



Transcriptional Landscape of Cotton Fiber Development and Its Alliance With Fiber-Associated Traits

Priti Prasad^{1,2}, Uzma Khatoon^{1,3†}, Rishi Kumar Verma^{1,2†}, Shahre Aalam¹, Ajay Kumar¹, Debashish Mohapatra⁴, Parthasarathi Bhattacharya⁴, Sumit K. Bag^{1,2*} and Samir V. Sawant^{1,2*}

¹ Division of Molecular Biology and Biotechnology, CSIR-National Botanical Research Institute, Lucknow, India, ² Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India, ³ Department of Botany, University of Lucknow, Lucknow, India, ⁴ Tierra Seed Science Pvt. Ltd., Hyderabad, India

OPEN ACCESS

Edited by:

Bhupendra Chaudhary,
Gautam Buddha University, India

Reviewed by:

Dhananjay K. Pandey,
Amity University, Ranchi, India
Harsh Chauhan,
Indian Institute of Technology Roorkee
(IITR), India

*Correspondence:

Sumit K. Bag
sumit.bag@nbri.res.in
Samir V. Sawant
samirsawant@nbri.res.in

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Bioinformatics,
a section of the journal
Frontiers in Plant Science

Received: 09 November 2021

Accepted: 10 January 2022

Published: 24 February 2022

Citation:

Prasad P, Khatoon U, Verma RK, Aalam S, Kumar A, Mohapatra D, Bhattacharya P, Bag SK and Sawant SV (2022) Transcriptional Landscape of Cotton Fiber Development and Its Alliance With Fiber-Associated Traits. *Front. Plant Sci.* 13:811655. doi: 10.3389/fpls.2022.811655

Cotton fiber development is still an intriguing question to understand fiber commitment and development. At different fiber developmental stages, many genes change their expression pattern and have a pivotal role in fiber quality and yield. Recently, numerous studies have been conducted for transcriptional regulation of fiber, and raw data were deposited to the public repository for comprehensive integrative analysis. Here, we remapped > 380 cotton RNAseq data with uniform mapping strategies that span ~400 fold coverage to the genome. We identified stage-specific features related to fiber cell commitment, initiation, elongation, and Secondary Cell Wall (SCW) synthesis and their putative cis-regulatory elements for the specific regulation in fiber development. We also mined Exclusively Expressed Transcripts (EETs) that were positively selected during cotton fiber evolution and domestication. Furthermore, the expression of EETs was validated in 100 cotton genotypes through the nCounter assay and correlated with different fiber-related traits. Thus, our data mining study reveals several important features related to cotton fiber development and improvement, which were consolidated in the “CottonExpress-omics” database.

Keywords: data-mining, stage-specific modules, cell commitment, cis-regulatory variation, fiber development, nCounter

INTRODUCTION

A transcriptome is a repertoire for protein-coding and non-coding RNAs of different tissues that are crucial in delineating the molecular basis of any phenotypic plasticity. Since the last decade, more than 175 thousand RNAseq data of several plant species have been deposited in Sequence Read Archive (SRA) alone. Thus, making sense of such public archives for reproducibility and reusability of data was highly recommended for providing new biological insights (Rung and Brazma, 2013). Recently, refining of *Arabidopsis* genome has been achieved through the reuse of tissue-specific online submitted RNAseq libraries, which yielded 635 novel genes and 452 obsolete protein-coding genes (Cheng et al., 2017). Similarly, the co-expression network analysis of publicly available transcriptome data in *Arabidopsis* resulted in identifying trait-related modules (Liu et al., 2019) and the nitrate transporter in different tissues (He et al., 2016), whereas the meiosis-related modules were identified in hexaploid wheat (Alabdullah et al., 2019). The reanalysis of RNAseq

datasets leads to the identification of “pan-network” and “core-network” for a condition-specific gene expression in *Arabidopsis* (He and Maslov, 2016; Mercatelli et al., 2020). In another study, pathogenesis-related RNAseq data were curated to provide cues for plant immunity responses in *Arabidopsis*, maize, wheat, and rice (Qi et al., 2018).

Cotton is one of the most important natural fiber crops, cultivated in more than 70 countries around the world, which contributed to 95% of global cotton production (Chen et al., 2007). The genus *Gossypium* is composed of 45 diploid and seven tetraploid species (Yuan et al., 2021). Tetraploid cotton species are natural allotetraploids formed by interspecific hybridization of A and D diploid genomes followed by polyploidization. The significant evolutionary pressure involving genetic and epigenetic modifications in the cotton genome is thought to be responsible for the domestications of modern-day cotton having much elongated and strengthened fiber (Paterson et al., 2012; Guan et al., 2014; Song et al., 2017; Li et al., 2019; Huang et al., 2021). Biologically, fibers are single-cell seed trichome that originates at the outer integument of the ovule surface in a highly polarized manner. Thus, they provide a model system for studying cell commitment, elongation, and cell wall deposition (Kim and Triplett, 2001). Fiber development involves four distinguished but overlapping stages, viz., initiation, elongation, secondary cell wall biosynthesis, and maturation. The fate of ovule epidermal cells is committed before and/or on the day of anthesis, which is coordinated by different regulation machinery (Qin and Zhu, 2011; Zhang et al., 2017; Prakash et al., 2020; Tian and Zhang, 2021). Thus, we named this stage as a fiber commitment stage, which ultimately decided the total fiber yield (Wang L. et al., 2021). Cotton fiber yield and its quality are essential agronomical traits, and thereby, significant efforts have been made to understand its biological mechanisms. The two independent groups sequenced the draft genome of *Gossypium hirsutum* genome in 2015 (Li et al., 2015; Zhang et al., 2015) that revolutionized the cotton research, and resequencing efforts have also significantly improvised the reference genome (Hu et al., 2019; Wang et al., 2019; Yang et al., 2019; Chen et al., 2020; Huang et al., 2020). The high-quality assembly provides reliable genomic information like genome architecture, gene annotation, gene expansion during the cotton speciation and domestication processes.

However, before the cotton genome sequencing projects, numerous studies unravel the molecular mechanisms of cotton fiber development (Arpat et al., 2004; Wilkins and Arpat, 2005; Samuel Yang et al., 2006; Shi et al., 2006; Udall et al., 2006, 2007; Taliencio and Boykin, 2007; Liu et al., 2012; Gilbert et al., 2013; Salih et al., 2016). The significant RNAseq efforts were also examined to understand the fiber-specific expression dynamics in contrasting cotton genotypes (Nigam et al., 2014; Qin et al., 2019), fuzzless/lintless mutant (*fl*) (Wang et al., 2010; Ma Q. F. et al., 2016), Ligon lintless (*Li1*), and other mutant lines (Ding et al., 2014; Salih et al., 2016). Additionally, transcript profiling was used to quantify the gene expression changes at different developmental stages (Naoumkina et al., 2014; Wang et al., 2015; Guo et al., 2016; Hu et al., 2018; Ando et al., 2021;

Li et al., 2021), which was further used for gene model prediction (Keilwagen et al., 2018). Thus RNAseq study can be broadly categorized before the availability of the draft cotton genome (Peng et al., 2014; Zhang et al., 2016; Parekh et al., 2018) and post-sequencing (Ma D. et al., 2016; Fang et al., 2017; Ma et al., 2018; Qin et al., 2019). The methodological variation in RNA sample preparation, sequencing technology, diverse mapping algorithm, and gene expression quantification leads to creating an obscure picture for defining the fiber dynamics at different fiber developmental stages (Corchete et al., 2020). Nevertheless, no coordinated efforts took into account to make use of publicly available RNAseq data to mine the fiber-specific features in a homogenous computational pipeline (Corchete et al., 2020).

In our current study, we used the high-quality reference genome of *Gossypium hirsutum* (Wang et al., 2019) and strengthened our understanding at the transcript level by reusing the publicly available cotton transcriptome data. We processed and clustered RNAseq data of different tissues by making use of a high-performance computing (HPC) facility. The clustering based on the coexpression network analysis identified fiber-specific genes that might have a distinct role at different stages of fiber development. The subgenome biased expression was also observed in our data with significant motif variation. Our comprehensive computational effort leads to identifying the stage-specific modules and Exclusively Expressed Transcripts (EETs). The expression level of identified EETs in 100 cotton genotypes and their correlation with different fiber quality and yield-related traits substantiates the importance of RNAseq data mining in cotton fiber development.

MATERIALS AND METHODS

Differential Quantification and Cluster Analysis of Cotton Transcriptome Data

More than 380 mock-treated or untreated cotton RNAseq datasets of 40 SRA accessions were collected from the NCBI (**Supplementary File 1**). These datasets, having 175 bp average read length, comprised 18 tissues and 34 cultivars. SRA data were quality checked followed by the adaptor trimming by the Trimmomatic tool (Bolger et al., 2014) wherever required. Eight billion filtered reads were mapped on the *Gossypium hirsutum* reference genome (Wang et al., 2019) through the STAR aligner (Dobin et al., 2013) with default settings. Three hundred thirty-six samples of selected five tissues (ovule, fiber, leaf, root, and seed) were grouped into two major tissues: OF (ovule + fiber) and LRS (leaf + root + seed) for fiber-specific differential quantification using the Ballgown R package (Pertea et al., 2016). Differentially expressed genes (DEGs) were further subjected to the Weighted Gene Co-expression Network Analysis (Langfelder and Horvath, 2008) that was categorized into different modules. Block-wise modules were used with the soft threshold power of 10 to generate the module-wise clustering where the minimum module size was set to 10 with unsigned Topological Overlap Matrix. Clustering was based on co-expression behavior among the genes. By defining the degree of centrality between genes, the hub gene was also identified for these modules, which

were eventually visualized through the Cytoscape plugin (Paul Shannon et al., 1971).

Consolidation of Different DPAs Into Four Overlapping Stages

To check the expression profiling of different modules in fiber development stages, we consolidated the uniquely mapped reads from publicly available SRA datasets (**Supplementary File 2**) in four major subsets according to the different DPAs, viz: “–3 DPA,” “–1 DPA,” “0 DPA” into commitment, “1 DPA,” “3 DPA” were merged into initiation stages, “5 DPA,” “8 DPA,” “10 DPA,” “12 DPA,” “14 DPA,” “15–16DPA,” “18 DPA,” “20 DPA” grouped for the elongation, while “21 DPA” “25–28 DPA,” “30 DPA,” “35 DPA,” and “40 DPA” were defined for SCW stages. All four subsets were assumed to be representative of fiber overlapping development stages. Furthermore, we quantitated the transcripts through the StringTie assembler (Pertea et al., 2016) for each stage. These subsets were used to check the *in silico* expression pattern for downstream analysis.

