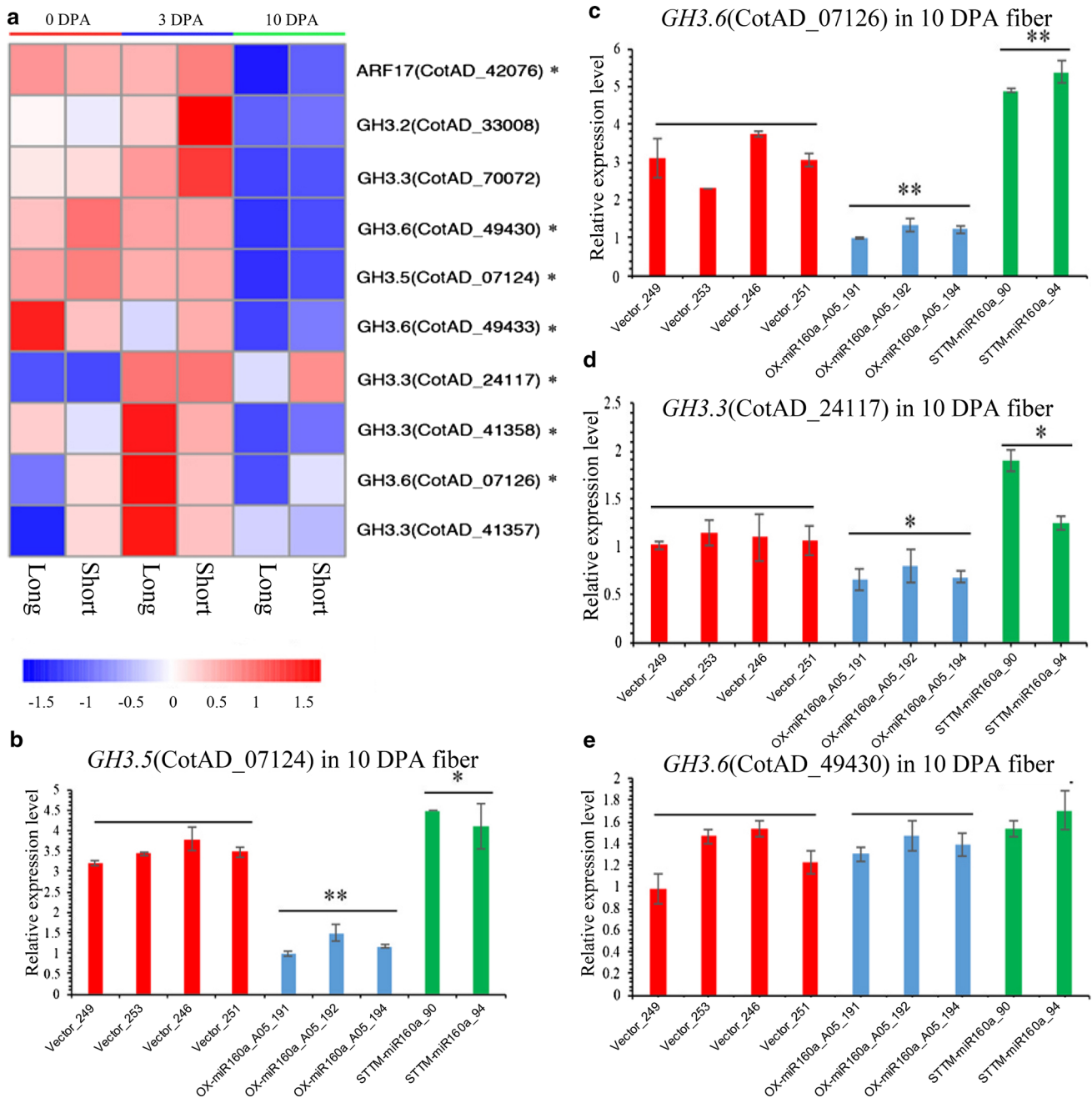
Previous studies identified several mature miR160 sequences in *G. hirsutum* and *G. barbadense*, but none have identified all *MIR160* genes. In this study, we identified and compared the *MIR160* family in both species. With the exception of *GhMIR160g\_A07*, which did not have a homolog in *G. barbadense*, the *MIR160* family sequences were conserved between the two species and the two (At and Dt) subgenomes. Based on mature miRNA sequences, the members were divided into four types, but only type I (which includes genes encoding miR160a) was differentially expressed. After mapping the *MIR160* genes onto the TM-1 genome, two *MIR160* genes were identified that colocalized with an FL QTL or QTL hotspot (*MIR160a* and *MIR160f*). However, only the precursor of *MIR160a* was differentially expressed between the “Long” and “Short” BILs, which is consistent with the sequencing result for miR160 in type I (miR160a). As shown in Fig. 2b, most *MIR160* homologs localized on

homologous chromosomes at similar locations, with the exception of *MIR160a\_D04* and *MIR160a\_A05*. This result indicates that certain regions on chromosomes 4 and 5 may have been exchanged or recombined during the evolution of tetraploid cotton from its diploid progenitors.

