



An analysis of lncRNAs related to fiber quality and the discovery of their target genes in a *Gossypium hirsutum* line with *Gossypium mustelinum* introgression

Baohua Wang¹ · Meijun Ji¹ · Hui Fang¹ · Haijing Gu¹ · Teame Gereziher Mehari¹ · Jinlei Han¹ · Wenxiang Feng¹ · Xuehan Huo² · Jingxia Zhang² · Yu Chen² · Jun Zhang² · Allah Ditta³ · Muhammad K. R. Khan³ · Andrew H. Paterson⁴ · Peng W. Chee⁵ · Kai Wang¹

Received: 4 October 2023 / Accepted: 4 January 2024 / Published online: 31 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Key message Analysis of fiber quality lncRNAs and their target genes from a pair of *Gossypium mustelinum* near-isogenic lines provide new prospects for improving the fiber quality of Upland cotton.

Abstract Long noncoding RNAs (lncRNAs) are an important part of genome transcription and play roles in a wide range of biological processes in plants. In this research, a pair of near-isogenic cotton lines, namely, a *Gossypium mustelinum* introgression line (IL9) with outstanding fiber quality and its recurrent Upland cotton parent (PD94042), were used as the experimental materials. Cotton fibers were selected for lncRNA sequencing at 17 and 21 days post-anthesis. A total of 2693 differentially expressed genes were identified. In total, 5841 lncRNAs were ultimately screened, from which 163 differentially expressed lncRNAs were identified. Target genes of the lncRNAs were predicted by two different methods: *cis* and *trans*. Some of the target genes were related to cell components, membrane components, plant hormone signal transduction and catalytic metabolism, and the results indicated that there might also be important effects on the development of fiber. Four differentially expressed target genes related to fiber quality (*Gomus.D05G015100*, *Gomus.A05G281300*, *Gomus.A12G023400* and *Gomus.A10G226800*) were screened through gene function annotation, and the functions of these four genes were verified through virus-induced gene silencing (VIGS). Compared to the negative controls, plants in which any of these four genes were silenced showed significant reductions in fiber strength. In addition, the plants in which the *Gomus.A12G023400* gene was silenced showed a significant reduction in fiber uniformity, whereas the plants in which *Gomus.A05G281300* was silenced showed a significant increase in fiber fineness as measured via micronaire. Our results showed that these genes play different roles during fiber development, impacting fiber quality.

Introduction

In addition to serving as one of the world's major cash crops, cotton has also been suggested to be an ideal model for studying allopolyploid plants (Wendel et al. 2012). Approximately 5–7 million years ago, diploid species differentiation occurred between the A genome originating from Africa and the D genome originating from Mexico. After the A and D species hybridized approximately 1–2 million

years ago, the new polyploid lineage ultimately gave rise to 7 tetraploid cotton species: *Gossypium hirsutum* L. (AD1), *Gossypium barbadense* L. (AD2), *Gossypium tomentosum* Nutt. ex Seem. (AD3), *Gossypium mustelinum* Miers ex G.Watt (AD4), *Gossypium darwinii* G.Watt (AD5), *Gossypium ekmanianum* Wittm. (AD6), and *Gossypium stephensii* (AD7; Gallagher et al. 2017). The main cultivated cotton varieties are tetraploid *G. hirsutum* and *G. barbadense*, with the improved form of *G. hirsutum* (Upland cotton) accounting for over 95% of global cotton fiber production due to its high yield and wide adaptation. After domestication and improvement, the growth cycle of Upland cotton has gradually shifted from a perennial semiwild species to an annual cultivated species. During the process of domestication, cotton fibers have also changed from the original short and sparse brown fibers to the current long and dense

Baohua Wang, Meijun Ji, and Hui Fang have contributed equally to this work.

Communicated by Tianzhen Zhang.

Extended author information available on the last page of the article

white fibers (Bao et al. 2019), with higher yields and easier production.

Cotton fiber quality traits are controlled by complex interactions of multiple genes. The evaluation of cotton fiber quality is mainly based on fiber length, strength, micronaire value (an indirect measure of fineness), uniformity, and elongation. Among these, fiber length and strength have been suggested to be the main characteristics determining fiber quality (Razzaq et al. 2022). Longer fibers, with low short fiber content, confer higher yarn count and strength and, in turn, higher textile quality. Higher fiber strength improves fabric strength, making the final product sturdy and durable. Current cotton varieties often suffer one or more defects in fiber quality that interfere with their processing and/or utility as textiles and are all inferior to synthetic fibers that lack the comforting texture of natural fibers. Therefore, it is extremely important to explore and identify candidate genes for excellent fiber quality and to breed cotton varieties with elite fiber quality.

The findings of previous research suggest that while the majority of genes in the genome are transcribed, a substantial portion of the transcripts may not encode proteins (Knoll et al. 2015). Noncoding RNAs belong to a class of functional RNAs with low protein-coding potential. According to their length, they can be divided into small noncoding RNAs (18–30 nt), medium-sized noncoding RNAs (31–200 nt), and long noncoding RNAs (lncRNAs) (> 200 nt) (Zhang et al. 2019). lncRNAs can be divided into long noncoding natural antisense transcripts (lncNATs), long intergenic noncoding RNAs (lincRNAs), overlapping lncRNAs, and long intron noncoding RNAs according to their genomic location and composition. lncRNAs are RNA transcripts with a length of over 200 bases that does not have significant protein coding ability (Salih et al. 2019). In early studies on the topic, lncRNA was considered a byproduct of RNA polymerase II during transcription, and no biological function was known. However, an increasing number of lncRNAs have been revealed to be involved during developmental and pathological stages, indicating their important regulatory roles in diseases and development. lncRNAs regulate the expression of related genes through different modes of action so that they can participate in various biological activities, mainly those related to regulatory networks at the transcription, post-transcription, translation, and epigenetic levels (Kim and Sung 2012).

To date, an increasing number of lncRNAs have been identified in plants. Studies have shown that lncRNAs can affect important biological processes, such as flowering, male and female differentiation, pollen development, gene silencing mechanisms, and abiotic stress tolerance. In research on the regulatory factors of grape responses to powdery and downy mildews, it was found that lncRNAs participated in the regulation of basic and specific defense

responses, such as cell wall reinforcement, active oxygen metabolism, disease-related protein accumulation, plant hormone signal transduction, and secondary metabolism (Bhatia et al. 2021). Dong et al. (2022) reported a novel lncRNA, *drought-induced intergenic lncRNA (DIR)*, which increases drought tolerance in cassava by modifying stress-related gene expression. Zhang et al. (2022) identified 2743 putative lncRNAs in rice under heat stress, of which 231 were differentially expressed lncRNAs (DELs).

Previous studies have shown that lncRNAs have a significant impact on cotton fiber development. Zou et al. (2016) systematically identified lncRNAs involved in fiber initiation and elongation processes, and the results suggested that rapid and dynamic changes in lncRNAs may contribute to fiber development in cotton. Salih et al. (2019) identified two lncRNAs (LNC_001237 and LNC_017085) that were significantly downregulated during fiber development using cotton mutants. Zheng et al. (2021) found that *lnc-Ga13g0352* was coexpressed with protein-coding genes during ovule and fiber development, and its lncRNA was verified to have dual functions in regulating target genes, both inhibiting and activating target gene transcription. Recently, it was found that lncRNA *MSTRG 2723.1* was associated with fat metabolism pathways and regulated cotton fiber initiation and development stages (Zou et al. 2022).

Although both *G. mustelinum* and *G. hirsutum* are tetraploid cottons that can be traced to the same polyploid ancestor, *G. mustelinum* is in the polyploid clade most distant from *G. hirsutum*. Our previous research showed that *G. mustelinum* had some unique fiber property alleles that can be applied to improve fiber quality in Upland cotton. A *G. mustelinum* introgression line, IL9, with excellent fiber quality was selected by marker-assisted selection, forming a pair of near isogenic lines with the recurrent Upland cotton parent PD94042 (Chen et al. 2021). In the current study, this pair of near-isogenic lines was used for lncRNA sequencing to explore candidate genes related to fiber quality. Cotton fibers at 17 and 21 days post-anthesis (dpa) were collected as samples, and RNA was extracted and used for lncRNA sequencing to identify lncRNAs and their target genes related to fiber quality and provide genetic resources for future cotton breeding.

Materials and methods

Library construction and sequencing

A *Gossypium mustelinum* introgressed line, IL9, with elite fiber quality and its recurrent Upland cotton parent, PD94042, were planted in the Botany Garden of Nantong University in 2021 and used in this experiment. The development of the *G. mustelinum* introgressed line IL9 has been

described elsewhere (Chen et al. 2020), briefly involving crossing *G. mustelinum* acc. AD 4–8 with the *G. hirsutum* line PD94042 followed by three backcrosses to PD94042 and marker-assisted selection from the BC₃F_{4.5} generation. The RNAprep Pure Plant Plus Kit (Polysaccharides & Polyphenolics-rich, Tiangen Biotech (Beijing) Co., LTD, China) was used to extract total RNA from cotton fibers at 17 and 21 dpa for lncRNA sequencing. Nanodrop was used to detect the purity of total RNA (OD260/280). An Agilent 2100 Bioanalyzer was used to measure the integrity of total RNA. Qubit was used for precise quantification of RNA concentration. Total RNA was sent for poly(A)⁺-type RNA-Seq library construction and high-throughput sequencing provided by Biomics (Beijing) Biotech Co., Ltd., China. The libraries were sequenced on an Illumina HiSeq2000 sequencer with 100-bp paired-end reads.