Functional Enrichment and *Cis*-Regulatory Elements Identification of Differently Clustered Modules

Gene ontology (GO) analysis of differentially regulated genes was carried out through the AgriGO database (Du et al., 2010). Significantly enriched, no-redundant GO Terms were visualized through the R package (R Core Team, 2013). Mapman analysis was carried out where data points in each bin were averaged, and the resulting *p*-value was adjusted to the Bonferroni. Motif enrichment analysis was carried out by the MEME suite for at most 10 motifs by using “–mod zoopsminw 4 –maxw 10” parameters for the DEGs and random datasets.

To check the regulatory elements in all six modules, promoter analysis was conducted through the PlantPAN3.0 database (Chow et al., 2019). We retrieved the genomic sequences from 1,000 bp upstream of TSS of each gene in all modules and scanned further for the identification of the *cis*-regulatory elements (CRE) binding site. The frequency binding site for CRE for each module was calculated as – number of TF binding sites in the promoter sequences of each module multiplied by the motif length, which is further divided by the total number of genes present in that module multiplied by 100.

Identification and *in silico* Validation of Exclusively Expressed Transcripts in Fiber Tissue

We accessed the SRA accessions “SRP049496” data and mapped the filtered reads on the reference genome to quantitate the exclusively expressed transcripts level at 0 DPA and 5 DPA in the natural fiberless mutant. Additionally, to scrutinize the expression status of exclusively expressed transcripts during the domestication process, we retrieved the RNAseq data of TM-1, domesticated cotton and TX2094, wild cotton cultivar at 10 DPA and 20 DPA fiber samples from the SRA accession “SRP017061.” All the accessed SRA data were processed in a similar way by which the previously downloaded SRA data were processed.

Assembly of Unmapped Reads

Approximately, 1 billion unmapped reads were *de novo* assembled independently by the trinity assembler (Haas et al., 2013) with a contig length of more than 300 bp. A total of 482,285 contigs were generated and blast with the nucleotide datasets to select the best hits. With >90% similarity on 80% of coverage length of the blast query, we selected 25 thousand contigs to cluster into Supertranscripts using Lace software (Davidson et al., 2017). Unmapped reads were further mapped on assembled supertranscripts. These supertranscripts were used as a reference for quantification of unmapped reads followed by the functional assignment by GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Ogata et al., 1999).

RNA Extraction and RT PCR Validation

Total RNA was isolated from cotton cultivar Coker-312 at different stages of ovules (–3 and 0 DPA), fibers (10, 21, and 30 DPA), and leaf using the SIGMA Spectrum™ total RNA kit according to the protocol of the manufacturer. DNase treatment was done using the Ambion TURBO™ DNase kit, and RNA integrity was checked by electrophoresis. About 2 μg of DNaseI treated total RNA of each developmental stage was taken for the first-strand cDNA synthesis using SuperscriptIII (Invitrogen). The PCR was performed on ABI 7500 Fast Real-Time PCR Machine (Applied Biosystems, United States) using Fast SYBR™ Green Master Mix (Applied Biosystems). Furthermore, the relative expression levels (fold change) were determined from analysis of relative gene expression data using real time quantitative PCR and the $2^{(-\Delta\Delta CT)}$ method (Schmittgen and Livak, 2001). UBQ14 (Ghir_D10G001850.1) was used as an endogenous control for normalization. Statistical analysis was done on two biological and three technical replicates for each developmental stage wherein the error bar represents the standard error. The list of primers used in this study was provided in **Supplementary File 3**.

Planting and Phenotyping of 100 Cotton Genotypes

Phenotyping of different fiber quality and fiber yield-related traits were carried out in the 100 naturally available cotton genotypes that were grown at two different locations (Hyderabad and Aurangabad) of India in the 2018 and 2019 growing seasons. All genotypes were planted in an experimental field with three replicates in April and were harvested in mid of October.

Thirty naturally opened bolls were randomly picked from each replicate for ginning processes. Fiber samples of 30- to 50-gram weight were prepared for fiber testing. First, the samples were conditioned at $25^{\circ}\text{C} \pm 2$ temperature and $65\% \pm 2$ relative humidity for 24 h. The preconditioned samples were put for the High Volume Instrument (HVI) test for different fiber parameters. Thus, the fiber length (mm), uniformity index (%), fiber strength (g/tex), fiber elongation (%), and micronaire (μg/inch) were recorded for the fiber quality traits. Fiber yield-related traits namely, boll number, boll weight (g), and cotton yield per plot (Kg) were also registered in our selected genotypes.

Nanostring nCounter Gene Expression Validation

Fiber samples from 0, 6, and 19 DPA were checked for RNA integrity, while a Nanodrop spectrophotometer was used to access the quantification of the total RNA. A total of 150 ng of total RNA was added to a master mix that contained a hybridization buffer, a gene-specific pool of probes, reported tags, and universal capture tags and hybridized at 65°C for 16 h. Gene-specific probes and reporter tags were designed by NanoString Technologies Inc., and provided in **Supplementary File 4**. The good quality samples were transferred to a nCounter cartridge followed by loading onto the Nanostring nCounter Sprint Profiler for hybridization and immobilization. Scanning of cartridge had been done by nCounter digital analyzer. The NanoString Norm package (Waggott et al., 2012) in R was used for quality check and data analysis. The mean expression value of three endogenous control genes, *viz.*, EF1a-5, UBQ-14, Actin-4 was used for normalizing the nCounter read count.

RESULTS

A Comprehensive Gene Expression Atlas of Cotton RNAseq for Extraction of Fiber-Specific Features

For the parallelization of billions of reads of 40 Sequence Read Archive (SRA) accessions (**Supplementary File 1**), we used multiple nodes on the HPC cluster and processed them with the same bioinformatics pipeline (**Figure 1A**). Such large data required a specific computation infrastructure to store, analyze, and retrieve the data. The SRA samples comprised of different cotton tissues, such as anther, boll, bud, calycle, corolla, cotyledon, fiber, flower, leaf, meristem, ovule, pericarp, pistil, root, seed, stamen, stem, and torus of the different cultivars (**Supplementary Figure 1**). We considered data only from five tissues (ovule, fiber, leaf, root, and seed) as those having the highest number of SRA samples (336 runs) (**Figure 1B**) to represent the fiber and non-fiber tissues. Out of approximately eight billion filtered reads, 6.1 billion reads were uniquely mapped to the cotton genome (**Figure 1C**) representing 400× coverage to ~2.8 Gb genome size. Approximately 0.8 million unmapped reads were also processed independently and assembled into super-transcripts (**Figure 1A**). Total 88 super-transcripts were quantified having the Reads Per Kilobase Million (RPKM) value > 1 (**Supplementary File 5**). The GO and KEGG pathway analysis indicates their role in metabolic pathways and oxidative phosphorylation, as well as associated with nutrient reservoir and seed maturation activity (**Supplementary Figure 2**).

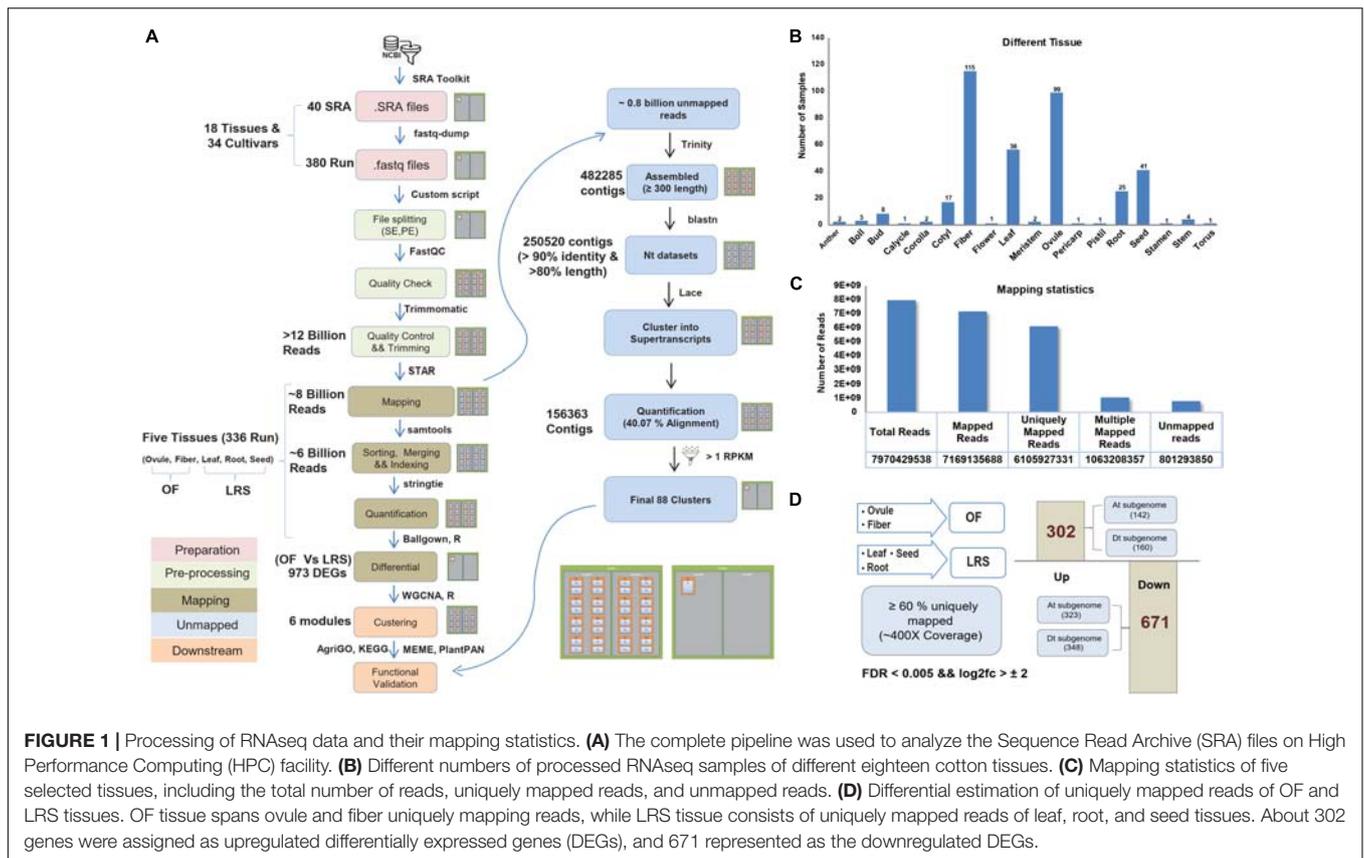
The primary aim of this study is to develop a comprehensive gene expression atlas to extract the fiber-specific features at different fiber development stages. We pulled uniquely mapped reads (**Supplementary File 2**) from the ovule and fiber tissues into one group and termed them as OF, while uniquely mapped reads from leaf, root, and seeds were pulled together and termed as LRS (**Supplementary File 6**). Differential expression analysis identified a total of 973 DEGs (FDR < 0.005 and log2fold