### Potential role for *MIR160a* in fiber elongation

Although miR160 in cotton has been shown to affect anther indehiscence at high temperature (Ding et al. 2017) and to affect fiber initiation (Pang et al. 2009; Wang et al. 2012), its role in determining cotton FL has not been reported. Auxin plays an important role in the early fiber-elongation stage (Liu et al. 2012). IAA can affect fiber-cell elongation in cotton ovules in vitro (Beasley and Ting 1974), and both IAA and related auxin metabolites have been detected in eight DPA fibers (Guinn and Brummett 1988). *ARF* participates in the auxin metabolic pathway and plays an important role in fiber elongation (Hagen and Guilfoyle 2002). The *ARF* family activates or represses the expression of auxin-responsive genes, including *GH3* (Mallory



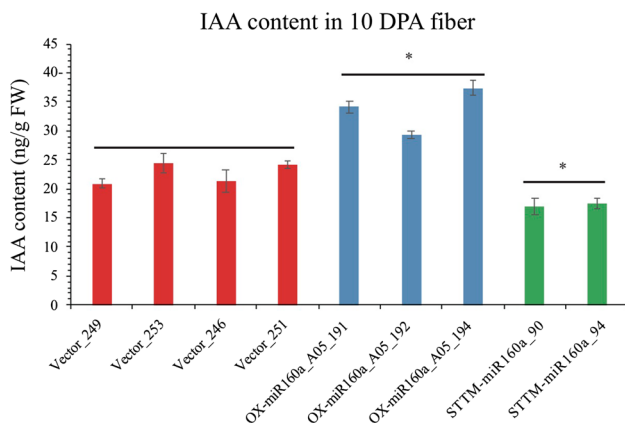


**Fig. 5** Expression of *GH3* members in “Long”, “Short”, and VIGS plants. **a** Heatmap analysis of group II *GH3* family in 0- and 3 DPA ovules and 10 DPA fibers of “Long” and “Short”. Blue, reduced expression; red, increased expression; \*indicates significant differences between “Long” and “Short” 10 DPA fibers. **b–e** qRT–

PCR analysis of transcript levels of *GH3.5* (*CotAD\_07124*), *GH3.6* (*CotAD\_07126* and *CotAD\_49430*), and *GH3.3* (*CotAD\_24117*) in 10 DPA fibers of VIGS and non-VIGS cotton plants. Error bars denote standard error (SE,  $n=6$ ); \* $P<0.05$  versus empty vector control; \*\* $P<0.001$  versus empty vector control

et al. 2005). In *Solanum lycopersicum*, *ARF10* overexpression inhibits leaflet blade outgrowth and reduces cell size compared with wild type (Hendelman et al. 2012). *ARF17* is cleaved by miR160, and the cleaved *ARF17* downregulates certain *GH3* members in *Arabidopsis* (Staswick et al. 2005) and rice (Yang et al. 2006). In rice, overexpression of *GH3.8* suppresses expression of expansin (Ding et al.

2008). The function of group II *GH3*-encoded proteins in *Arabidopsis* is to conjugate IAA to amino acids (e.g., Asp and Glu), which is viewed as a commitment step in IAA catabolism. Overexpressing *GH3.6* increases the accumulation of IAA-Asp and decreases the cellular IAA level (Staswick et al. 2005). In cotton, Zhang et al. (2011) showed that expressing a bacterial IAA synthase gene in



**Fig. 6** IAA content (ng/g fresh weight) in 10 DPA fibers as determined by HPLC. Error bars denote standard error (SE,  $n=3$ ); \* $P < 0.05$  versus empty vector control

ovules at the fiber-initiation stage increases the IAA content, fiber initiation, and lint percentage.