Quality control of sequencing

To ensure the accuracy of the subsequent analysis, strict quality control was performed on the data to ensure that these reads had sufficiently high quality. The filtering method used on the raw reads was to remove the reads containing adaptors and to remove low-quality reads, i.e., reads with a ratio of N greater than 10%, and reads with more than 50% bases having Q -score ≤ 10 . After the series of quality control measures mentioned above, high-quality clean reads were obtained. The quality score or Q -score is an integer mapping of the probability of base calling errors. The commonly used Phred base Q -score formula is:

$Q_{\text{score}} = -10 \times \log_{10} P$, where P is the probability of base calling errors.

The higher the base Q -score is, the more reliable and accurate the base recognition.

Read mapping and transcript assembly

Reference genome sequences of Upland cotton were obtained from the Joint Genome Institute database (https://phytozome-next.jgi.doe.gov/info/Ghirsutum_v2_1). Afterward, HISAT2 was used to perform sequence alignment between clean reads and the reference genome to obtain the position information on the reference genome or gene, as well as the unique sequence feature information of the sequencing samples. The aligned reads were then assembled into transcripts using StringTie (Kim et al. 2015) software for final quantitative analysis.

Identification of lncRNAs

By comparison with known mRNAs and utilizing the class_Code information from the results of gffcompare, candidate transcripts of long intergenic noncoding RNA

(lincRNA), intronic lncRNA, and antisense lncRNA types were screened. lncRNA is a type of long-stranded noncoding RNA with a length of > 200 bp. lncRNAs can be divided into intergenic lncRNAs (referred to as lincRNAs), intronic lncRNAs, antisense lncRNAs, sense lncRNAs, bidirectional lncRNAs, and other types based on their positional relationship with coding sequences. Among these, lincRNAs accounted for the highest proportion, and the first three types of screening were mainly conducted at this point. We set a series of strict screening conditions based on the structural characteristics and noncoding functional characteristics of lncRNAs. Based on the StringTie and Scripture assembly results, we conducted the following screening steps. The screened lncRNAs were used as the final candidate lncRNA set for subsequent analysis. The screening mainly consisted of five parts. First, transcripts with a length of ≥ 200 bp and an exon count of ≥ 2 were selected. Second, the read coverage of each transcript was calculated through StringTie, and transcripts with a minimum coverage ≥ 3 were selected. Third, transcripts of known nonlncRNA and nonmRNA types (rRNA, tRNA, snRNA, snoRNA, pre-miRNA, pseudogenes, etc.) of the species were compared, and transcripts that were similar or identical to the aforementioned known transcripts were identified. Fourth, the coding potential was screened, transcripts with coding potential (CPC, CNCI, Pfam) were removed, coding potential screening was performed, and coding potential calculator (CPC) analysis (Kong et al. 2007), coding–noncoding index (CNCI) analysis (Sun et al. 2013), and Pfam protein domain analysis (Finn et al. 2014) were performed to remove transcripts with coding potential. Finally, if there was small RNA information for the species in miRbase, BLAST was performed to remove the small RNA precursor sequence.

Identification of DEGs and DELs

Transcriptome sequencing can be simulated as a random sampling process; that is, sequence fragments can be randomly selected independently from any nucleic acid sequence of a sample transcript. The number of fragments extracted from a gene (or transcript) follows a beta negative binomial distribution. Based on this mathematical model, the expression levels of transcripts and genes were quantified using HTSeq-count software by mapping the position information of reads on genes.

The number of fragments extracted from a transcript is related to the amount of sequencing data (or mapped data), transcript length, and transcript expression level. For the number of fragments to truly reflect the transcript expression level, it is necessary to normalize the number of mapped reads and the transcript length in the sample. Htseq-count, which uses fragments per kilobase of script per million

fragments mapped (FPKM) as an indicator to measure transcript or gene expression levels, was utilized.

DESeq2 was used for differential expression analysis between sample groups to identify a set of differentially expressed genes (DEGs) and differentially expressed lncRNAs (DELs) between any two biological conditions, with $|\text{Fold Change}| \geq 1.5$ and $\text{FDR} < 0.05$ being used as screening criteria. Fold change represents the ratio of expression levels between two samples (groups), whereas the false discovery rate (FDR) is determined by correcting for the p value of significant differences.

Predicting the target genes of the lncRNAs

Because lncRNA itself does not have a coding function, its principle of action is usually achieved by acting on protein coding target genes in two ways, namely *cis* and *trans*, and another approach is playing its role based on complementary pairing of lncRNA and mRNA. On the basis of this principle, we used the target gene prediction tool LncTar (McKenna et al. 2010) to predict the target genes of the lncRNAs. The basic principle of *cis* target gene prediction is that the function of lncRNAs is related to the protein-coding genes near their coordinates, so the protein-coding genes located approximately 100 kb upstream and downstream of lncRNAs are selected as their target genes. However, the prediction of the *trans* effect on its related target gene prediction is different from that of the *cis* effect. The basic principle of its prediction is that the function of lncRNA is not related to the positional relationship between itself and the coding gene but is related to its coexpressed protein-coding genes. Its target genes can be predicted through correlation analysis or coexpression analysis of the expression of lncRNAs and protein-coding genes among samples.

Validation of the expression of DELs and their target genes

The relative quantities of lncRNAs and their target genes were determined using qRT-PCR analysis. Table S1 shows the information for all the primers. In addition, actin was used as an internal reference gene. qRT-PCR was performed with ChamQ SYBR qPCR Master Mix (Low ROX) from Vazyme Biotech Co., Ltd., according to the manufacturer's protocol. Primer5 software was used to design qRT-PCR-specific primers, using the Upland cotton actin gene as a control for normalization between samples. qRT-PCR was performed in an ABI 7500 real-time fluorescence quantitative PCR instrument (Applied Biosystems, Singapore), and the relative expression of candidate genes was calculated using $2^{-\Delta\Delta C_t}$ (Livak and Schmittgen 2001). GraphPad Prism v.9.0.0 software was used to visualize the graphing results.

Virus-induced gene silencing (VIGS) experiment

The gene sequence of the target gene was uploaded to the SGN-VIGS Tool (<http://vigs.solgenomics.net/>) for sequence analysis, the optimal 200–350 bp region was selected as the VIGS target gene fragment, the cDNA of the *G. mustelinum* introgression line IL9 was used as the amplification template, and Primer5 software was used to design primers for the target gene (Table S2). If the amplification band showed a single target band, a Tiangen agarose gel DNA recovery kit (DP209-03) for gel recovery was used to recover the target segment.

Two suitable restriction endonuclease sites, *Spe* I and *Asc* I, were selected by analyzing the pCLCrV vector (Fig. S1) using SnapGene software. The digested vector was purified by agarose gel electrophoresis to obtain the corresponding digested purified vector.

The recovered target fragment was ligated to the purified pCLCrV-A vector with a ClonExpress II One Step Cloning Kit (www.vazyme.com, Vazyme Biotech Co., Ltd). The plasmid containing the VIGS fragment together with pCLCrV-A was transformed into *Agrobacterium tumefaciens* LBA4404 according to the *Agrobacterium tumefaciens* transformation instructions.

In the VIGS protocol, pCLCrV::00 was used as the negative control, and pCLCrV::*GhPDS* was used as the positive control. The LBA4404 strain carrying the pCLCrV::Target gene, pCLCrV::00 (negative control) and pCLCrV::*GhPDS* (positive control) was mixed with the strain containing pCLCrV-B (helper vector) at a 1:1 ratio to achieve a final $\text{OD}_{600} = 1.5$. The needle of a sterile syringe was used to prick a small hole on the back of two fully unfolded cotyledons of cotton material of IL9, and then, the *Agrobacterium* bacterial solution was injected into the same batch of cotton plants accordingly. The treated plants were transferred to a growth chamber with a photoperiod of 16/8 h light/dark, 28 °C light and 22 °C dark. The plants responded to infiltration at approximately 15 days after transfection. The *GhPDS* gene was silenced in the positive control, resulting in bleaching of the true leaves. To ensure that the target gene in the experimental group was silenced, the total RNAs of the negative control and experimental group leaves and 17 and 21 dpa fibers were extracted for quantitative analysis, where the negative control plants were used as the reference.

Cotton fiber quality traits were tested for three biological replicates in the control group and the experimental group to investigate whether the target gene is related to fiber quality. Fifteen VIGS plants for each candidate gene and 15 control plants were used to collect cotton bolls. Cotton fiber ginned from 15 bolls was collected as a biological sample, and three biological samples for each gene and control were tested. Fiber quality traits, including upper-half mean length (mm), fiber strength (cN/tex), fiber uniformity index (%),

fiber micronaire, and fiber elongation (%), were measured by an HVI900 fiber quality tester at the Fiber Quality Supervising and Testing Center, Ministry of Agricultural and Rural Affairs, China.

Results

High-throughput sequencing and DEG analysis

After confirming the validity and completeness of the sequenced raw data, screening and filtering were conducted, and various aspects of the high-quality clean data output were statistically analyzed and are described in Table 1.

The Q30 content in each sample was greater than 94%, whereas the Q20 content was greater than 98% (Table 1), indicating the high reliability and accuracy of our sequencing data. Among the data, the percentage of GC content was between 44 and 48%, which is within the 30–50% range typical of terrestrial plants. More than 85% of the clean reads were aligned to the reference genome (Table S3), and those corresponding to exon regions accounted for most of the total (Fig. S2). The intergenic region with incomplete genome annotation accounted for only a small part, which showed that the comparison results were reliable, and the accuracy of subsequent analysis and research was also ensured.