change > 2) between OF and LRS, out of which 302 genes were upregulated and 671 genes were downregulated (**Supplementary File 7**). At the subgenome level, 142 and 160 genes were upregulated, whereas 323 and 348 genes were downregulated for cotton At and Dt subgenome, respectively (**Figure 1D**). The functional relevance of 302 upregulated DEGs for fiber development was validated by comparing their expression in the wild cotton (Xu142) and its fiberless mutant (Xu142 *fl*) at 0 DPA and 5 DPA fiber developmental stages. The expression of 302 upregulated DEGs was found significantly lower at 0 DPA and 5 DPA in Xu142 *fl* as compared to Xu142, substantiating their role in fiber development (**Supplementary Figure 3**).

The Distinct *Cis*-Regulatory Sequence Governs the Fiber-Specific Expression of the At and Dt Genomes

The conserved motifs that could be responsible for regulating the fiber-specific expression of 302 upregulated DEGs were identified in the promoter sequences (1,000 bp upstream of TSS) using MEME suite (Bailey et al., 2015). The significance of identified motifs was ascertained by E-value in MEME and determining its significance (*p*-value) in a random-promoter dataset of a similar number. The MEME of the 302 gene's promoter identified three statistically significant motifs, and their enrichment was also significant over the random dataset. The annotation of identified motifs indicates that the topmost significantly identified motif YYYCAHCHCC showed the consensus bHLH transcription factor (TF) binding site, while the other three motifs, *viz.*, YTTTCTTTYT and RRAAAGAAR belong to the consensus C2H2 zinc finger TFs binding site (**Figure 2A**). We next identified conserved motifs belonging to At and Dt subgenome-related upregulated DEGs, separately, that identified three statistically significant motifs for each (**Figures 2B,C**). Interestingly, the topmost significant motif RAARAAGAAR showed the consensus C2H2 zinc finger TFs binding site in At subgenome (**Figure 2B**). The other two conserved motifs, *viz.*, GRKGGKGGWG and CYCYBCCTCC, belong to the consensus AP2/ERF TFs binding site, which were also significant with the random promoter dataset. The most significant motif, RAAAAGAAA in Dt subgenome, was somewhat similar to the topmost identified motif in At subgenome and belongs to the consensus C2H2 zinc finger TFs binding site. The other two motifs, *viz.*, TYTYTYTCYY and CHCCAYCHC, belonging to consensus C2H2 and AP2/ERF TFs binding sites, were quite different from the motif identified in At subgenome (**Figure 2B**). In Dt subgenome-upregulated DEGs, RAAAAGAAA and CHCCAYCHC motifs were also statistically significant over a similar number of random promoter datasets.

The differences in the conserved motifs identified for At and Dt subgenomes prompted us to correlate *cis*-regulatory element bias with the expression of At and Dt prominent genes. We thus identified 64 out of 142 At prominent and 76 out of 160 Dt prominent genes whose gene expression ratio with their respective homeologs was > 1 (**Supplementary Files 8, 9**). The 64 At prominent genes showed four statistically significant motifs, *viz.*, GGKGGKGGAG (C2H2), RAARAAGAA (MADS-box), YTTYHTTTC (HMG factors), and SCMMHGCC (AP2/ERF).



However, Dt homeolog of these 64 At prominent genes showed three conserved motifs, *viz.*, TTYTTTTTCT, RRARAARAAG, and CTCCYYCTYC, belonging to the consensus C2H2 zinc finger TF binding site (Figure 2D). In the case of Dt prominent 76 genes, two statistically significant motifs, *viz.*, CHYCHCACC (Myb/SANT) and AAAAAARAAA (MADS-box), were very distinct from the consensus motif identified in their At homeolog, namely TYTTTTYTYTC (C2H2), RAAAAWGAAA (C2H2), and CHCCWHWCWCC (AP2/ERF) (Figure 2E).

Clustering of Differentially Expressed Genes at Different Fiber Developmental Stages

Differential expression analysis was further considered for the clustering based on the weighted gene co-expression network analysis. Out of 302 upregulated genes, 8 genes were excluded as outliers, and the remaining 294 genes were clustered with the soft thresholding power of 10 to construct a block-wise module (Figure 3A). As a result of this, we obtained six modules represented in red (18 genes), yellow (27 genes), green (25 genes), brown (35 genes), turquoise (128 genes), and blue (61 genes) colored schemes. Based on eigengene relationships, block-wise clustered six modules were further grouped into two clades, Clade I representing red and yellow modules with brown and green module outliers and Clade II consisting of blue and turquoise modules (Figure 3B). The expression profiling of genes belonging to Clade I (especially for red and yellow modules) reflects a higher

expression level at the fiber commitment stage (Figure 3C), and Clade II (blue and turquoise) showed higher expression at fiber elongation and SCW stages (Figure 3E). Cytochrome P450 gene represented as a hub gene for the yellow module, whereas the red module containing genes was connected to the MADS-box family protein (Figure 3B). In the outlier modules of Clade I, brown and green modules also showed their relatively higher expression at the initiation and elongation stages (Figure 3D) with the central key players as glucose-methanol-choline oxidoreductase family protein and subtilase family protein, respectively. Glycosyl hydrolases family protein and hydroxyproline-rich glycoprotein family were the central key players for turquoise and blue modules of Clade II, respectively (Figure 3B), which represent the entire module. The downregulated genes were also clustered into four major modules (Supplementary Figure 4) and have a functional role in photosystem-related pathways (Supplementary Figure 5) and in response to abiotic and temperature stimuli (Supplementary Figure 6).

Functional Validation of Upregulated Clustered Modules and Their Cis-Regulatory Enrichment

To check the functional enrichment, GO analysis was carried out for each clade, individually suggesting the involvement of Clade I in different reproductive organ development and in TF activities, whereas Clade II was significantly enriched for the lipid transport, localization, and lipid-binding processes (Figure 4A).

