Available online at www.sciencedirect.com

ScienceDirect

Journal of Genetics and Genomics 41 (2014) 553−556

JGG

LETTER TO THE EDITOR

# Identification and Profiling of Known and Novel Fiber MicroRNAs during the Secondary Wall Thickening Stage in Cotton (*Gossypium hirsutum*) *via* High-Throughput Sequencing

Upland cotton (*Gossypium hirsutum* L.) is an allotetraploid species originated from interspecific hybridization between AA-genome diploid (*G. arboretum*) and DD-genome diploid (*G. raimondii*) (Wendel et al., 1992). Cotton fibers are single-celled trichomes that emerge from the ovule epidermal cells. Indexed by the number of days post-anthesis (dpa), fiber morphogenesis includes four distinct but overlapping steps: initiation (0−3 dpa), elongation (3−20 dpa), secondary cell wall thickening (15−45 dpa) and maturation (40−60 dpa) (Yang et al., 2008, Du et al., 2013). The efficiency and duration of each morphogenesis stage is important to the quality attributes of the mature fiber. Cell elongation is critical for fiber length, whereas secondary cell wall thickening is important for fiber fineness and strength (Meinert and Delmer, 1977).

Cotton microRNAs (miRNAs) are endogenous, non-coding small RNAs that play crucial roles in posttranscriptional gene regulation *via* the degradation of target mRNAs (Guan et al., 2014). Many studies have investigated cotton miRNAs and their regulatory functions in developing fiber cells (Abdurakhmonov et al., 2008; Pang et al., 2009). Notably, Li et al. (2012) employed ncRNA-seq technology to identify 562 candidate miRNAs from 7-dpa fibers of allotetraploid cotton, and Wang et al. (2012a) annotated 348 miRNA candidate sequences in cotton DD-genome shotgun sequences. In our previous studies, seven fiber initiation-related miRNAs were characterized by comparative miRNAome analysis in developing cotton ovules (Wang et al., 2012b). Subsequently, from 4 small RNA libraries constructed from upland cotton fibers ranging from 5 to 20 dpa, eight miRNAs and one trans-acting small interfering RNA (tasiRNA) were found to be related to cotton fiber elongation in *G. hirsutum* (Xue et al., 2013). These studies demonstrate potential regulatory functions for these small RNAs during cotton fiber initiation and elongation. However, no study to date has validated the distribution of cotton miRNAs and their expression patterns in secondary wall thickening (SWT) fiber cells.

In this study, we employed high-throughput sequencing to examine known and novel cotton miRNAs expressed in SWT fibers. A small RNA library was constructed using total RNA extracted from 25-dpa fibers of an allotetraploid cotton cultivar CRI 35 and sequenced with an Illumina HiSeq 2000 analyzer. A total of 15,748,452 high-quality reads were obtained from the library, generating 14,886,698 clean reads (2,686,236 unique sequences) for further analysis. The redundant and unique reads >18 nucleotides long are shown in Fig. 1A. The majority (62.26%) of the clean reads were 21−24 nt long; of these, a length of 24-nt (17.91%) was the most abundant class of total clean reads, followed by lengths of 21-nt (17.27%), 22-nt (13.63%) and 23-nt (13.45%). As small RNAs in the 21-nt class typically represent miRNAs, the majority of small RNAs in SWT fiber cells are putative miRNAs. Regarding unique reads, the 24- and 23-nt classes ranked first (41.24%) and second (17.76%), respectively, indicating that the 24-nt unique reads were the most abundant of all unique reads. In *Arabidopsis*, small 24-nt RNAs are mainly small interfering RNAs (siRNAs) (Lu et al., 2005); thus, the high percentage of unique small 24-nt RNA sequences in fiber cells suggests that a large number of siRNA molecules may be enriched during the SWT stage.