The number of DEGs among different samples is presented in a Venn diagram (Fig. 1). The number of DEGs between IL9–17 dpa and IL9–21 dpa was the highest, with a total of 1720 DEGs, of which 840 were upregulated and 880 were downregulated. The number of DEGs between PD94042–17 dpa and IL9–17 dpa was the lowest, with a total

of 66 DEGs, of which 55 were upregulated and 11 were downregulated.

Identification and characterization of lncRNAs

Due to the inherent structure of lncRNAs and their lack of coding ability, it was necessary to analyze the coding function of the transcripts in three databases, namely CNCI, CPC, and Pfam protein domains, after passing strict screening conditions. As shown in Fig. S3, we can observe some differences in the predicted results in different structural domains, perhaps due to the different calculation methods and relevant principles used by each software. Therefore, we selected the intersection of the results of these three databases and identified them as the lncRNA databases of 17 and 21 dpa for Upland cotton fibers for the convenience of subsequent analysis, and 5841 lncRNAs were ultimately identified through rigorous screening.

The identified lncRNAs and mRNAs were compared and analyzed in terms of structure. Specifically, the following two aspects were analyzed. First, the length of lncRNA was mainly distributed in the 400–800 nt concentration (Fig. 2a, 42%), with only a small portion distributed in ≥ 3000 nt (2%); in contrast, mRNA was concentrated at ≥ 3000 nt (accounting for 11%), while the remaining distribution was relatively balanced. The number of exons contained (Fig. 2b) was usually 2–3, accounting for 86% of the total number of lncRNAs, and the largest number of exons among all the lncRNAs was 10. However, there were relatively more genes containing a single exon in mRNA, accounting for 53% of the total number, and approximately 7% of the total lncRNAs had more than 10 exons.

Differential expression analysis was conducted on the identified lncRNAs, and a total of 163 DELs were identified

Table 1 General information about the clean data of the sequenced samples

Sample ID	Read sum	Base sum	GC (%)	Q20 (%)	Q30 (%)
PD-17-1	34,422,583	10,326,774,900	44.59	98.40	95.13
PD-17-2	37,685,740	11,305,722,000	44.25	98.36	95.02
PD-17-3	36,087,928	10,826,378,400	47.01	98.29	94.93
PD-21-1	35,241,348	10,572,404,400	44.77	98.38	95.10
PD-21-2	37,689,892	11,306,967,600	44.50	98.36	94.99
PD-21-3	40,236,469	12,070,940,700	44.54	98.35	95.01
IL9-17-1	33,353,557	10,006,067,100	44.42	98.35	94.99
IL9-17-2	36,279,293	10,883,787,900	43.80	98.36	94.96
IL9-17-3	34,634,917	10,390,475,100	43.80	98.38	95.05
IL9-21-1	40,553,895	12,166,168,500	44.13	98.30	94.86
IL9-21-2	35,466,917	10,640,075,100	44.51	98.36	95.04
IL9-21-3	37,416,885	11,225,065,500	44.19	98.32	94.89

Note: SampleID: Sample name for sequencing; ReadSum: The total number of pair-end reads in clean data; BaseSum: Total base number of clean data; GC: GC content of clean data, which is the percentage of G and C bases in clean data to the total base; Q20 (%): The proportion of data with a base recognition accuracy of 99%; Q30 (%): The proportion of data with a base recognition accuracy of 99.9%

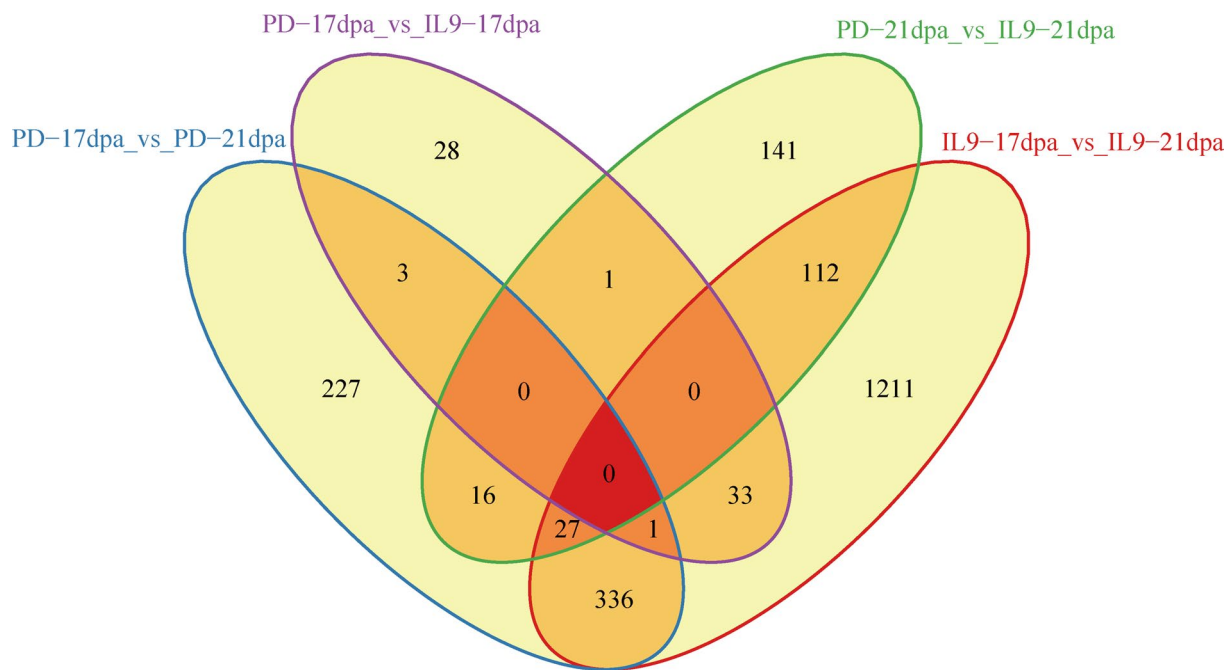


Fig. 1 Venn diagram for differentially expressed genes

(Fig. 3), consistent with the trend in the mRNA results. The highest number of differences was found between IL9–17dpa and IL9–21dpa, with a total number of 120, whereas the lowest number of 8 was found between PD94042–17dpa and IL9–17dpa.

Among the DELs, four lncRNAs related to fiber development were selected for expression validation, and the results are shown in Fig. 4. The relative expression levels of these four lncRNAs in cotton fibers at 17 and 21 dpa were basically consistent with the sequencing results, indicating that our sequencing results are accurate and reliable and have potential functions during fiber development.

Functional analysis of DEGs

To further analyze the potential functions of the DEGs, relevant Gene Ontology (GO) annotation analyses were conducted. The GO annotation system is a directed acyclic graph that contains three main branches, namely the biological process, molecular function, and cell component branches. The GO classification for the DEGs is shown in Fig. S4. The GO enrichment regions of the DEGs were relatively similar among different groups, e.g., the metabolic process and cellular process branches accounted for the most in all four groups (Fig. S4).

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used as the main public database of pathways (Kanehisa et al. 2004), providing not only all possible metabolic pathways but also comprehensive annotations of the

enzymes catalyzing each step of the reaction, including amino acid sequences and links to the Protein Data Bank (PDB).

In this research, it was found that 787, 37, 113, and 286 DEGs were associated with different pathways in IL9–17dpa and IL9–21dpa, PD94042–17dpa and IL9–17dpa, PD94042–21dpa and IL9–21dpa, and PD94042–17dpa and PD94042–21dpa, respectively. Twenty significantly correlated pathways were selected and plotted as KEGG enrichment scatter plots (Fig. S5).

In the IL9–17dpa and IL9–21dpa group, 34 DEGs were related to phenylpropane biosynthesis, and 33 and 26 genes were related to glycolysis and glucose metabolism, respectively. In the PD94042–17dpa and IL9–17dpa group, there were four DEGs associated with starch and sucrose metabolism and flavonoid biosynthesis. In the PD94042–21dpa and IL9–21dpa group, as well as the PD94042–17dpa and PD94042–21dpa group, there were also significant correlations between DEGs and phenylpropane biosynthesis and starch and sucrose metabolism processes.