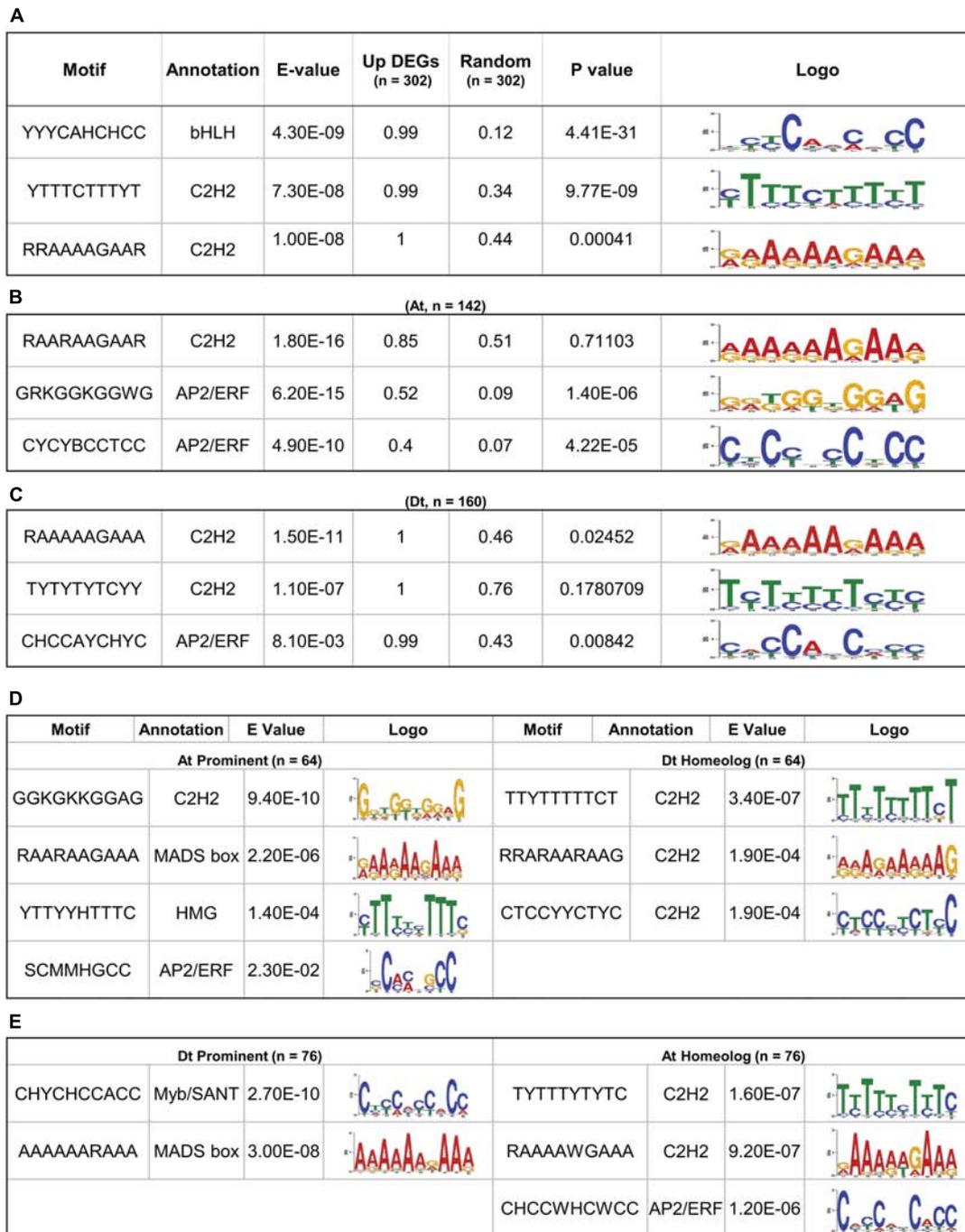
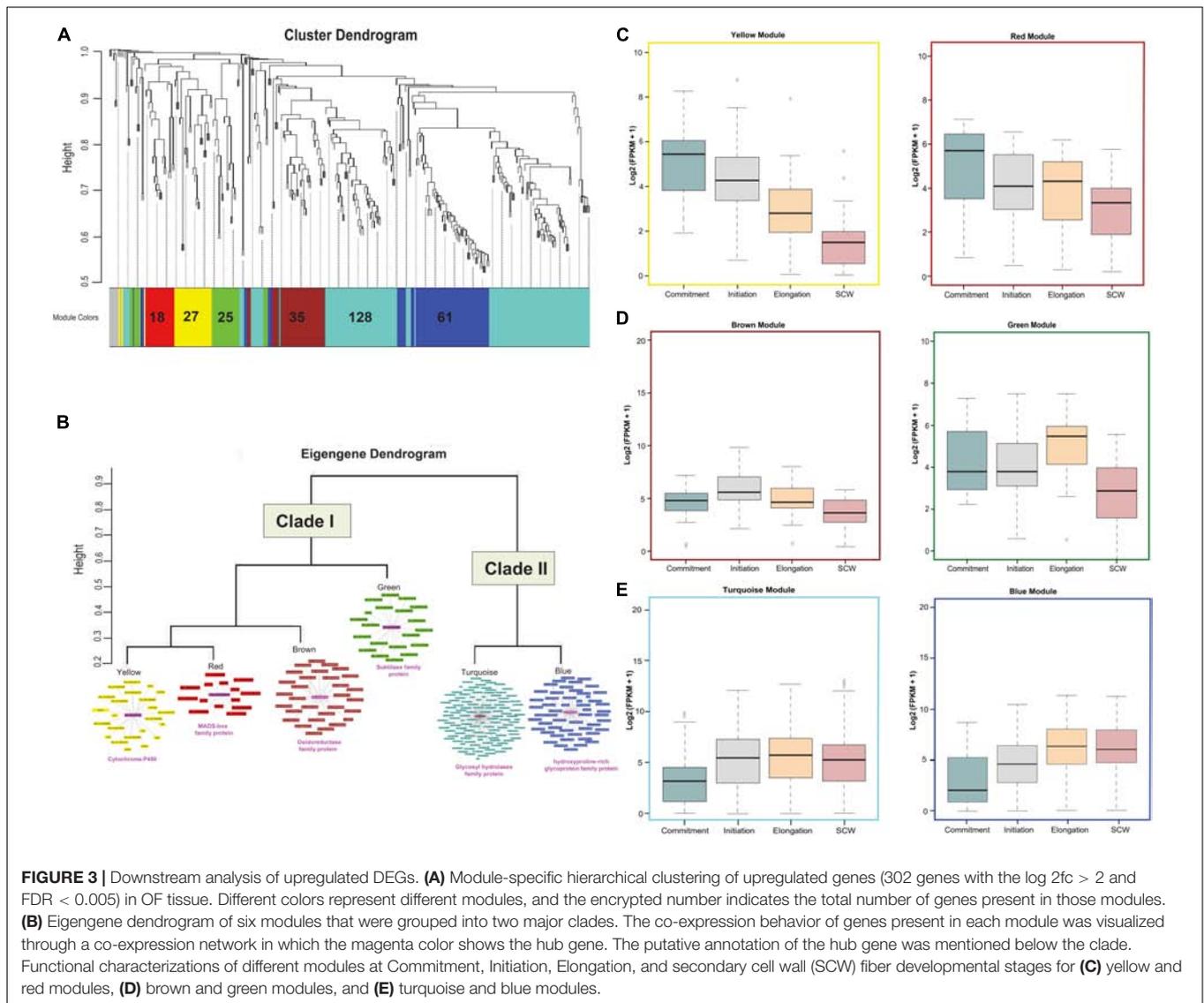


FIGURE 2 | Significant motif enrichment analysis of upregulated DEGs with the random datasets and their homeolog. **(A)** The three enrichment motifs of 302 upregulated DEGs with the E-value < 0.05 and their significant frequency of occurrences ($P < 0.05$) with a similar number of random datasets. **(B)** Significantly enriched motif of 142 At-containing DEGs and **(C)** 160 Dt-containing DEGs with the random datasets. **(D)** Motif enrichment of 64 out of 142, At prominent DEGs with their respective Dt homeologous genes and **(E)** motif enrichment of 76 out of 160 Dt prominent DEGs with their respective At homeologous genes.

Additionally, the central key player of each identified module in the form of a hub gene was validated through the real-time expression analyses (**Figure 4B**). The relatively higher expression level of hub genes at -3 DPA for red, yellow, and green modules, suggesting the gene present in the respective modules, was

involved and co-expressed for the fiber commitment process. The dual expression of cytochrome P450 (Ghir_A11G0010730.1) was validated where it was relatively expressed at fiber commitment (-3 DPA) as well as in SCW stages (21 DPA). The hub genes of the other three modules namely, brown, turquoise, and



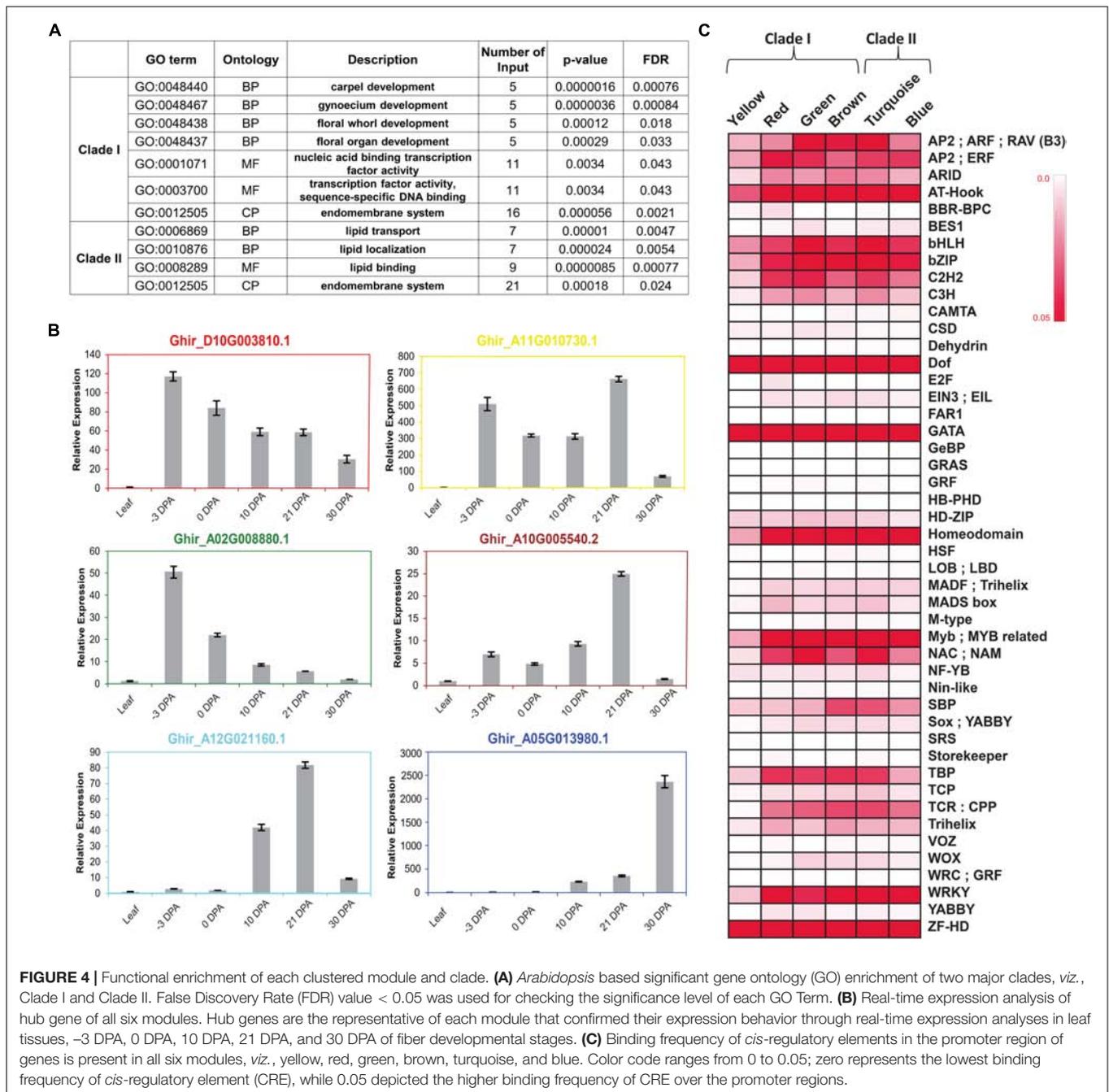
blue modules were validated for the fiber elongation and SCW stages, respectively.

The different expression levels of each module at different fiber developmental stages might be regulated by the various CREs present in their promoter region. We identified the binding frequency of the CRE in 1,000 bp upstream of Transcriptional Start Site (TSS) using the PlantPAN3.0 database (Chow et al., 2019). The higher binding frequency of CRE in each module represents its pivotal role to regulate the expression of its respective genes. A total of 47 TF binding sites were identified in which Dof, GATA, and ZF-HD TFs-binding frequency were very prominent in all six modules (**Figure 4C**). Homeodomain, MYB/Myb related, NAC, and WRKY TFs-binding sites frequency were comparatively higher in all modules except for yellow. Four modules, viz., turquoise, red, green, and brown modules have high TATA Binding Protein (TBP) frequency while low in blue and yellow

modules. Squamosa promoter binding protein (SBP) has a higher binding frequency in turquoise and brown modules, which was designated as an elongation-specific module (**Figures 3E, 4C**).