Clean reads were mapped to the diploid cotton *G. raimondii* genome (Wang et al., 2012a), generating 9,069,837 genome-matched reads (60.92% of clean reads). The genome-matched small RNA sequences were clustered into several RNA classes including known miRNAs, rRNA, tRNA, small nuclear RNA/small nucleolar RNA (snRNA/snoRNA), mRNA and repeats (Table S1). However, only 25.88% of the unique reads were mapped to the diploid DD-genome, suggesting that the remaining 74.12% of unique reads might be from another allotetraploid cotton AA-genome or from evolved AADD-genomes. Furthermore, the majority (93.23%) of the unique sequences comprised unclassified small RNA reads (Table S1), potentially containing new regulatory small RNAs and novel miRNA sequences, and should be further studied in detail.

http://dx.doi.org/10.1016/j.jgg.2014.08.002

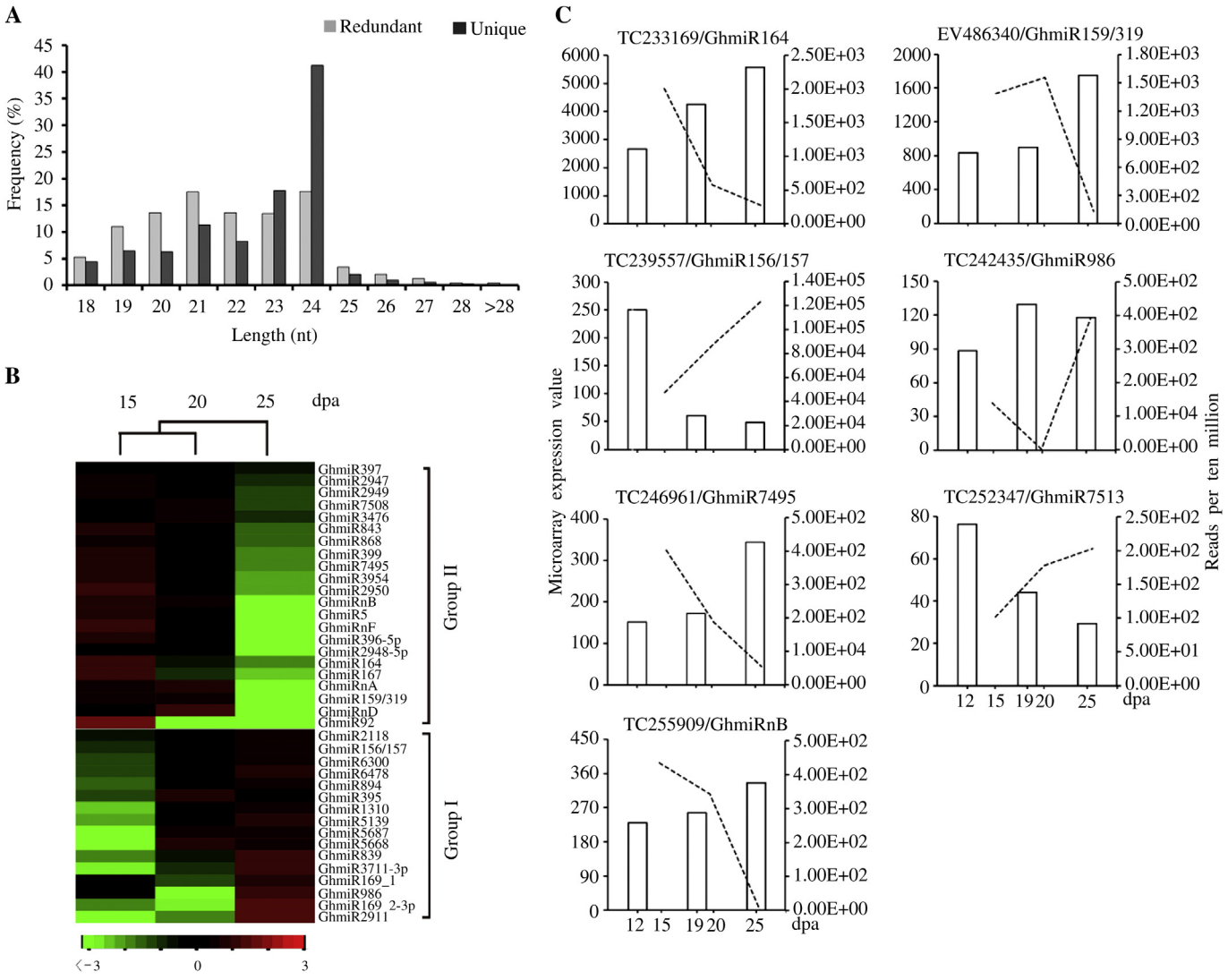


Fig. 1. Characteristic analyses of small RNAs from a 25-dpa cotton fiber library.

**A:** Length distribution and abundance of small RNA sequences in the 25-dpa cotton fiber library. **B:** Differentially expressed miRNA profiles during the cotton fiber secondary wall thickening stage. A complete linkage hierarchical cluster analysis of differentially expressed miRNAs in the cotton fibers (15−25 dpa) was performed by comparing the RPTM of the miRNAs in every library to the average of the three cotton fiber sRNA libraries. The color indicates the $\log_2$-fold change from high (red) to low (green), as indicated in the color scale. The names of the miRNAs and the clusters to which they belong are shown on the right side of the panel, and the dendrogram of the 15-, 20- and 25-dpa small RNA libraries generated by the cluster analysis is shown above. **C:** Expression analysis of seven differentially expressed miRNAs and their targets based on microarray data. The data represent the mean values of three replicates. The dots represent the corresponding abundance on the right axis in the three small RNA libraries. TC233169: UDP-glucuronic acid decarboxylase; EV486340: beta-ketoacyl-CoA synthase; TC246961: a hypothetical protein; TC255909: a transaldolase family protein; TC239557: squamosa promoter-binding-like protein; TC252347: proline-rich receptor-like protein kinase; TC242435: choline kinase 3-like.