Prediction of fiber development-related lncRNA target genes and their functional annotation

The GO annotation for target genes of *cis* differentially expressed lncRNAs is shown in Fig. S6. In the IL9–17dpa and IL9–21dpa group, metabolic processes, cellular processes, and biological regulation were mainly enriched in terms of biological processes, with 281, 268, and 110 genes,

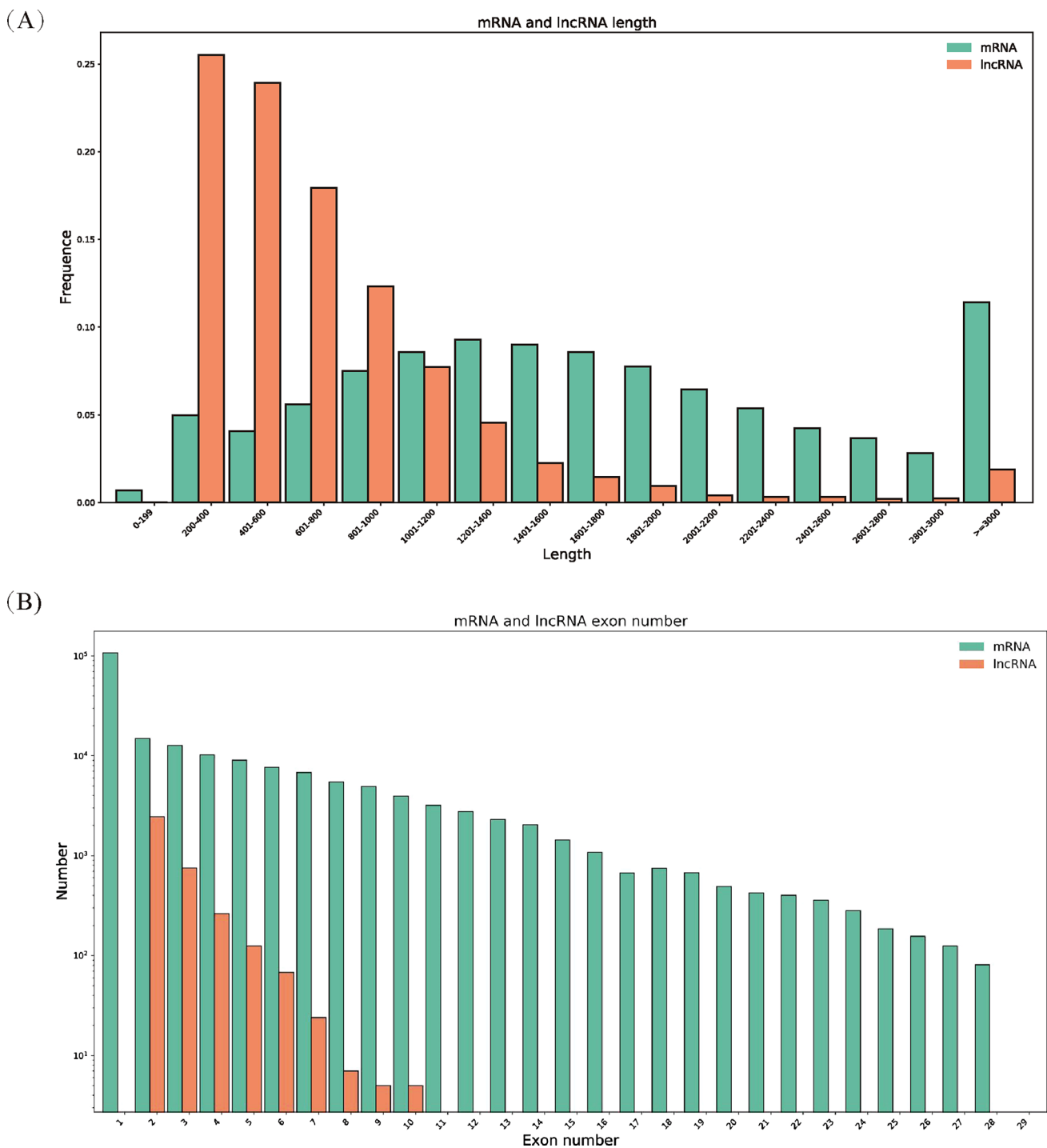


Fig. 2 Basic feature analysis of lncRNAs. **a** Comparison of length between lncRNA and mRNA; **b** Comparison of the number of exons between lncRNA and mRNA

respectively, related to these three aspects, accounting for 72.1% of the total number.

In terms of cellular components, a total of 568 target genes of *cis* DELs were associated with cells and cell parts, accounting for 37%; 519 target genes of *cis* DELs were enriched in membrane parts, accounting for 34.4%.

In terms of molecular function, there were many target genes of *cis* DELs concentrated and distributed in both catalytic activity and binding, with 314 genes related to binding and 290 genes related to catalytic activity. The classification and enrichment distribution among the other

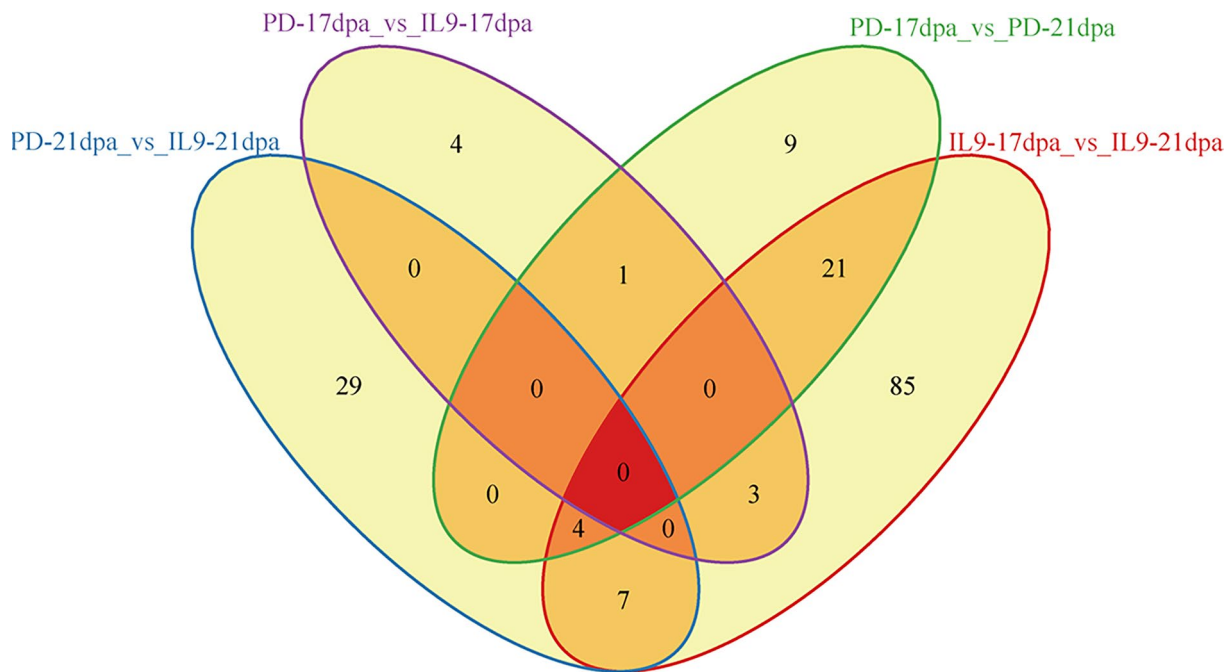


Fig. 3 Venn diagram for differentially expressed lncRNAs

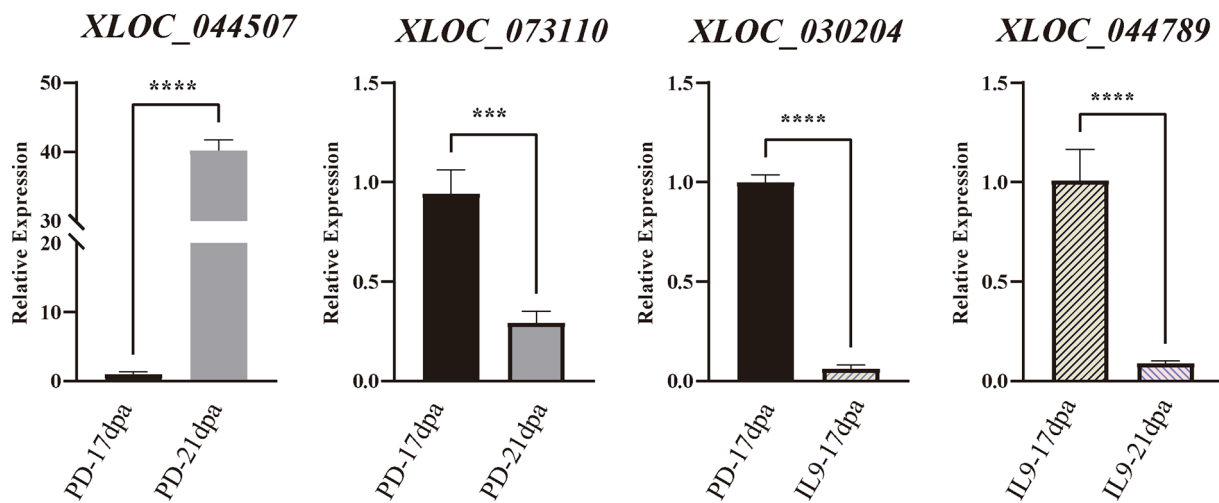


Fig. 4 Expression analysis of differentially expressed lncRNAs related to fiber development. Note: The asterisk above the bar chart indicates statistically significant differences (*** $P < 0.001$, **** $P < 0.0001$)

three groups were relatively consistent with those of the IL9–17dpa and IL9–21dpa group.

Through the KEGG analysis, it was found that 364, 83, and 99 genes were enriched in the IL9–17dpa and IL9–21dpa, PD94042–21dpa and IL9–21dpa, and PD94042–17dpa and PD94042–21dpa groups, respectively. However, there were fewer differentially expressed genes between the PD94042–17dpa and IL9–17dpa group, and there was no significant metabolic pathway or other gene enrichment. Therefore, the gene enrichment of

the other three groups was plotted in a scatter diagram (Fig. S7).