Consistent with these studies, our results reveal that “Long” and its longer fiber *G. barbadense* parent line had higher levels of pri-miR160a and mature miR160a expression, along with lower transcript levels for its target gene *ARF17*, compared with “Short” and the shorter fiber *G. hirsutum* parent line. Although other pathways may also regulate *ARF17* expression, miR160a and *ARF17* had clearly opposite expression trends. Moreover, most members of the group II *GH3* family were also expressed at lower levels in “Long” than in “Short”. Furthermore, overexpression of the precursor gene *MIR160a\_A05* by VIGS reduced the expression of *ARF17* as well as *GH3.3*, *GH3.5*, and *GH3.6*, resulting in increased levels of active IAA and increased FL, whereas silencing *MIR160a* had the opposite effect. Therefore, our results suggest that mature miR160a (from *MIR160a\_A05*) downregulates *ARF17* expression during cotton-fiber development in “Long”, resulting in lower expression of three *GH3* genes, decreased the accumulation of IAA-Asp and IAA-Glu, and increased the level of free IAA, which in turn increases fiber elongation. Although this is consistent with previous studies, showing that IAA affects fiber elongation, further experiments are needed to better understand the regulatory mechanisms involved. Such experiments include generating transformed cotton plants overexpressing miR160a in “Short” or knocking down miR160a in “Long”, ascertaining why *ARF17* does not regulate all group II *GH3* genes, examining the differences among group II *GH3* gene promoters, and determining how IAA promotes fiber elongation. Nevertheless, this study demonstrates that *MIR160a\_A05* and its target *ARF17* play important roles in determining FL in the “Long” and “Short” BILs.

## Conclusions

Members of the *MIR160* family were identified in *G. hirsutum* and *G. barbadense*. *MIR160a\_A05* (on chromosome A05) was identified as a candidate gene for partially determining cotton FL. These results provide an important foundation for further studies of the molecular- and genomic-level genetics of fiber elongation when FL genes are introduced from *G. barbadense* into *G. hirsutum* for improvement of fiber quality in cotton.

**Author contribution statement** JY and SY conceived and designed the experiments. GL, JL, WP, XL, JM, NW, JZ, and XZ performed the experiments. GL analyzed the data. GL, JY, and MW wrote the paper.

**Acknowledgements** We thank Dr. Yuqiang Sun, College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou, and Dr. Xueping Zhou, Institute of Biotechnology, Zhejiang University, Hangzhou, for providing the CLCrV-based vectors.

**Funding** The research was supported by grants from the National Natural Science Foundation of China (grant no. 31621005), National Key Research and Development Program of China (grant no. 2018YFD0100402 and 2018YFD0100300), and the National Research and Development Project of Transgenic Crops of China (grant no. 2016ZX08005005).

## Compliance with ethical standards

**Conflicts of interest** The authors have no conflict of interest to declare.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Beasley C, Ting IP (1974) Effects of plant growth substances on in vitro fiber development from unfertilized cotton ovules. *Am J Bot* 61(2):188–194
- Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, Arioli T, Brubaker C, Cantrell RG, Lacape JM, Ulloa M, Chee P, Gingle AR, Haigler CH, Percy R, Saha S, Wilkins T, Wright RJ, Van Deynze A, Zhu Y, Yu S, Abdurakhmonov I, Katageri I, Kumar PA, Mehboob Ur R, Zafar Y, Yu JZ, Kohel RJ, Wendel JF, Paterson AH (2007) Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol* 145(4):1303–1310
- Ding X, Cao Y, Huang L, Zhao J, Xu C, Li X, Wang S (2008) Activation of the indole-3-acetic acid-amido synthetase GH3-8 suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in rice. *Plant Cell* 20(1):228–240