Currently, miRBase (version 20.0, released on June 24, 2013) lists 84 mature miRNAs belonging to 55 miRNA families expressed in cotton (*Gossypium* spp.). To identify these known miRNAs in our data, clean reads were blasted against known miRNAs in miRBase 20.0. Following a BlastN search and further sequence analysis, nearly 1800 unique sequences belonging to 98 families were found to be homologs of known miRNAs from other plant species. Expression among different families was found to be largely divergent, with abundance ranging from 1 to 183,036 reads (Table S2). Based on a normalized read count per ten million (RPTM), the five known miRNA families with greatest abundance in our data are 156/157, 894, 2911, 165/166 and 167, accounting for 46.5%, 28.6%, 9.2%, 7.4% and 3.1% of all known miRNA reads, respectively. However, the expression of some miRNA families (such as 398, 528 and 858) showed low abundance, with only several reads each (Table S2). Such variable expression levels in different families suggest that these miRNA genes are differentially transcribed during the fiber cell SWT stage.

Because there are no available genome sequences for allotetraploid cotton (*G. hirsutum*), we also examined known miRNA precursors by mapping known miRNA sequences to *G. raimondii* genome sequences. Among the known miRNAs, a total of 116 sequences from 26 miRNA families were mapped to the genome (Table S3), all of which are able to adopt hairpin structures resembling the fold-back structures of

miRNA precursors. Stem-loop structures for pre-GhmiR172-1.2, pre-GhmiR396-1.1, pre-GhmiR5225-1, pre-GhmiR1520-1.2 and pre-GhmiR1520-2.1 were identified for the first time (Fig. S1). In addition, the number of members per miRNA family ranged from 1 to 7 members. The miR156, miR482 and miR477 families were the largest, with 7, 6 and 5 members, respectively, whereas 13 miRNA families had only a single member (Table S3). Interestingly, different members of the same miRNA family were found to be derived from different precursors. For example, in the seven-member miR156 family, miR156-1 and miR156-2 are derived from six and four different gene loci, respectively, though each of the other family members are derived from only one precursor gene (Table S3).