The GO annotation for target genes of *trans* differentially expressed lncRNAs is shown in Fig. S8. The distribution trend of the target genes was consistent with the distribution trend of all genes in the IL9–17dpa and IL9–21dpa group, being relatively balanced across the biological processes, molecular function, and cell component branches. Among the remaining three groups, there was a significant difference in the distribution trend for target genes compared

to that of all genes. For example, in the PD94042–17dpa and IL9–17dpa group, target genes were concentrated in metabolic processes, cellular processes, biological regulation, response to stimulus, localization, cellular component organization or biogenesis, and developmental processes within the biological processes category. In terms of cell components, the DEGs were mainly distributed in the membrane, cell, cell part, membrane part, organelle, and cell junction but not in other parts. In terms of molecular function, binding, catalytic activity, transporter activity, and transcriptional regulator activity were the only ones that were enriched. The distribution within the PD94042–21 dpa and IL9–21 dpa group was relatively consistent with that in the PD94042–17 dpa and IL9–17 dpa group, while it was different from that in the PD94042–17 dpa and PD94042–21 dpa group.

Through the KEGG (Fig. S9) analysis, it was found that the *trans* and *cis* enrichment differences were mainly concentrated between the IL9–17dpa and IL9–21dpa group and the PD94042–17dpa and IL9–17dpa group. Many target genes were identified through different methods and principles. To further clarify how they play roles in the process of fiber development, the relevant gene annotation information from different online databases (<https://phytozome-next.jgi.doe.gov/>; NCBI: <https://www.ncbi.nlm.nih.gov/>) combined with the sequencing data was used to compare and analyze the differentially expressed target genes, and finally, the target genes related to fiber development were identified for subsequent verification.

Functional verification of four candidate genes by VIGS

Functional verification of four candidate genes by VIGS

To further identify the association between target genes of DELs and cotton fiber development, we performed colocalization analysis of DEL target genes and DEGs to narrow down the important candidate genes. A total of 103 DEGs were found to be DEL target genes (Fig. 5). The results of the GO functional annotation analysis showed that these genes were not significantly enriched in any biological or molecular pathway, so we selected three candidate genes (*Gohir.D05G014550*, *Gohir.A10G205300*, and *Gohir.A12G022400*) related to fiber development based on the relevant annotation information, plus another target gene related to fiber development, although it was not among the DEGs (*Gohir.A05G263300*). A total of four candidate genes were selected for functional validation, and their functional annotations are shown in Table 2.

Fig. 5 Colocalization analysis between target genes of the lncRNA and DEGs

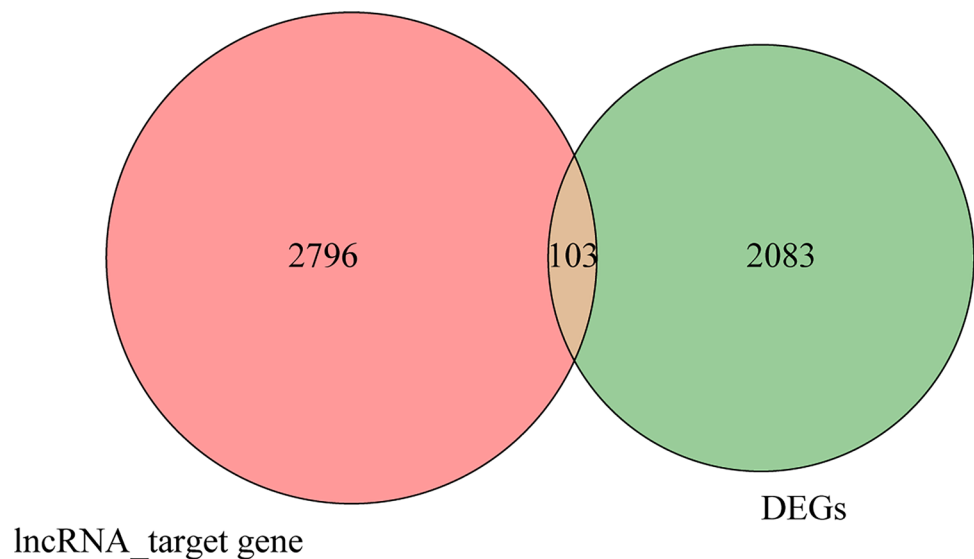


Table 2 Functional annotation information for the target genes

lncRNA	Target genes in <i>Gossypium hirsutum</i>	Homologous genes in <i>Gossypium mustelinum</i>	Functional annotation
<i>XLOC_057472</i>	<i>Gohir.D05G014550</i>	<i>Gomus.D05G015100</i>	Pyrophosphate-energized vacuolar membrane proton pump
<i>XLOC_030245</i>	<i>Gohir.A10G205300</i>	<i>Gomus.A10G226800</i>	Auxin response factor
<i>XLOC_036712</i>	<i>Gohir.A12G022400</i>	<i>Gomus.A12G023400</i>	Phosphoglycerate dehydrogenase activity
<i>XLOC_022539</i>	<i>Gohir.A05G263300</i>	<i>Gomus.A05G281300</i>	Ubiquitin–protein transferase activity

The homologous *G. mustelinum* gene sequences for the relevant target genes were used to carry out qRT-PCR verification. The expression of these four target genes was significantly up- or downregulated to varying degrees (Fig. 6), and the transcriptome data showed the same expression trend.

To determine whether the target genes of the identified lncRNAs affect fiber quality, four target genes, namely *Gohir.A05G263300*, *Gohir.A10G205300*, *Gohir.D05G014550*, and *Gohir.A12G022400*, with annotation information and verified by qRT-PCR, were selected, and their homologous genes in *G. mustelinum*, namely *Gomus.A05G281300*, *Gomus.A10G226800*, *Gomus.D05G015100*, and *Gomus.A12G023400*, were used to construct VIGS vectors for the subsequent experiments. The target fragments were amplified by PCR, and then, the purified target fragments were ligated to the pCLCrV vector. Primers with *Spe* I and *Asc* I restriction endonuclease sites were used for colony PCR identification. After colony PCR identification, target fragments of approximately 272, 271, 278, and 270 bp in size were obtained (Fig. S10). After successful validation, the colonies were expanded and cultured and then sent for sequencing. The sequencing results validated that the recombinant plasmid was successfully transferred into *Agrobacterium tumefaciens*.

After 15 days of treatment with VIGS, infection occurred in the Upland cotton seedlings. As shown in Fig. 7, the negative control showed only a slight wilting state. In the positive control, the synthesis of chlorophyll was disrupted, resulting in photobleaching in the plants with the silenced phytoene desaturase (PDS) gene, such that their true leaves were bleached (Fig. 7), which demonstrated that the VIGS was successful. After confirming the successful injection of the bacterial solution, to ensure that the normal growth and

development of the cotton plants would not later be affected, the plants were moved into large pots with a diameter of 21 cm (Fig. S11).

After the albino phenotype was observed in pCLCrV::*GhPDS*, total RNA was extracted from the leaves and from the fibers at 17 dpa and 21 dpa of the silenced plants, including pCLCrV::*Gomus.D05G015100*, pCLCrV::*Gomus.A05G281300*, pCLCrV::*Gomus.A10G226800*, and pCLCrV::*Gomus.A12G023400*. After reverse transcription, qRT-PCR experiments were performed using the same primers listed in Table S2. As shown in Fig. 8, the expression level of the target genes showed a significant decrease compared to the negative control in both leaves and cotton fibers. The results indicated that VIGS infection was successful and correct and that the target gene was effectively silenced.

After 5–6 months of growth of VIGS plants, mature cotton fibers were collected, with three biological replicates for each candidate gene, and fiber quality traits were measured for each fiber sample (Table 3). The gene-silenced plants showed different fiber quality compared to the negative controls. The fiber strength values of gene-silenced plants for all the four target genes were significantly lower than those of the negative control group, with the gene *Gomus.A12G023400* having the lowest fiber strength of 31.0 cN/tex. Regarding the two other important fiber traits, fiber length and fiber elongation, there were no significant differences between the gene-silenced plants and the negative control. However, the negative control group had a fiber uniformity index of 89.2%, whereas the *Gomus.A12G023400* gene-silenced plants had a fiber uniformity index of 85.8%, which was significantly lower than that of the negative control group. In terms of the fiber micronaire value, the measured

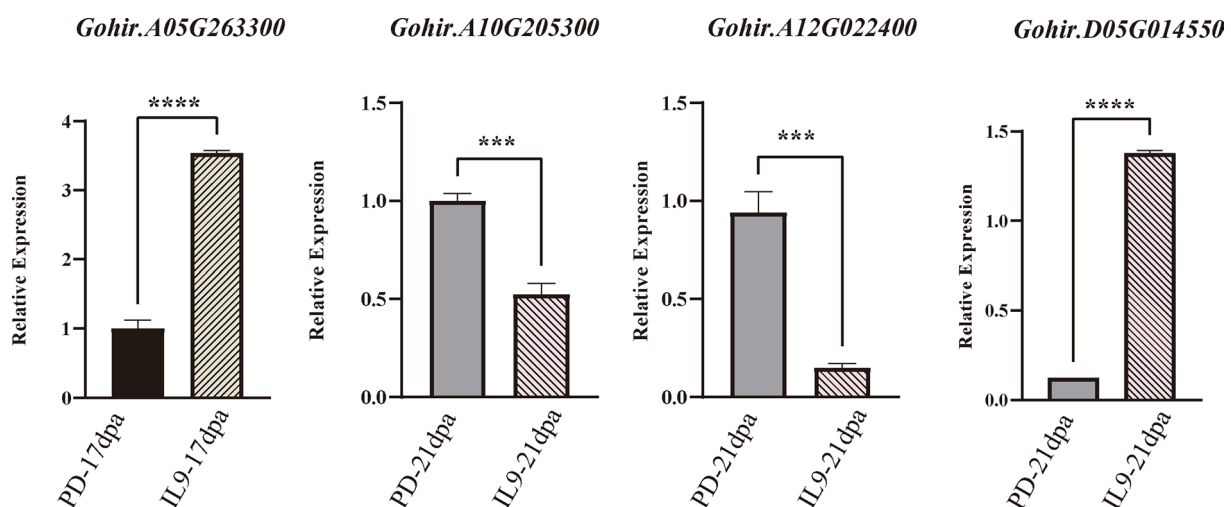


Fig. 6 Expression analysis for target genes of lncRNAs related to fiber development. Note: The asterisks above the bar indicate significant differences (*** $P < 0.001$, **** $P < 0.0001$).