- Ding YH, Ma YZ, Liu NA, Xu J, Hu Q, Li YY, Wu YL, Xie S, Zhu LF, Min L, Zhang XL (2017) microRNAs involved in auxin signalling modulate male sterility under high-temperature stress in cotton (*Gossypium hirsutum*). *Plant J* 91(6):977–994
- Gokani SJ, Thaker VS (2002) Role of gibberellic acid in cotton fibre development. *J Agr Sci* 138(3):255–260
- Gou J, Strauss SH, Tsai CJ, Fang K, Chen Y, Jiang X, Busov VB (2010) Gibberellins regulate lateral root formation in *Populus* through interactions with auxin and other hormones. *Plant Cell* 22(3):623–639
- Gu ZH, Huang CJ, Li FF, Zhou XP (2014) A versatile system for functional analysis of genes and microRNAs in cotton. *Plant Biotechnol J* 12(5):638–649
- Guan XY, Pang MX, Nah G, Shi XL, Ye WX, Stelly DM, Chen ZJ (2014) miR828 and miR858 regulate homoeologous *MYB2* gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat Commun* 5:3050
- Guinn G, Brummett DL (1988) Changes in abscisic acid and indoleacetic acid before and after anthesis relative to changes in abscission rates of cotton fruiting forms. *Plant Physiol* 87(3):629–631
- Hagen G, Guilfoyle T (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol Biol* 49(3–4):373–385
- Hawkins RS, Serviss GH (1930) Development of cotton fibers in the Pima and Acala varieties. *J Agric Res* 40(11):1017–1029
- Hendelman A, Buxdorf K, Stav R, Kravchik M, Arazi T (2012) Inhibition of lamina outgrowth following *Solanum lycopersicum* AUXIN RESPONSE FACTOR 10 (*SIARF10*) derepression. *Plant Mol Biol* 78(6):561–576
- Hu H, Yu D, Liu H (2015) Bioinformatics analysis of small RNAs in pima (*Gossypium barbadense* L.). *PLoS One* 10(2):e0116826
- Kelly CM, Hequet EF, Dever JK (2012) Interpretation of AFIS and HVI fiber property measurements in breeding for cotton fiber quality improvement. *J Cotton Sci* 16:1–16
- Kim HJ, Triplett BA (2001) Cotton fiber growth in planta and in vitro Models for plant cell elongation and cell wall biogenesis. *Plant Physiol* 127(4):1361–1366
- Kwak PB, Wang QQ, Chen XS, Qiu CX, Yang ZM (2009) Enrichment of a set of microRNAs during the cotton fiber development. *BMC Genomics* 10(1):457
- Lee JJ, Woodward AW, Chen ZJ (2007) Gene expression changes and early events in cotton fibre development. *Ann Bot* 100(7):1391–1401
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25(15):1966–1967
- Li Q, Jin X, Zhu YX (2012) Identification and analyses of miRNA genes in allotetraploid *Gossypium hirsutum* fiber cells based on the sequenced diploid *G-rainmondii* genome. *J Genet Genomics* 39(7):351–360
- Li XH, Wu M, Liu GY, Pei WF, Zhai HH, Yu JW, Zhang JF, Yu SX (2017) Identification of candidate genes for fiber length quantitative trait loci through RNA-Seq and linkage and physical mapping in cotton. *BMC Genomics* 18(1):427
- Liu K, Sun J, Yao LY, Yuan YL (2012) Transcriptome analysis reveals critical genes and key pathways for early cotton fiber elongation in Ligon lintless-1 mutant. *Genomics* 100(1):42–50
- Liu N, Tu L, Tang W, Gao W, Lindsey K, Zhang X (2014) Small RNA and degradome profiling reveals a role for miRNAs and their targets in the developing fibers of *Gossypium barbadense*. *Plant J* 80(2):331–344
- Liu GJ, Xiao GH, Liu NJ, Liu D, Chen PS, Qin YM, Zhu YX (2015) Targeted lipidomics studies reveal that linolenic acid promotes cotton fiber elongation by activating phosphatidylinositol and phosphatidylinositol monophosphate biosynthesis. *Mol Plant* 8(6):911–921
- Liu N, Tu L, Wang L, Hu H, Xu J, Zhang X (2017) MicroRNA 157-targeted *SPL* genes regulate floral organ size and ovule production in cotton. *BMC Plant Biol* 17(1):7
- Liu G, Wu M, Pei W, Li X, Wang N, Ma J, Zang X, Yu S, Zhang J, Yu J (2019) A comparative analysis of small RNAs between two Upland cotton backcross inbred lines with different fiber length: expression and distribution. *Crop J* 7(2):198–208
- Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of *Arabidopsis AUXIN RESPONSE FACTOR17* is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* 17(5):1360–1375
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen XM, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi YJ, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu JK (2008) Criteria for annotation of plant microRNAs. *Plant Cell* 20(12):3186–3190
- Naoumkina M, Thyssen GN, Fang DD, Hinchliffe DJ, Florane CB, Jenkins JN (2016) Small RNA sequencing and degradome analysis of developing fibers of short fiber mutants Ligon-lintless-1 (*Li<sub>1</sub>*) and -2 (*Li<sub>2</sub>*) revealed a role for miRNAs and their targets in cotton fiber elongation. *BMC Genomics* 17(1):360
- Pang MX, Woodward AW, Agarwal V, Guan XY, Ha M, Ramachandran V, Chen XM, Triplett BA, Stelly DM, Chen ZJ (2009) Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biol* 10(11):R122
- Qin YM, Zhu YX (2011) How cotton fibers elongate: a tale of linear cell-growth mode. *Curr Opin Plant Biol* 14(1):106–111
- Ruan YL, Llewellyn DJ, Furbank RT (2001) The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K<sup>+</sup> transporters and expansin. *Plant Cell* 13(1):47–60
- Ruan YL, Xu SM, White R, Furbank RT (2004) Genotypic and developmental evidence for the role of plasmodesmatal regulation in cotton fiber elongation mediated by callose turnover. *Plant Physiol* 136(4):4104–4113
- Said JI, Song MZ, Wang HT, Lin ZX, Zhang XL, Fang DD, Zhang JF (2015) A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum* x *G. barbadense* populations. *Mol Genet Genomics* 290(3):1003–1025
- Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, Zhu YX (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18(3):651–664
- Shindy WW, Smith OE (1975) Identification of plant hormones from cotton ovules. *Plant Physiol* 55(3):550–554
- Staswick PE, Serban B, Rowe M, Tiryaki I, Maldonado MT, Maldonado MC, Suza W (2005) Characterization of an *Arabidopsis* enzyme family that conjugates amino acids to indole-3-acetic acid. *Plant Cell* 17(2):616–627
- Sun Y, Allen RD (2005) Functional analysis of the *BIN2* genes of cotton. *Mol Genet Genomics* 274(1):51–59
- Sun Y, Fokar M, Asami T, Yoshida S, Allen RD (2004) Characterization of the *Brassinosteroid insensitive 1* genes of cotton. *Plant Mol Biol* 54(2):221–232
- Sun Y, Veerabomma S, Abdel-Mageed HA, Fokar M, Asami T, Yoshida S, Allen RD (2005) Brassinosteroid regulates fiber development on cultured cotton ovules. *Plant Cell Physiol* 46(8):1384–1391
- Wang JW, Wang LJ, Mao YB, Cai WJ, Xue HW, Chen XY (2005) Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* 17(8):2204–2216