Module-Specific Pathway Enrichment of Upregulated Genes in Different Fiber Developmental Stages

Mapman pathway analysis revealed that the modules present in Clade II were involved in cellular respiration as well as in carbohydrate, sucrose, starch, and lipid metabolism (**Figures 5A,B**). Thus, it was associated with cell wall formation through the upregulated expression of pectin and many glycoproteins (**Figure 5D**). Many cytoskeletons-related components like alpha, beta-tubulin, and actin filament were also upregulated in Clade II (**Figure 5C**) with oxidoreductase, transferases, and hydrolases like enzymatic activities (**Figure 5H**).



The turquoise module of Clade II was involved in chromatin modification (Figure 5E) and the secondary metabolism pathway (Figure 5G), whereas the blue module assists protein modification and phosphorylation *via* tyrosine kinases (TKL) (Figure 5F). The upregulated genes of these two modules were involved in solute transport through the involvement of ABC, VHP PPase, MFS, and TOC transporter (Figure 5I) while DMT and MFS transporters were only upregulated in the Clade I module. In the presence of the extensin gene (Figure 5D), the yellow module is involved in sucrose degradation and thus in cell wall formation (Figure 5A).

The auxin hormone pathway was also enriched in the yellow module (Figure 5J), along with the MYB superfamily TFs (Figure 5K). Other TFs like the homeobox superfamily, MADS-box, and trihelix were also aligned with this module, whereas bZIP TFs were upregulated for the green module and AP2/ERF and homeobox superfamily for the turquoise module. Moreover, the brown module's genes facilitated the cuticle formation, phytosterol conjugation, and pectin modification (Figure 5D) with ethylene hormone responsive genes, while the red module manifested the cytokinin perception and signal transduction (Figure 5J).

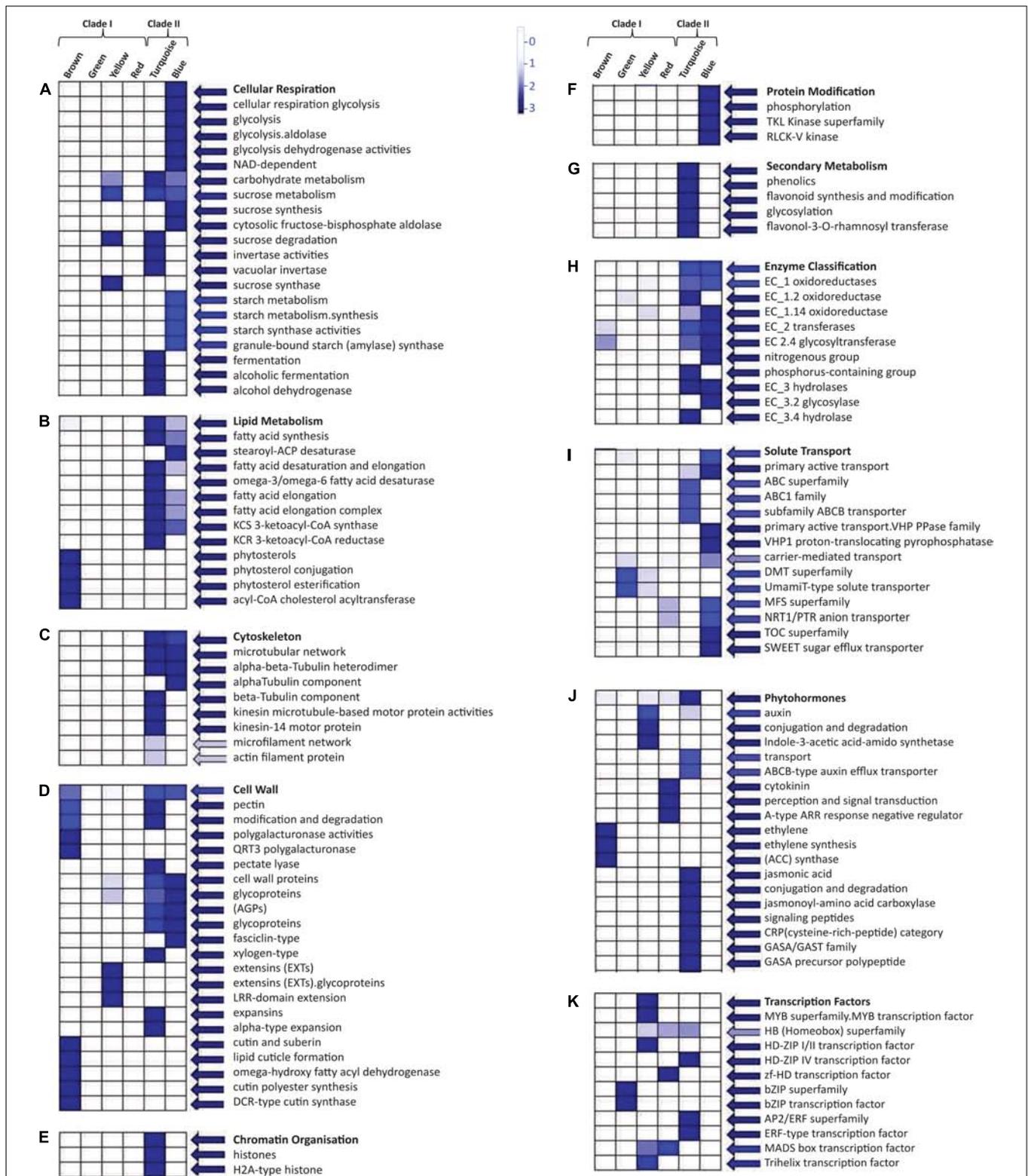


FIGURE 5 | Mapman based significant functional classification of two different clades. These two clades consist of six modules that were involved in different pathways: **(A)** cellular respiration and carbohydrate metabolism, **(B)** lipid metabolism, **(C)** cytoskeleton molecules, **(D)** cell wall-related functioning, **(E)** chromatin organization, **(F)** protein modification, **(G)** secondary metabolism pathway, **(H)** enzymatic action, **(I)** solute transportation, **(J)** phytohormone, and **(K)** transcription factors. BINs color code ranging from 0 to 3 with the vertical bar that represents the log₂ expression value of each module.

The Exclusively Expressed Gene Variants in Cotton Fiber Development

Besides 302 upregulated genes in OF, we identified 20 transcripts that expressed exclusively in OF as compared to LRS. Expression of these transcripts was very significant as we processed approximately $\sim 400\times$ uniquely mapped read coverage to the cotton reference genome. It is very interesting to note that the expression of these gene variants was very specific (Table 1) to the OF tissues as compared to their other variants, which showed somewhat higher expression in non-fiber tissue (LRS). All these gene variants belong to the ELO gene, PDF2, WRKY, and bHLH TFs, several enzymes (glycerol-3 phosphatase dehydrogenase and benzoquinone reductase), and many hypothetical proteins. The real-time PCR analyses represent the typical expression as determined by *in silico* transcriptome analysis (Supplementary Figure 7). In real-time expression analysis, hypothetical protein (Ghir_D10G025770.3) expressed at -3 DPA and 0 DPA and serine threonine-protein kinase transcript (Ghir_D05G003410.3) shows significant expression at the elongation stage with null expression value in the leaf tissue.

Evolution of Exclusively Expressed Transcripts and Their Validation in 100 Cotton Genotypes

Exclusive expressed transcripts (EETs) were very specific to the fiber structures, so to substantiate their importance in the cotton genome, we checked the collective expression of these transcripts in fiberless mutant, Xu142*fl*, and wild cotton cultivar, Xu142 (Figure 6A). In fiberless mutant cotton, these transcripts show null expression at 0 DPA and 5 DPA fiber developmental stages, indicating their potential role in various stages of fiber development, especially at the initiation stage. We also checked the expression profiling of these 20 gene variants in wild cotton, TX2094, and domesticated cotton, TM-1 (Figure 6B). The cumulative expression of 20 EETs was significantly low at 10 DPA and 20 DPA in wild variety than the domesticated variety. These observations further corroborate the functional importance of these genes in the cotton fiber developmental stages during the different evolutionary processes.

We selected five highly expressed transcripts [$\log_2(\text{FPKM} + 1) > 1$], *viz.*, the ELO/SUR4 family, PDF2, NAC domain-containing protein 62 like, glycerol-3-phosphate dehydrogenase, benzoquinone reductase, and one low expressed transcript [$\log_2(\text{FPKM} + 1) < 0.1$] bifunctional epoxide hydrolase 2 for expression and correlation study in 100 cotton genotypes (Supplementary File 10). Different fiber quality and fiber-yield traits were estimated for 2 consecutive years, *viz.* 2018 at Hyderabad, 2019 at Aurangabad and Hyderabad. Aurangabad and Hyderabad are two different cities of Maharashtra and Telangana states, respectively, which were reported as the leading states in cotton acreage in India (Nagrare et al., 2020). The phenotypic relevance of these fiber quality traits was highly correlated with each other in two different seasons (2018 and 2019) (Supplementary Figure 8) and different locations (Hyderabad and Aurangabad) (Supplementary Figure 9). Expression of selected exclusively expressed transcripts was

further examined using nanostring nCounter technology at 0 DPA, 6 DPA, and 19 DPA in 100 genotypes of cotton that were grown in 2018 at Hyderabad city of India (Figure 6C). The nCounter expression analysis confirmed that PDF2, NAC domain-containing protein 62 like, and benzoquinone reductase were highly expressed at 0 DPA, while ELO/SUR4, glycerol-3-phosphate dehydrogenase, and bifunctional epoxide hydrolase 2 like gene were significantly highly expressed at 6 DPA and 19 DPA. The expression of exclusive transcripts observed in the nCounter assay confirmed their expression pattern, which was obtained through *in silico* analysis and thus further validated our results.