The unannotated small RNAs of 2,504,484 unique reads were mapped to the *G. raimondii* genome and subjected to a rigorous secondary structure analysis of their precursors. Using the Mireap software developed by BGI (Shenzhen, China) and the three criteria suggested by Meyers et al. (2008), 21 potentially novel miRNAs from 28 miRNA gene loci were identified (Tables S4 and S5; Fig. S2). Moreover, 3 (GhmiRn002, GhmiRn007, and GhmiRn010) of the 21 novel miRNAs were linked to more than one miRNA gene locus, whereas the remaining 18 novel miRNAs were associated with a single locus (Tables S4 and S5); 21- and 24-nt long miRNAs were the dominant species (8 and 8, respectively) of the 21 novel miRNAs. Notably, more than 40% of these novel miRNAs have a 5′ terminal uridine nucleotide (Table S6). Understandably, the abundance of these novel miRNAs was lower than that of the known miRNAs in SWT fiber cells and was detected at very low levels, ranging from 1 to 16 reads (Table S4). In addition, the minimum free energy for the hairpin structures of these novel miRNA precursors was lower than −29.5 kcal/mol, and the lengths of the new miRNA precursors ranged from 86 to 457 nt (Fig. S2 and Table S5), consistent with previous studies (Bonnet et al., 2004). Furthermore, we can predict that there will be more novel miRNA loci identified in the larger A sub-genome (∼1.7 Gb of haploid *G. arboretum*) and that the allotetraploid *G. hirsutum* genome might contain many more miRNA loci.

To elucidate the expression patterns of miRNAs during the SWT stage, we compared our current data from 25-dpa fiber cells with previously reported data from 15- and 20-dpa fibers (Xue et al., 2013) because they successively represent the elongation stage and transition stage from elongation to SWT. To minimize noise and improve accuracy, only those miRNA reads with more than 100 RPTM in at least one library were used for the comparison. After Student's *t*-test ($P < 0.01$) and cluster analysis, the remaining 38 known miRNAs exhibited differential expression with more than 2-fold changes of up- or down-regulation during the fiber developmental process and clustered into two groups with similar profiles (Fig. 1B and Table S7). Of these, 16 miRNAs, including GhmiR156/157, GhmiR395 and GhmiR2911 of Group I, showed a gradual increase in read number and peaked at 25-dpa, a point at which SWT occurs. Thus, Group I miRNAs most likely play negative regulatory roles in fiber SWT. In contrast, 22 miRNAs in Group II, including GhmiR92, GhmiRnD and GhmiR159/319, exhibited a gradual decrease and maintained a relatively low level at 25-dpa (Fig. 1B and Table S7), implying that these miRNAs might serve as positive regulators of fiber SWT. Taken together, all the differentially regulated miRNAs, including Groups I and II, demonstrated dynamic changes in read number in response to different developmental stages and may be considered as SWT-related miRNAs in cotton fibers.

To understand the functions of SWT-related miRNAs, we predicted their targets using the psRNATarget Web server (http://plantgrn.noble.org/psRNATarget/). As a result, 299 EST sequences were obtained as candidate target genes of 38 differentially expressed miRNAs. These targets include many important proteins of regulatory factors and metabolic enzymes, e.g., myeloblastosis (MYB) transcription factor, LIM (Lin11, Isl-1 and Mec-3) domain protein, NAC (NAM, ATAF1, 2 and CUC2) domain protein, sucrose synthase, 3-ketoacyl-CoA synthase and sugar transport protein, suggesting that these miRNAs mediate fiber SWT by regulating their targets (Table S8).