Fig. 7 Leaf color phenotype of VIGS plants at the seedling stage. Note: The first column in the figure is pCLCrV::00 (negative control, empty vector); the second column is pCLCrV::GhPDS (positive control with the *GhPDS* gene silenced), with leaves of all the positive controls showing albinism; the third column represents the experimental group, and the target genes in the A–E group are *Gomus.A05G281300*, *Gomus.D05G015100*, *Gomus.A10G226800*, and *Gomus.A12G023400*, respectively

value of the *Gomus.A05G281300* gene-silenced plants was 4.3, which was significantly higher than that of the negative control group.

Discussion

With the rapid development of omics sequencing technology, lncRNAs have been investigated in many plant species. For example, 238 lncRNAs related to drought stress have been identified in rice (Li et al. 2019a, b), and 17,674 lncRNAs related to flowering and fruiting during growth and development have been identified in tomato (Wang et al. 2021a, b). lncRNAs have been found to play multiple regulatory roles in transcription, posttranscription, translation and chromatin modification and are involved in transcriptional interference, alternative splicing, protein modification and

DNA methylation regulation (Jia et al. 2023). In plants, lncRNAs have been found to play roles in developmental regulation, reproductive development, immunity, and abiotic or biotic stress responses (Seo et al. 2017; Yatusевич et al. 2017). *GhDANI* gene silencing resulted in higher drought tolerance and increased glutathione and proline levels in roots (Tao et al. 2021). There has been a large-scale discovery of lncRNAs during the cotton fiber development stage using publicly available RNA-seq data from different cotton varieties and chain-specific RNA-seq data from the ovules or fibers of Upland cotton, which identified tens of thousands of lncRNAs (Wang et al. 2015). Based on coexpression analysis, several lncRNAs with potential functions in cotton fiber initiation and elongation were also identified (Wang et al. 2015). In this study, we systematically identified and analyzed lncRNAs in cotton fibers to discover novel lncRNAs and their target genes related to fiber development. These data can provide new data resources for the functional study of lncRNAs regulating cotton fiber development.

In this research, we conducted a comprehensive analysis of mRNA, lncRNA, and target genes with sequencing data. A total of 2693 DEGs were identified from mRNA analysis, among which a higher number of upregulated genes were found in the comparison of different cotton lines at the same period, indicating the possibility of genes with a positive response to fiber development in IL9 during fiber development. When compared at different stages of the same line, it was found that there were more downregulated genes, indicating that many genes may have a certain degree of influence in the early stage of fiber development. In the GO classification of the DEGs, less than 30% of the genes were unannotated, while the remaining genes were annotated and were mainly concentrated in cellular processes, cell part, cell junction, biological regulation, metabolic processes, and catalytic activity.

In the lncRNA analysis, a total of 163 DELs were identified, with more upregulated lncRNAs than downregulated lncRNAs in the IL9–17 dpa and IL9–21 dpa group and more downregulated lncRNAs in the PD94042–17 dpa and PD94042–21 dpa group. The number of DELs between PD94042 and IL9 at 17 dpa was much lower than that between PD94042 and IL9 at 21 dpa, indicating that lncRNAs play a significant role in different cotton lines and fiber development processes at different stages.

lncRNAs are key regulatory factors in studying gene expression at both the genetic and epigenetic levels (Kaelik et al. 2019). They can act not only as *cis*-acting factors around RNA synthesis sites but also as *trans*-acting factors and play a role at a distance from the synthesis site (Suksamran et al. 2020). In this research, on the basis of the GO classification and KEGG enrichment analysis of the target genes, the target genes predicted by *cis* action were mainly concentrated in the terms cell, cell part, membrane, membrane

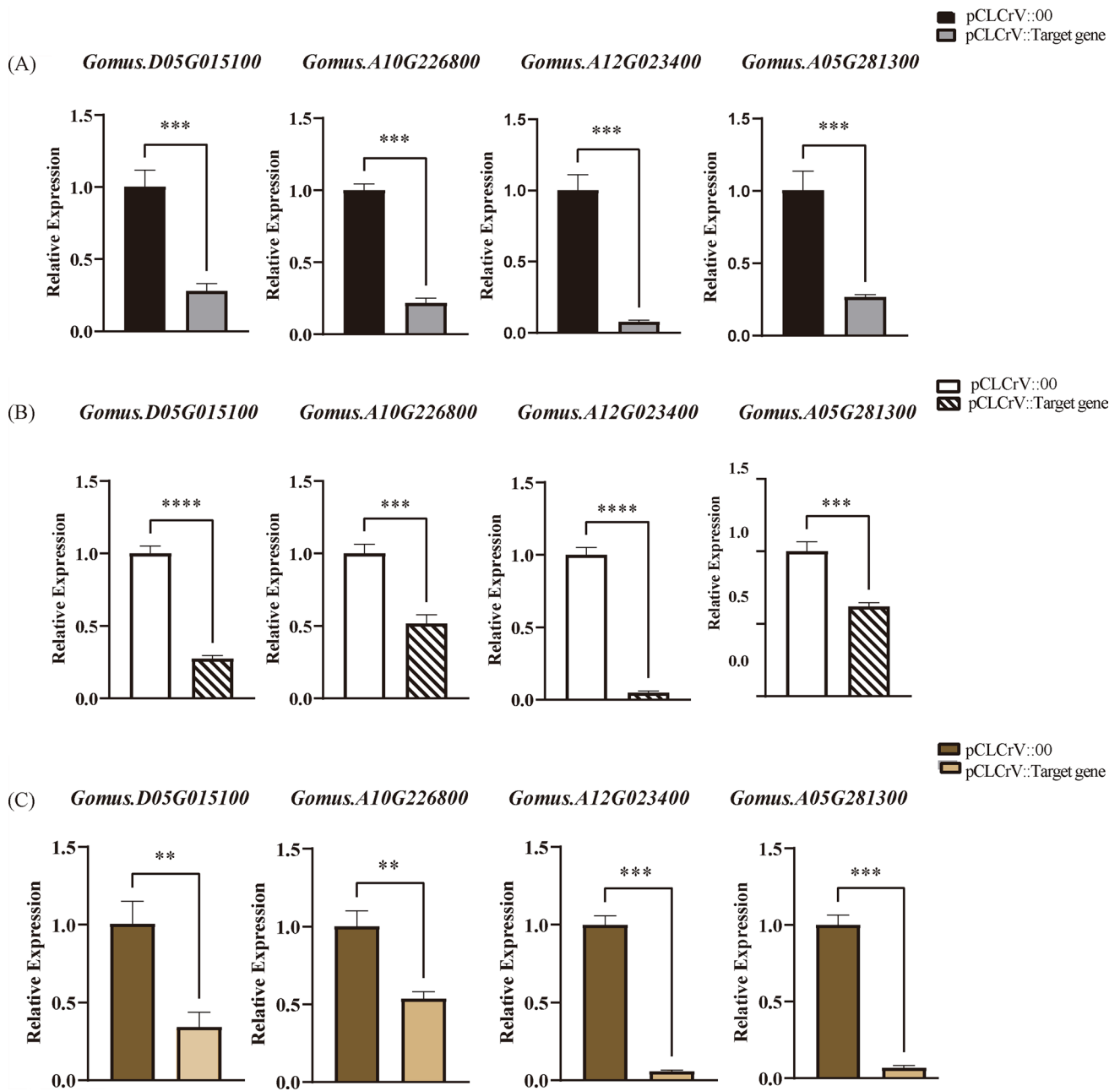


Fig. 8 Gene expression validation of VIGS plants. Note: **a** Leaf; **b** 17 dpa cotton fiber; **c** 21 dpa cotton fiber. The asterisk above the bar chart indicates statistically significant differences (** $P < 0.01$, *** $P < 0.001$)

Table 3 Fiber quality performance of the VIGS plants

Silenced genes	Upper half mean length (mm)	Fiber strength (cN/ tex)	Fiber uniformity index (%)	Fiber micronaire	Fiber elongation (%)
pCLCrV::00	30.8	36.6	89.2	3.7	7.7
<i>Gomus.D05G015100</i>	31.5	33.6*	86.3	3.7	7.4
<i>Gomus.A05G281300</i>	29.8	33.1**	87.3	4.3*	7.0
<i>Gomus.A12G023400</i>	31.4	31.0**	85.8*	3.4	7.1
<i>Gomus.A10G226800</i>	31.2	33.3**	86.8	4.1	7.4

Note: The fiber quality of the plants with silenced genes was compared with that of the pCLCrV::00 plants. * $P < 0.05$, ** $P < 0.01$

part, catalytic activity and binding, while the target genes predicted by *trans* action were not only concentrated in these terms but also were concentrated in plant hormone signal transduction.