The Expression of Exclusively Expressed Transcripts Correlates With the Quality Parameters of Cotton Fiber

In correlation and statistical analysis of nanostring expression and fiber quality traits in 100 genotypes, the ELO/SUR4 family gene variant exhibits a significant positive correlation with fiber strength, micronaire, boll number, and cotton yield at 0 DPA (Figures 7A,B). This transcript also correlated with elongation, micronaire, and boll weight traits at 6 DPA and only with the boll weight at 19 DPA. An unlikely significant negative correlation was seen with the micronaire at 6 DPA in 2019 at Hyderabad city. The NAC domain-containing protein 62 like exhibits a negative significant correlation with the micronaire trait at 0 DPA and 6 DPA, while the positive correlation with fiber length at 6 DPA in 2 consecutive years (2018, 2019) as well as in two different locations (Hyderabad, Aurangabad). Contrasting behavior of the NAC domain gene was observed in 2018_H and 2019_A data at 0 DPA where we observed significant negative and positive correlations, respectively, with the fiber elongation trait. At 0 DPA, HD zipper protein protodermal factor 2 transcripts showed a significant positive correlation with the fiber length, uniformity index, and fiber strength, while a significant negative correlation with boll weight at 6 DPA for 2019_H. Glycerol-3-phosphate dehydrogenase and bifunctional epoxide hydrolase 2-like transcripts showed significant positive and negative correlations with elongation and micronaire, respectively, at 0 DPA (Figure 7A). The bifunctional epoxide hydrolase 2-like transcript was also positively correlated with the fiber elongation at 0 DPA and 6 DPA and cotton yield at 6 DPA. The benzoquinone reductase transcript displayed a significant negative correlation with the micronaire at 0 DPA, fiber length and boll weight at 6 DPA, and fiber length, uniformity index, fiber strength, elongation, boll number, and cotton yield at 19 DPA (Figures 7A,B). All selected transcripts show the significant negative correlation with micronaire at 0 DPA and 6 DPA, exceptional ELO at 0 DPA, and benzoquinone reductase at 6 DPA. All the mentioned correlations were significantly affected by the fiber-related traits either positively or negatively.

“CottonExpress-omics” Database

CottonExpress-omics represents the comprehensive gene expression atlas of cotton transcriptome data to unravel the fiber-specific expression features at different developmental stages

TABLE 1 | Detail information of Exclusively Expressed Transcripts (EETs), including their *Arabidopsis*-based annotation and log₂ (FPKM + 1) value in OF and LRS tissues.

Gene	Annotation	Transcripts	OF	LRS
			Log ₂ (FPKM + 1)	
Ghir_D01G000100	ELO/SUR4 family	Ghir_D01G000100.1	8.247181314	0
		Ghir_D01G000100.2	5.849377026	1.3980718
Ghir_A10G001030	HD zipper protein protodermal factor 2	Ghir_A10G001030.1	0.285908775	0.0775192
		Ghir_A10G001030.2	1.082713488	1.0809891
		Ghir_A10G001030.3	2.907309878	0
Ghir_D09G019020	NAC domain-containing protein 62-like	Ghir_D09G019020.1	2.640708144	3.0826249
		Ghir_D09G019020.2	1.314970249	1.8542175
		Ghir_D09G019020.3	1.44367189	1.9737428
		Ghir_D09G019020.4	2.6234633	0
Ghir_A09G012990	glycerol-3-phosphate dehydrogenase	Ghir_A09G012990.1	0.087107046	0.4593403
		Ghir_A09G012990.2	1.204346712	0.4166236
		Ghir_A09G012990.3	0.875061288	0.8913321
		Ghir_A09G012990.4	2.617127776	0
Ghir_D01G006540	Benzoquinone reductase	Ghir_D01G006540.1	6.268913584	3.9051894
		Ghir_D01G006540.2	0.999345589	0.690696
		Ghir_D01G006540.3	1.020565494	0
Ghir_D10G025770	Hypothetical protein	Ghir_D10G025770.1	0.92436369	0.0527247
		Ghir_D10G025770.2	1.094223231	0.2255871
		Ghir_D10G025770.3	0.534289745	0
		Ghir_D10G025770.4	1.084769108	0.1948385
Ghir_A05G038240	WRKY transcription factor 1	Ghir_A05G038240.1	4.778754111	1.9936975
		Ghir_A05G038240.2	2.929963278	0.137295
		Ghir_A05G038240.3	0.634027231	0.033738
		Ghir_A05G038240.4	0.940046443	0.0363838
		Ghir_A05G038240.5	1.671006617	0.0342636
		Ghir_A05G038240.6	0.451709621	0
Ghir_D05G003410	Serine-threonine protein kinase	Ghir_D05G003410.1	2.98741965	0.6356541
		Ghir_D05G003410.2	0.303920993	0.0050133
		Ghir_D05G003410.3	0.280408817	0
		Ghir_D05G003410.4	0.096578968	0.0057262
Ghir_D11G022110	bHLH61 like transcription factor	Ghir_D11G022110.1	0.235100837	0
		Ghir_D11G022110.2	2.182144669	1.0477708
		Ghir_D11G022110.3	1.617503963	0.7769445
Ghir_A08G022470	WVD1 like 1 isoform	Ghir_A08G022470.1	1.282175262	0.2576479
		Ghir_A08G022470.2	0.095796175	0.2643491
		Ghir_A08G022470.3	0.099241334	0.8651611
		Ghir_A08G022470.4	0.589878515	0.8751675
		Ghir_A08G022470.5	0.227819923	0
		Ghir_A08G022470.6	0.202128056	0
Ghir_A11G025670	Hypothetical protein	Ghir_A11G025670.1	1.320330436	0.0007414
		Ghir_A11G025670.2	2.417842256	0.7236608
		Ghir_A11G025670.3	0.188235059	0
Ghir_D01G005620	Hypothetical protein	Ghir_D01G005620.1	0.238984911	0.3046851
		Ghir_D01G005620.2	2.047156551	1.1493537
		Ghir_D01G005620.3	0.153845536	0
Ghir_D01G020410	Hypothetical protein	Ghir_D01G020410.1	0.888239423	0.7282053
		Ghir_D01G020410.1	0.073779111	0
Ghir_D04G015610	bifunctional epoxide hydrolase 2 like gene	Ghir_D04G015610.1	0.053649379	0
		Ghir_D04G015610.2	5.945628677	1.2868179
Ghir_A09G014480	Hypothetical protein	Ghir_A09G014480.1	0.0413986	0

(Continued)

TABLE 1 | (Continued)

Gene	Annotation	Transcripts	OF	LRS
			Log ₂ (FPKM + 1)	
Ghir_A12G003890	Hypothetical protein	<i>Ghir_A09G014480.2</i>	1.144999148	0.6043513
		Ghir_A12G003890.1	0.011841959	0
Ghir_A09G013880	Hypothetical protein	Ghir_A12G003890.2	4.091971001	0.926886
		Ghir_A09G013880.1	0.740766125	0.2872205
		Ghir_A09G013880.2	3.114859556	1.437379
		Ghir_A09G013880.3	0.781812072	0.3974635
Ghir_D12G002480	LAG1 longevity assurance homolog2-like	Ghir_A09G013880.4	0.007735155	0
		Ghir_D12G002480.1	0.001137838	0
Ghir_D05G007320	Hypothetical protein	Ghir_D12G002480.2	2.875775739	0.410619
		Ghir_D05G007320.1	8.80017E-05	0
		Ghir_D05G007320.2	7.705065032	3.6233814
		Ghir_D05G007320.3	0.616319112	0.0705858

Identified EETs were highlighted with the olive green-colored cell whose expression was null in LRS tissue. Some identified transcripts have identical nucleotide sequences, which were represented by the bold and italic fonts in the transcripts column.

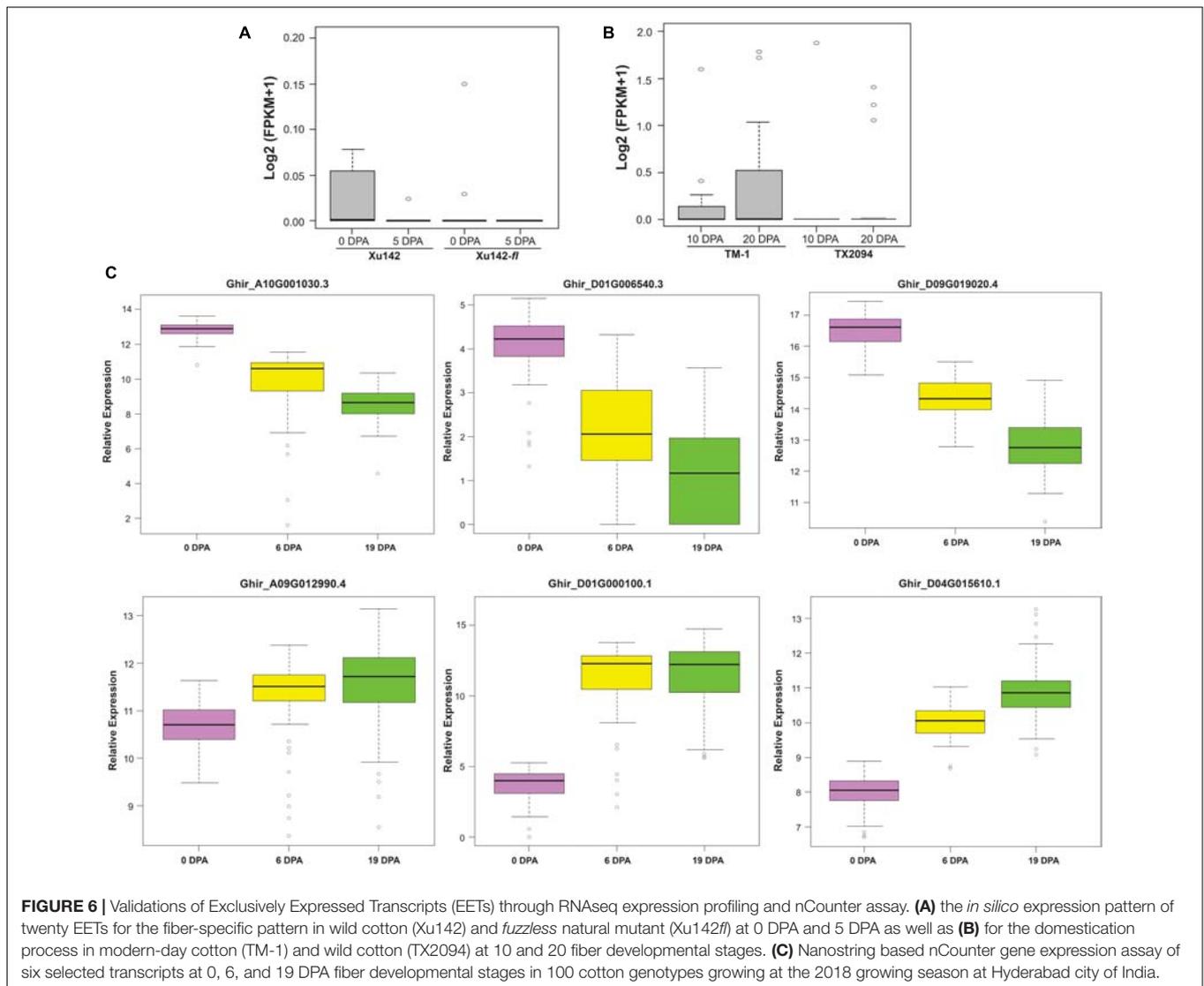
and is available at the <http://14.139.63.169/cottonexpressomics>. The database aids researchers with a comprehensive account of fiber-specific DEGs, EETs, clustered modules, and conserved motifs that were mined in the present study. Cotton community can obtain the expression value of the gene of interest from diverse cotton cultivars or for different fiber developmental stages under the “FPKM” subsection of the “Mapping Statistics” menu. The fold change expression value of OF vs. LRS and the annotation of different members of clustered modules can also be searched by users under the “Downstream” menu. Thus, this database provides elaborative information of different fiber-specific features under one roof that will be helpful for the cotton researchers and breeders for cotton fiber improvement.