In general, the expression of miRNAs is negatively correlated with the expression of their targets. To further examine the relationships between these miRNAs and their targets, we downloaded microarray data from the NCBI Gene Expression Omnibus (GEO) database with the accession number GSE36228. Only data from three stages (12-, 19-, and 25-dpa) were selected for further analysis to determine whether a negative correlation exists between the accumulation of miRNAs and their targets during fiber development. As shown in Fig. 1C, the expression levels of seven target genes were inversely correlated with the read abundances of their corresponding miRNAs, which is in accordance with the gene-silencing function of miRNAs. For the targets of GhmiR164, 159/319, 7495 and GhmiRnB, with RPTMs decreased from 15- to 25-dpa, the expression levels increased from 12- to 25-dpa. Conversely, during fiber developmental stages, GhmiR156/157, 7513 and 986 were expressed at relatively lower levels in 25-dpa *vs* 15-dpa fibers, whereas, as expected, their target genes TC239557 (squamosa promoter-binding-like protein), TC252347 (proline-rich receptor-like protein kinase) and TC242435 (choline kinase 3-like) were expressed at relatively higher levels in 25-dpa *vs* 15-dpa fibers (Fig. 1C). These results indicate that miRNAs might regulate the expression of their target genes during the SWT stage of cotton fiber development.

SWT is the process of cellulose biosynthesis in cotton fiber cells. It has been reported that a series of enzymes, including sucrose synthase (Brill et al., 2011), cellulose synthase (Jacob-Wilk et al., 2006) and glycosyltransferases (Scheible and Pauly 2004), play key roles in the pathways of cellulose and sugar metabolism and are significantly correlated with SWT. In this study, several miRNAs were found to have differential expression that may impact their target genes during cellulose and sucrose metabolism. For example, GhmiR164, GhmiR2911, GhmiR395 and GhmiR986 were predicted to target UDP-glucuronic acid decarboxylase, sucrose synthase, cellulose synthase-like G2 and sugar transport protein,

respectively, and GhmiR159/319 was predicted to target 3-ketoacyl-CoA synthase (Fig. 1C). Therefore, these differentially expressed miRNAs and their sugar metabolism-related targets might play a role in cotton fiber SWT, and further analysis of their relationships and regulatory interactions is needed. Together, these miRNAs and their targets might play important roles in various biochemical processes to regulate fiber SWT.

## ACKNOWLEDGEMENTS

We are grateful to members of the Laboratory of Molecular Biology at Tsinghua University for comments and participation in discussions. This work was supported by the grants from the State Key Basic Research and Development Plan (No. 2010CB126003) and the National Transgenic Animals and Plants Research Project (Nos. 2011ZX08005-003 and 2011ZX08009-003).

## SUPPLEMENTARY DATA

Supplementary data related to this article can be found at doi:10.1016/j.jgg.2014.08.002.

*Dingwei Yu*[a,b], *Yanmei Wang*[a], *Wei Xue*[a], *Shuli Fan*[b], *Shuxun Yu*[b,*], *Jin-Yuan Liu*[a,*]

[a]*Laboratory of Plant Molecular Biology, Center for Plant Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China*

[b]*State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China*

*Corresponding authors. Tel: +86 10 6277 2243 (J.-Y. Liu); Tel: +86 37 2252 5363 (S. Yu).
*E-mail addresses:* yu@cricaas.com.cn (S. Yu); liujy@mail.tsinghua.edu.cn (J.-Y. Liu)

Received 3 March 2014
Revised 8 July 2014
Accepted 6 August 2014
Available online 19 August 2014

## REFERENCES

Abdurakhmonov, I.Y., Devor, E.J., Buriev, Z.T., Huang, L., Makamov, A., Shermatov, S.E., Bozorov, T., Kushanov, F.N., Mavlonov, G.T., Abdukarimov, A., 2008. Small RNA regulation of ovule development in the cotton plant, *G. hirsutum* L. BMC Plant Biol. 8, 93.