- Wang ZM, Xue W, Dong CJ, Jin LG, Bian SM, Wang C, Wu XY, Liu JY (2012) A comparative miRNAome analysis reveals seven fiber initiation-related and 36 novel miRNAs in developing cotton ovules. *Mol Plant* 5(4):889–900
- Wang M, Tu L, Yuan D, Zhu Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q, Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, Zhang X (2019) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet* 51(2):224–229
- Wu MF, Tian Q, Reed JW (2006) *Arabidopsis* microRNA167 controls patterns of *ARF6* and *ARF8* expression, and regulates both female and male reproduction. *Development* 133(21):4211–4218
- Xue W, Wang ZM, Du MJ, Liu YD, Liu JY (2013) Genome-wide analysis of small RNAs reveals eight fiber elongation-related and 257 novel microRNAs in elongating cotton fiber cells. *BMC Genomics* 14:629
- Yan J, Gu YY, Jia XY, Kang WJ, Pan SJ, Tang XQ, Chen XM, Tang GL (2012) Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. *Plant Cell* 24(3):1301
- Yang JH, Han SJ, Yoon EK, Lee WS (2006) Evidence of an auxin signal pathway, microRNA167-*ARF8-GH3*, and its response to exogenous auxin in cultured rice cells. *Nucleic Acids Res* 34(6):1892–1899
- Yu JW, Zhang K, Li SY, Yu SX, Zhai HH, Wu M, Li XL, Fan SL, Song MZ, Yang DG, Li YH, Zhang JF (2013) Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. *Theor Appl Genet* 126(1):275–287
- Zhang M, Zheng XL, Song SQ, Zeng QW, Hou L, Li DM, Zhao J, Wei Y, Li XB, Luo M, Xiao YH, Luo XY, Zhang JF, Xiang CB, Pei Y (2011) Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat Biotechnol* 29(5):453
- Zhang TZ, Hu Y, Jiang WK, Fang L, Guan XY, Chen JD, Zhang JB, Sasaki CA, Scheffler BE, Stelly DM, Hulse-Kemp AM, Wan Q, Liu BL, Liu CX, Wang S, Pan MQ, Wang YK, Wang DW, Ye WX, Chang LJ, Zhang WP, Song QX, Kirkbride RC, Chen XY, Dennis E, Llewellyn DJ, Peterson DG, Thaxton P, Jones DC, Wang Q, Xu XY, Zhang H, Wu HT, Zhou L, Mei GF, Chen SQ, Tian Y, Xiang D, Li XH, Ding J, Zuo QY, Tao LN, Liu YC, Li J, Lin Y, Hui YY, Cao ZS, Cai CP, Zhu XF, Jiang Z, Zhou BL, Guo WZ, Li RQ, Chen ZJ (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol* 33(5):531–537

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