Among these target genes, *Gohir.A10G205300* belongs to the ARF (auxin response factor) gene family, and auxin response factors belong to a group of plant transcription factors that are composed of the conserved N-terminal DNA binding domain (DBD), the most conserved C-terminal dimerization domain (CTD), and the nonconserved middle region (MR) (Guilfoyle and Hagen 2007). The MR region is considered to have the function of inhibiting or activating structural domains (Tiwari et al. 2003). Previous studies have reported that *Arabidopsis* contains 23 ARF genes (Okushima et al. 2005), while rice contains 25 ARF genes (Wang et al. 2007). It has been reported that ARF2 negatively regulates plant growth in *Arabidopsis* (Wang et al. 2011; Schruoff et al. 2006) and tomato (Breitel et al. 2016), but the function of transcription factors varies depending on the tissue and is more diverse in polyploid species. It also has a certain impact on the development of cotton fiber cells. Auxin-dependent cell expansion is key to the formation of cotton fiber cells, which determines the yield and quality of cotton fibers (Zeng et al. 2019). Zhang et al. (2021) found that in Upland cotton, most ARF genes are expressed in multiple cotton tissues, with *GhARF2b* negatively regulating cotton fiber elongation. By precise regulation of auxin synthesis genes, synchronous improvement of cotton fiber quality and yield can be achieved (Zeng et al. 2019).

Gohir.D05G014550 belongs to the H₂PPase (H⁺-pyrophosphate) gene family. Pyrophosphate is produced by nearly 200 different metabolic reactions and is a byproduct of macromolecular synthesis. It is released during the production process of DNA, RNA, proteins, and polysaccharides and can be used as a phosphate donor or energy donor for each other under limited ATP levels (Wimmer et al. 2021). H⁺-PPases and PPsase1/PECP2 play important roles in the early development and dynamic balance of pyrophosphate in *Arabidopsis* (Tojo et al. 2023), and they can also enhance the low nitrogen stress tolerance of transgenic *Arabidopsis* and wheat by interacting with receptor-like protein kinases (Zhang et al. 2023). By introducing the single gene *ZmVPPI*, which encodes vacuolar H⁺-pyrophosphatase, into maize, the drought resistance of maize can be improved (Liu et al. 2023). In cotton, overexpression of the *Arabidopsis* vacuolar pyrophosphatase gene (*AVPI*) can improve drought and salt tolerance, and under dryland conditions, overexpression of *AVPI* in cotton results in a fiber yield increase of at least 20% compared to the wild type (Zhang et al. 2011). Under salt stress, adjusting the application amount of potassium fertilizer can affect the activities of H⁺-PPase and H-ATPase, thereby affecting the metabolism of osmotic fluid during cotton fiber

development and ultimately affecting cotton fiber elongation (Yu et al. 2023).

In this study, four genes related to cotton fiber quality development were identified by sequencing and analysis of lncRNAs, and these candidate genes were verified by VIGS functional verification. The results of the gene expression analysis showed that in both the leaves and fibers, the gene expression levels of gene-silenced cotton lines significantly decreased compared to negative controls, indicating that VIGS effectively silenced the related candidate genes. After the relevant genes were silenced, the fiber quality phenotype of each plant was tested. The results showed that gene-silenced plants for all the four target genes had significantly lower fiber strength than the negative control plants, indicating that these four genes had positive effects on improving cotton fiber strength. In terms of fiber uniformity, the *Gomus.A12G023400* gene-silenced line was significantly lower than that of the negative control, indicating a positive response of this gene in improving fiber uniformity. Similarly, in fiber micronaire, the *Gomus.A05G281300* gene-silenced line was significantly higher than that of the negative control, indicating that the gene had a positive regulatory effect on reducing micronaire value and improving fiber quality. The silencing of the candidate genes revealed different mechanisms for regulating fiber development and fiber quality formation, and these four genes can be used to breed cotton lines with improved fiber quality.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-024-04541-z>.

Author Contribution statement KW, BW, AHP, and PWC designed the research. MJ, HF, HG, TGM, JH, and WF performed the bioinformatics analysis. MJ, XH, JXZ, YC, and JZ performed the gene function validation experiments. BW and MJ wrote the manuscript. KW, AHP, and PWC commented on the manuscript. AD and MKRK participated in the data analysis. All the authors have read and approved the final manuscript.

Funding We appreciate financial help from the National Key R&D Program of China (2021YFE0101200), the Pakistan Science Foundation, PSF/CRP/18th Protocol (07); the National Natural Science Foundation of China (32070544, 31000729); and the Key Research and Development Project of Jiangsu Province, China (Modern Agriculture, BE2022364).

Data Availability All data generated in the current study are included in this published article and its supplementary information files.

Declarations

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

References

Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF (2019) Unraveling *cis* and *trans* regulatory evolution during cotton


- domestication. *Nat Commun* 10(1):5399. <https://doi.org/10.1038/s41467-019-13386-w>
- Bhatia G, Upadhyay SK, Upadhyay A, Kashmir Singh K (2021) Investigation of long non-coding RNAs as regulatory players of grapevine response to powdery and downy mildew infection. *BMC Plant Biol* 21(1):26. <https://doi.org/10.1186/s12870-021-03059-6>
- Breitel DA, Chappell-Maor L, Meir S, Panizel I, Puig CP, Hao Y, Yifhar T, Yasuor H, Zouine M, Bouzayen M, Richart AG, Rogachev I, Aharoni A (2016) AUXIN RESPONSE FACTOR 2 intersects hormonal signals in the regulation of tomato fruit ripening. *PLoS Genet* 12(3):e1005903. <https://doi.org/10.1371/journal.pgen.1005903>
- Chen Q, Wang W, Khanal S, Han J, Zhang M, Chen Y, Li Z, Wang K, Paterson AH, Yu J, Chee PW, Wang B (2021) Transcriptome analysis reveals genes potentially related to high fiber strength in a *Gossypium hirsutum* line IL9 with *Gossypium mustelinum* introgression. *Genome* 64(11):985–995. <https://doi.org/10.1139/gen-2020-0177>
- Chen Q, Wang W, Wang C, Zhang M, Yu J, Zhang Y, Yuan B, Ding Y, Jones DC, Paterson AH, Chee PW, Wang B (2020) Validation of QTLs for fiber quality introgressed from *Gossypium mustelinum* by selective genotyping. *G3-Genes Genomes Genetics* 10(7):2377–2384. <https://doi.org/10.1534/g3.120.401125>
- Dong S, Xiao L, Li Z, Shen J, Yan H, Li S, Liao W, Peng M (2022) A novel long non-coding RNA, *DIR*, increases drought tolerance in cassava by modifying stress-related gene expression. *J Integr Agric* 21(9):2588–2602. <https://doi.org/10.1016/j.jia.2022.07.022>
- Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <https://doi.org/10.1093/nar/gkt1223>
- Gallagher JP, Grover CE, Rex K, Moran M, Wendel JF (2017) A new species of cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Syst Bot* 42(1):115–123. <https://doi.org/10.1600/036364417X694593>
- Guilfoyle TJ, Hagen G (2007) Auxin response factors. *Curr Opin Plant Biol* 10(5):453–460. <https://doi.org/10.1016/j.pbi.2007.08.014>
- Jia YQ, Zhao HM, Niu YN, Wang Y (2023) Identification of birch lncRNAs and mRNAs responding to salt stress and characterization of functions of lncRNA. *Hortic Resh* 10(2):277. <https://doi.org/10.1093/hr/uhac277>
- Kaelik E, Ari S, Gozukirmizi N (2019) LncRNAs: genetic and epigenetic effects in plants. *Biotechnol Biotechnol Equip* 33(1):429–439. <https://doi.org/10.1080/13102818.2019.1581085>
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280. <https://doi.org/10.1093/nar/gkh063>
- Kim ED, Sung SB (2012) Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends Plant Sci* 17(1):16–21. <https://doi.org/10.1016/j.tplants.2011.10.008>
- Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12(4):357–360. <https://doi.org/10.1038/nmeth.3317>
- Knoll M, Lodish HF, Sun L (2015) Long non-coding RNAs as regulators of the endocrine system. *Nat Rev Endocrinol* 11(3):151–160. <https://doi.org/10.1038/nrendo.2014.229>
- Kong L, Zhang Y, Ye ZQ, Liu X, Zhao S, Wei L, Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 35:W345–W349. <https://doi.org/10.1093/nar/gkm391>
- Li P, Yang H, Wang L, Liu H, Huo H, Zhang C, Liu A, Zhu A, Hu J, Lin Y, Liu L (2019a) Physiological and transcriptome analyses reveal short-term responses and formation of memory under drought stress in rice. *Front Genet* 10:55. <https://doi.org/10.3389/fgene.2019.00055>
- Li Y, Qin T, Dong N, Wei C, Zhang Y, Sun R, Dong T, Chen Q, Zhou R, Wang Q (2019b) Integrative analysis of the lncRNA and mRNA transcriptome revealed genes and pathways potentially involved in the anther abortion of cotton (*Gossypium hirsutum* L.). *Genes* 10(12):947. <https://doi.org/10.3390/genes10120947>
- Liu SX, Liu XH, Zhang XM, Chang S, Ma C, Qin F (2023) Co-expression of *ZmVPP1* with *ZmNAC111* confers robust drought resistance in maize. *Genes* 14(1):8. <https://doi.org/10.3390/genes14010008>
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Onodera C, Quach H, Smith A, Yu G, Theologis A (2005) Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of *ARF7* and *ARF19*. *Plant Cell* 17(2):444–463. <https://doi.org/10.1105/tpc.104.028316>
- Razzaq A, Zafar MM, Ali A, Hafeez A, Sharif F, Guan X, Deng X, Li P, Shi Y, Haroon M, Gong W, Ren M, Yuan Y (2022) The pivotal role of major chromosomes of sub-genomes A and D in fiber quality traits of cotton. *Front Genet* 12:642595. <https://doi.org/10.3389/fgene.2021.642595>
- Salih H, Gong W, He S, Xia W, Odongo MR, Du X (2019) Long non-coding RNAs and their potential functions in Ligon-lintless-1 mutant cotton during fiber development. *BMC Genomics* 20(1):661. <https://doi.org/10.1186/s12864-019-5978-5>
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ (2006) The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133(2):251–261. <https://doi.org/10.1242/dev.02194>
- Seo JS, Sun H, Park BS, Huang CH, Yeh SD, Jung C, Chua NH (2017) ELF18-induced long-noncoding RNA associates with mediator to enhance expression of innate immune response genes in *Arabidopsis*. *Plant Cell* 29(5):1024–1038. <https://doi.org/10.1105/tpc.16.00886>
- Suksamran R, Saithong T, Thammarongtham C, Kalapanulak S (2020) Genomic and transcriptomic analysis identified novel putative cassava lncRNAs involved in cold and drought stress. *Genes* 11(4):366. <https://doi.org/10.3390/genes11040366>
- Sun L, Luo HT, Bu DH, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* 41(17):e166. <https://doi.org/10.1093/nar/gkt646>
- Tao X, Li M, Zhao T, Feng S, Zhang H, Wang L, Han J, Gao M, Lu K, Chen Q, Zhou B, Guan X (2021) Neofunctionalization of a polyploidization-activated cotton long intergenic non-coding RNA *DANI* during drought stress regulation. *Plant Physiol* 186(4):2152–2168. <https://doi.org/10.1093/plphys/kiab179>
- Tiwari SB, Hagen G, Guilfoyle T (2003) The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell* 15(2):533–543. <https://doi.org/10.1105/tpc.008417>
- Tojo H, Tabeta H, Gunji S, Hirai MY, David P, Javot H (2023) Ferjani A (2023) Roles of type II H⁺-Ppases and PpsPase1/PECP2 in early developmental stages and PPI homeostasis of *Arabidopsis thaliana*. *Front Plant Sci* 14:1031426. <https://doi.org/10.3389/fpls.2023.1031426>