Bonnet, E., Wuyts, J., Rouze, P., Van de Peer, Y., 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics 20, 2911−2917.

Brill, E., van Thournout, M., White, R.G., Llewellyn, D., Campbell, P.M., Engelen, S., Ruan, Y.L., Arioli, T., Furbank, R.T., 2011. A novel isoform of sucrose synthase is targeted to the cell wall during secondary cell wall synthesis in cotton fiber. Plant Physiol. 157, 40−54.

Du, S.J., Dong, C.J., Zhang, B., Lai, T.F., Du, X.M., Liu, J.Y., 2013. Comparative proteomic analysis reveals differentially expressed proteins correlated with fuzz fiber initiation in diploid cotton (*Gossypium arboreum* L.). J. Proteomics 82, 113−129.

Guan, X., Pang, M., Nah, G., Shi, X., Ye, W., Stelly, D.M., Chen, Z.J., 2014. miR828 and miR858 regulate homoeologous *MYB2* gene functions in *Arabidopsis* trichome and cotton fibre development. Nat. Commun. 5, 3050.

Jacob-Wilk, D., Kurek, I., Hogan, P., Delmer, D.P., 2006. The cotton fiber zinc-binding domain of cellulose synthase A1 from *Gossypium hirsutum* displays rapid turnover *in vitro* and *in vivo*. Proc. Natl. Acad. Sci. USA 103, 12191−12196.

Li, Q., Jin, X., Zhu, Y.X., 2012. Identification and analyses of miRNA genes in allotetraploid *Gossypium hirsutum* fiber cells based on the sequenced diploid *G. raimondii* genome. J. Genet. Genomics 39, 351−360.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., Green, P.J., 2005. Elucidation of the small RNA component of the transcriptome. Science 309, 1567−1569.

Meinert, M.C., Delmer, D.P., 1977. Changes in biochemical composition of the cell wall of the cotton fiber during development. Plant Physiol. 59, 1088−1097.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., Griffiths-Jones, S., Jacobsen, S.E., Mallory, A.C., Martienssen, R.A., Poethig, R.S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., Zhu, J.K., 2008. Criteria for annotation of plant microRNAs. Plant Cell 20, 3186−3190.

Pang, M., Woodward, A.W., Agarwal, V., Guan, X., Ha, M., Ramachandran, V., Chen, X., Triplett, B.A., Stelly, D.M., Chen, Z.J., 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, R122.

Scheible, W.R., Pauly, M., 2004. Glycosyltransferases and cell wall biosynthesis: novel players and insights. Curr. Opin. Plant Biol. 7, 285−295.

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S., Zou, C., Li, Q., Yuan, Y., Lu, C., Wei, H., Gou, C., Zheng, Z., Yin, Y., Zhang, X., Liu, K., Wang, B., Song, C., Shi, N., Kohel, R.J., Percy, R.G., Yu, J.Z., Zhu, Y.X., Yu, S., 2012a. The draft genome of a diploid cotton *Gossypium raimondii*. Nat. Genet. 44, 1098−1103.

Wang, Z.M., Xue, W., Dong, C.J., Jin, L.G., Bian, S.M., Wang, C., Wu, X.Y., Liu, J.Y., 2012b. A comparative miRNAome analysis reveals seven fiber initiation-related and 36 novel miRNAs in developing cotton ovules. Mol. Plant 5, 889−900.

Wendel, J.F., Brubaker, C.L., Percival, A.E., 1992. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am. J. Bot. 79, 1291−1310.

Xue, W., Wang, Z., Du, M., Liu, Y., Liu, J.Y., 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.

Yang, Y.W., Bian, S.M., Yao, Y., Liu, J.Y., 2008. Comparative proteomic analysis provides new insights into the fiber elongating process in cotton. J. Proteome Res. 7, 4623−4637.