DISCUSSION

The cotton genome was refined over the years by different independent groups to get the well-annotated genes and unravel the evolution dynamics for fiber-specific features (Yang et al., 2020). Additionally, to understand the molecular and biological functions of genes involved in cotton fiber growth and development, the multi-omics approach was also taken into consideration (Pang et al., 2009; Rai et al., 2013; Song et al., 2015, 2017; Zou et al., 2016; Kumar et al., 2018; Ayubov et al., 2019; Li et al., 2019). In recent years, many researchers have reannotated the cotton genome sequence by third and fourth generation sequencing techniques to get the high contiguity genome (Pan et al., 2020). Prior to cotton genome sequence being made available, several transcriptome data (Wang et al., 2010; Nigam et al., 2014; Peng et al., 2014; Cheng et al., 2016; Parekh et al., 2018) were generated that would not have been mapped properly due to the unavailability of a well-furnished reference genome. Recent efforts have been carried out for the identification of important molecular signatures for fiber-related traits, in which around fivefold to sevenfold coverage transcriptome data were generated and mapped on the draft genome of cotton

(Fang et al., 2017; Ma et al., 2018). The low coverage on the draft genome may not be used for the identification of low expressed genes and its variants that might be crucial for fiber development. Similarly, in an eQTL identification study, only 10 billion reads were used for 251 cotton accessions (Li et al., 2020) and only focused on the initiation of secondary cell wall synthesis at 15 DPA. Thus far, no studies have been available that used high coverage cotton transcriptome data on the newly assembled genome (Hu et al., 2019; Wang et al., 2019; Yang et al., 2019; Chen et al., 2020; Huang et al., 2020) for all fiber developmental stages. In our current study, we elucidated many fiber-specific features for different fiber developmental stages, categorizing them into commitment, initiation, elongation, and SCW synthesis stages where we processed and reanalyzed the 356 publicly available cotton transcriptome data on the HPC cluster. The HPC cluster accomplished the computation requirement for such a large dataset, which was processed through a similar bioinformatics pipeline in the batch script. The high-coverage data helped us in the identification of important fiber features that were not reported yet for different fiber developmental stages, which subsequently correlated with the different fiber-related traits.

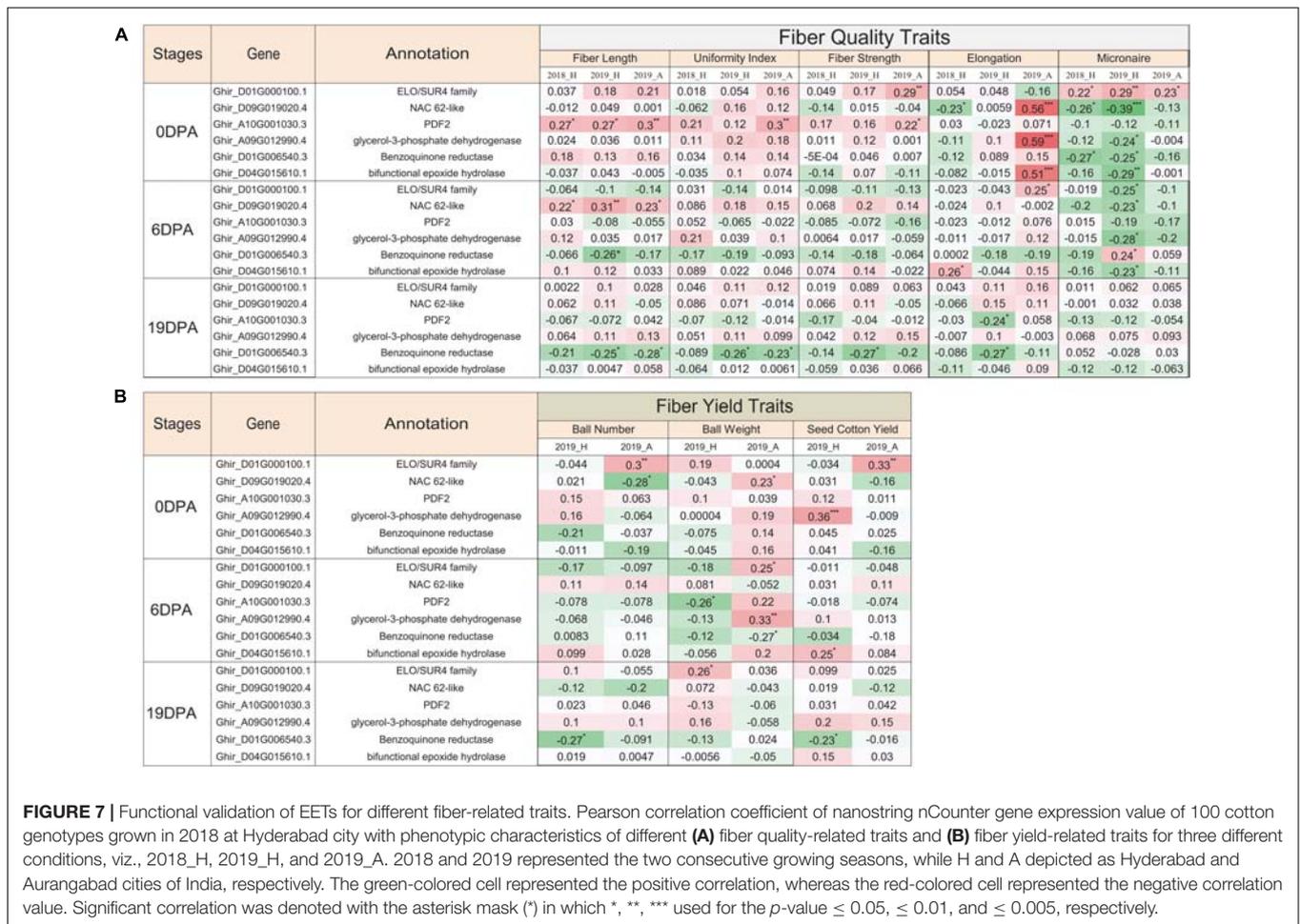
Motif enrichment analysis of 302 differentially expressed fiber-specific genes revealed a significant enrichment of *cis*-regulatory motifs belonging to bHLH and C2H2 zinc finger TFs (Figure 2A). The TFs that interact with these two motifs are required for different transcriptional regulations at fiber initiation and elongation stages (Wang L. et al., 2021; Wang N. N. et al., 2021; Wu et al., 2021). However, in *At* subgenome (142) and *Dt* subgenome specific (160) DEGs, instead of bHLH, AP2/ERF TF showed significant motif enrichment over the random datasets (Figures 2A,C), which was previously reported to control the fiber length (Mittal et al., 2015). The enrichment of distinct motifs for *At*/*Dt* subgenome indicates the distinctness of the transcriptional regulatory apparatus that interacts by which they are expressed specifically during cotton fiber development. Many reports also indicated the similar gene expression biasness at the



subgenome level during fiber development and its connection with the *cis*-regulatory motifs (Song et al., 2017; Zhao et al., 2018). Thus, we were also keen to link the biased expression between At and Dt subgenomes to *cis*-regulatory motifs in our analysis (Figures 2D,E). The MEME revealed a very distinct set of motifs that were enriched in 64 At and 76 Dt prominent genes corresponding to their homeologs, which reassure the distinct regulatory mechanisms for fiber-specific gene expression. Taken together, our motif analysis indicates fiber-specific DEGs expression that was governed by the different *cis*-regulatory mechanisms involving distinct sets of TFs. Thus, it will be interesting to explore further how the coordination between At and Dt subgenomes operates transcriptionally to achieve fiber specificity.

The comprehensive clustering of fiber-specific genes based on their expression resulted in six distinct modules, which are further grouped into two prominent clades to represent the stage-specific expression (Figure 3). Clade I significantly

represents the commitment-specific module involved in different organ developments, while Clade II module involved in the fiber cell elongation and maturation through lipid metabolism processes (Figure 4A). The yellow module's genes were annotated as MYB, HD-ZIP, MADS, and Trihelix TFs (Figure 5K and Supplementary Figure 4). The role of these TFs like MYB/MYB like (MacHado et al., 2009; Walford et al., 2011) and homeodomain leucine zipper gene (Walford et al., 2012) in the fiber initials further affirmed the specificity of this module for commitment-specific expression. Interestingly, although the expression of cell fate-determining factors - MYB and HD-ZIP is significantly high in the yellow module; however, promoter analysis indicates the poor enrichment of their binding site in the promoter of the same module (Figure 4C). This indicates that the expression of MYB and HD-ZIP is governed by different regulators at the commitment stage, but they regulate the expression of genes of other modules as represented through CRE frequency in the promoter regions (Figure 4C). Furthermore,



the significance of the yellow module is substantiated by the presence of genes belonging to the auxin and indole-3-acetic acid amido synthase pathway whose role was well established for cell fate determination (Zhang et al., 2011; Xu et al., 2019). The role of auxin efflux and influx is well documented in the fiber initials at -1 DPA (Petrásek et al., 2006; Zhang et al., 2017), and fiber protrudes through the upregulation of genes belonging to extension-related glycoproteins. Another module of Clade I, viz., red module, shows the upregulation of the cytokinin and cytokinin perception and signal transduction genes, including A type ARR negative regulator (Figure 5J). Thus, fiber development may follow a cascade where the antagonistic behavior of auxin and cytokinin (Zeng et al., 2019) has been maintained by the upregulation of A-type ARRs. The functional significance of genes identified in each module was also ascertained since many genes fall in the genic and intergenic regions that were associated with several fiber-related traits as shown in previously identified genome-wide association study (Supplementary Table 1; Ma et al., 2018).