- Wang D, Pei K, Fu Y, Sun Z, Li S, Liu H, Kan T, Han B, Tao Y (2007) Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* 394(1):13–24. <https://doi.org/10.1016/j.gene.2007.01.006>
- Wang L, Hua D, He J, Duan Y, Chen Z, Hong X, Gong Z (2011) *Auxin response factor2 (ARF2)* and its regulated homeodomain gene *hb33* mediate abscisic acid response in *Arabidopsis*. *PLoS Genet* 7(7):e1002172. <https://doi.org/10.1371/journal.pgen.1002172>
- Wang JY, Feng Y, Ding XT, Huo J, Nie W (2021a) Identification of long non-coding RNAs associated with tomato fruit expansion and ripening by strand-specific paired-end RNA sequencing. *Horticulturae* 7(12):522–538. <https://doi.org/10.3390/horticulturae7120522>
- Wang L, Han J, Lu K, Li M, Gao M, Cao Z, Zhao T, Chen X, Tao X, Chen Q, Guan X (2021b) Functional examination of lncRNAs in allotetraploid *Gossypium hirsutum*. *BMC Genomics* 22(1):443. <https://doi.org/10.1186/s12864-021-07771-3>
- Wang M, Yuan D, Tu L, Gao W, Zhang X (2015) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol* 207(4):1181–1197. <https://doi.org/10.1111/nph.13429>
- Wendel JF, Flagel LE, Adams KL (2012) Jeans, genes, and genomes: cotton as a model for studying polyploidy. *Polyploidy Genome Evolut* 8(2):181–207. https://doi.org/10.1007/978-3-642-31442-1_10
- Wimmer JE, Kleinermanns K, Martin WF (2021) Pyrophosphate and irreversibility in evolution, or why PPI is not an energy currency and why nature chose triphosphates. *Front Microbiol* 12:759359. <https://doi.org/10.3389/fmicb.2021.759359>
- Yatusevich R, Fedak H, Ciesielski A, Krzyczmonik K, Kulik A, Dobrowolska G, Swiezewski S (2017) Antisense transcription represses *Arabidopsis* seed dormancy QTL *DOG1* to regulate drought tolerance. *EMBO Reports* 2017 18(12):2186–2196. <https://doi.org/10.15252/embr.201744862>
- Yu K, Ju FY, Wang Z, Sun L, Huo Y, Zhu J, Pang J, Ali S, Chen W, Wang S, Zhou Z, Tang Q, Chen B (2023) Potassium ameliorates cotton (*Gossypium hirsutum* L.) fiber length by regulating osmotic and K⁺/Na⁺ homeostasis under salt stress. *Physiol Plant* 175(1):e13842. <https://doi.org/10.1111/ppl.13842>
- Zeng JY, Zhang M, Hou L, Bai W, Yan X, Hou N, Wang H, Huang J, Zhao J, Pei Y (2019) Cytokinin inhibits cotton fiber initiation via disrupting PIN3a-mediated IAA asymmetric accumulation in ovule epidermis. *J Exp Bot* 70(12):3139–3151. <https://doi.org/10.1093/jxb/erz162>
- Zhang H, Shen G, Kuppu S, Gaxiola R, Paxton Payton P (2011) Creating drought and salt-tolerant cotton by overexpressing a vacuolar pyrophosphatase gene. *Plant Signal Behav* 6(6):861–863. <https://doi.org/10.4161/psb.6.6.15223>
- Zhang BB, Zhang XX, Zhang M, Guo L, Qi T, Wang H, Tang H, Qiao X, Shahzad K, Xing C, Wu J (2019) Transcriptome analysis implicates involvement of long noncoding RNAs in cytoplasmic male sterility and fertility restoration in cotton. *Int J Mol Sci* 20(22):5530. <https://doi.org/10.3390/ijms20225530>
- Zhang XF, Cao JF, Huang CC, Zheng Z, Liu X, Shangguan X, Wang L, Zhang Y, Chen Z (2021) Characterization of cotton ARF factors and the role of *GhARF2b* in fiber development. *BMC Genomics* 22(1):202. <https://doi.org/10.1186/s12864-021-07504-6>
- Zhang Z, Zhong H, Nan B, Xiao B (2022) Global identification and integrated analysis of heat-responsive long non-coding RNAs in contrasting rice cultivars. *Theor Appl Genet* 135:833–852. <https://doi.org/10.1007/s00122-021-04001-y>
- Zhang HJ, Chen M, Xu CJ, Liu R, Tang W, Chen K, Zhou Y, Xu Z, Chen J, Ma Y, Chen W, Sun D, Fan H (2023) H⁺-pyrophosphatases enhance low nitrogen stress tolerance in transgenic *Arabidopsis* and wheat by interacting with a receptor-like protein kinase. *Front Plant Sci* 14:1096091. <https://doi.org/10.3389/fpls.2023.1096091>
- Zheng XM, Chen YZ, Zhou YF, Shi K, Hu X, Li D, Ye H, Zhou Y, Wang K (2021) Full-length annotation with multistrategy RNA-seq uncovers transcriptional regulation of lncRNAs in cotton. *Plant Physiol* 185(1):179–195. <https://doi.org/10.1093/plphys/kiaa003>
- Zou C, Wang Q, Lu C, Yang W, Zhang Y, Cheng H, Feng X, Prosper MA, Song G (2016) Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). *Sci China Life Sci* 59(2):164–171. <https://doi.org/10.1007/s11427-016-5000-2>
- Zou X, Ali F, Jin S, Li F, Wang Z (2022) RNA-Seq with a novel glabrous-ZM24f reveals some key lncRNAs and the associated targets in fiber initiation of cotton. *BMC Plant Biol* 22(1):61. <https://doi.org/10.1186/s12870-022-03444-9>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Baohua Wang¹  · Meijun Ji¹ · Hui Fang¹ · Haijing Gu¹ · Teame Gereziher Mehari¹ · Jinlei Han¹ · Wenxiang Feng¹ · Xuehan Huo² · Jingxia Zhang² · Yu Chen² · Jun Zhang² · Allah Ditta³ · Muhammad K. R. Khan³ · Andrew H. Paterson⁴ · Peng W. Chee⁵ · Kai Wang¹

✉ Baohua Wang
bhwang@ntu.edu.cn

✉ Andrew H. Paterson
paterson@uga.edu

✉ Peng W. Chee
pwchee@uga.edu

✉ Kai Wang
kwang5@ntu.edu.cn

¹ School of Life Sciences, Nantong University,
Nantong 226019, Jiangsu, China

² Key Laboratory of Cotton Breeding and Cultivation
in Huang-Huai-Hai Plain, Ministry of Agriculture and Rural
Affairs of China, Institute of Industrial Crops, Shandong
Academy of Agricultural Sciences, Jinan 250100, Shandong,
China

³ Plant Breeding and Genetics Division, Nuclear Institute
for Agriculture and Biology, Faisalabad 38000, Pakistan

⁴ Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA

⁵ Department of Crop and Soil Sciences, University of Georgia, Tifton, GA 31793, USA