High-coverage transcriptome data were used to identify the full sequence diversity of the genome, including lowly expressed transcripts (Fu et al., 2014). In the present study, analysis of approximately six million uniquely mapped reads

leads to identify 20 EETs which showed unique fiber specific expression. The importance of exclusivity of these transcripts in fiber tissue needs to be explored further; however, tissue-specific variants have been shown to regulate the flowering at different developmental stages in *Arabidopsis* (Rosloski et al., 2013; Ghelli et al., 2018). The tissue-specific expression of variants may be due to the presence of tissue-specific promoter architecture, which requires a tightly regulated transcriptional machinery (Srivastava et al., 2018; Das and Bansal, 2019; Schmitz et al., 2021). During the domestication of cultivated cotton species, these transcripts might be positively selected in modern cotton for fiber-specific features (Figures 6A,B). The comprehensive expression analysis of EETs using nCounter gene expression assay involving 100 genotypes and three developmental stages (0 DPA, 6 DPA, and 19 DPA) confirmed the different stage-specific expressions of EETs (Figure 6C) that correlated with different fiber-related traits (Figure 7). The significant correlations further ascertain the importance of EETs in cotton fiber development identified in our study.

One of the tested EETs, PROTODERMAL FACTOR 2 (Ghir_A10G001030.3), significantly correlated with the fiber length (Figure 7). This was consistent with the previous report wherein the role of PROTODERMAL FACTOR 1 and its

mechanism in fiber development was well documented (Deng et al., 2012). Yet another EET that belongs to the NAC family gene correlates with the fiber length (Figure 7A) and has been documented for positive and negative regulations at different developmental stages (Shang et al., 2016; Zhang et al., 2018; Sun et al., 2020). ELO transcript (Ghir_D01G000100.1) also correlates positively with the micronaire at 0 DPA that putatively involved in the synthesis of very long-chain fatty acids (VLCFs) (Qin et al., 2007a; Zhang et al., 2015), which ultimately acts as a precursor for the sphingolipid biosynthesis to promote cotton fiber elongation (Qin et al., 2007a,b). Our nCounter gene expression suggests that ELO's expression initiated early at 0 DPA; thus, it is very interesting to validate further how the early expression of ELO influences the micronaire trait. Our correlation study established an inverse relationship between fiber elongation and micronaire, which may be crucial for sustaining the yarn strength (Mathangadeera et al., 2020) by combining properties of fiber length, density, and maturity (Long et al., 2013).

Thus, our extensive and comprehensive reanalyzing of RNAseq resulted in the identification of several new genes whose expression might be crucial for fiber development, which was validated through the different approaches. Data mining identifies several important features, which were supported by the previous results and thus substantiates our findings. All the associated result of this study was systemized in the "CottonExpress-omics" database for the cotton researchers and breeders that will help the cotton community to quickly identify genes that are expressed at different developmental stages.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

REFERENCES

- Alabdullah, A. K., Borrill, P., Martin, A. C., Ramirez-Gonzalez, R. H., Hassani-Pak, K., Uauy, C., et al. (2019). A Co-expression network in hexaploid wheat reveals mostly balanced expression and lack of significant gene loss of homeologous meiotic genes upon polyploidization. *Front. Plant Sci.* 10:1325. doi: 10.3389/fpls.2019.01325
- Ando, A., Kirkbride, R. C., Jones, D. C., Grimwood, J., and Chen, Z. J. (2021). LCM and RNA-seq analyses revealed roles of cell cycle and translational regulation and homoeolog expression bias in cotton fiber cell initiation. *BMC Genomics* 22:309. doi: 10.1186/s12864-021-07579-1
- Arpat, A. B., Waugh, M., Sullivan, J. P., Gonzales, M., Frisch, D., Main, D., et al. (2004). Functional genomics of cell elongation in developing cotton fibers. *Plant Mol. Biol.* 54, 911–929. doi: 10.1007/s11103-004-0392-y
- Ayubov, M. S., Mirzakhmedov, M. H., Sripathi, V. R., Buriev, Z. T., Ubaydullaeva, K. A., Usmonov, D. E., et al. (2019). Role of MicroRNAs and small RNAs in regulation of developmental processes and agronomic traits in *Gossypium* species. *Genomics* 111, 1018–1025. doi: 10.1016/j.ygeno.2018.07.012
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Chen, Z. J., Scheffler, B. E., Dennis, E., Triplett, B. A., Zhang, T., Guo, W., et al. (2007). Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* 145, 1303–1310. doi: 10.1104/pp.107.107672
- Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., et al. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52, 525–533. doi: 10.1038/s41588-020-0614-5
- Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tpj.13415
- Cheng, W. H., Zhu, H. G., Tian, W. G., Zhu, S. H., Xiong, X. P., Sun, Y. Q., et al. (2016). De novo transcriptome analysis reveals insights into dynamic homeostasis regulation of somatic embryogenesis in upland cotton (*G. hirsutum* L.). *Plant Mol. Biol.* 92, 279–292. doi: 10.1007/s11103-016-0511-6
- Chow, C. N., Lee, T. Y., Hung, Y. C., Li, G. Z., Tseng, K. C., Liu, Y. H., et al. (2019). Plantpan3.0: a new and updated resource for reconstructing transcriptional regulatory networks from chip-seq experiments in plants. *Nucleic Acids Res.* 47, D1155–D1163. doi: 10.1093/nar/gky1081
- Corchete, L. A., Rojas, E. A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N. C., and Burguillo, F. J. (2020). Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci. Rep.* 10:19737. doi: 10.1038/s41598-020-76881-x

and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

PP and SS conceptualized the study and wrote the manuscript. PP performed all related analyses. PP and SB conceived the web-portal design. SA and PP executed the web-portal design. SB and SS helped in data interpretation. RV provided the RNA samples for qRT-PCR. UK validated the relative expression results. DM and PB provided phenotypic data of 100 genotypes of cotton. AK helped in nCounter analysis. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We acknowledge the NCBI team and all those researchers who submitted the RNAseq data to the SRA database without which it is impossible to conduct the research. We thank Council of Scientific and Industrial Research (CSIR), New Delhi, Government of India for providing the financial support to carry out this research. We also duly acknowledge the director, CSIR-NBRI, for his encouragement and support. We also thank Ratnesh Tripathi for providing a helping hand in the nCounter assay experiment. Institution manuscript number allotted to this manuscript is CSIR-NBRI_MS/2021/02/08.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.811655/full#supplementary-material>

- Das, S., and Bansal, M. (2019). Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLoS One* 14:e0212678. doi: 10.1371/journal.pone.0212678
- Davidson, N. M., Hawkins, A. D. K., and Oshlack, A. (2017). Supertranscripts: a data driven reference for analysis and visualisation of transcriptomes. *Genome Biol.* 18:148. doi: 10.1186/s13059-017-1284-1
- Deng, F., Tu, L., Tan, J., Li, Y., Nie, Y., and Zhang, X. (2012). GbPDF1 is involved in cotton fiber initiation via the core cis-element HDZIP2ATATHB2. *Plant Physiol.* 158, 890–904. doi: 10.1104/pp.111.186742
- Ding, M., Jiang, Y., Cao, Y., Lin, L., He, S., Zhou, W., et al. (2014). Gene expression profile analysis of Ligon lintless-1 (Li1) mutant reveals important genes and pathways in cotton leaf and fiber development. *Gene* 535, 273–285. doi: 10.1016/j.gene.2013.11.017
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64–W70. doi: 10.1093/nar/gkq310
- Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., et al. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* 49, 1089–1098. doi: 10.1038/ng.3887
- Fu, G. K., Xu, W., Wilhelm, J., Mindrinos, M. N., Davis, R. W., Xiao, W., et al. (2014). Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1891–1896. doi: 10.1073/pnas.1323732111
- Ghelli, R., Brunetti, P., Napoli, N., De Paolis, A., Cecchetti, V., Tsuge, T., et al. (2018). A newly identified flower-specific splice variant of AUXIN RESPONSE FACTOR8 regulates stamen elongation and endothecium lignification in *Arabidopsis*. *Plant Cell* 30, 620–637. doi: 10.1105/tpc.17.00840
- Gilbert, M. K., Turley, R. B., Kim, H. J., Li, P., Thyssen, G., Tang, Y., et al. (2013). Transcript profiling by microarray and marker analysis of the short cotton (*Gossypium hirsutum* L.) fiber mutant Ligon lintless-1 (Li1). *BMC Genomics* 14:403. doi: 10.1186/1471-2164-14-403
- Guan, X., Song, Q., and Chen, Z. J. (2014). Polyploidy and small RNA regulation of cotton fiber development. *Trends Plant Sci.* 19, 516–528. doi: 10.1016/j.tplants.2014.04.007
- Guo, K., Du, X., Tu, L., Tang, W., Wang, P., Wang, M., et al. (2016). Fibre elongation requires normal redox homeostasis modulated by cytosolic ascorbate peroxidase in cotton (*Gossypium hirsutum*). *J. Exp. Bot.* 67, 3289–3301. doi: 10.1093/jxb/erw146
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- He, F., Karve, A. A., Maslov, S., and Babst, B. A. (2016). Large-scale public transcriptomic data mining reveals a tight connection between the transport of nitrogen and other transport processes in *Arabidopsis*. *Front. Plant Sci.* 7:1207. doi: 10.3389/fpls.2016.01207
- He, F., and Maslov, S. (2016). Pan-and core-network analysis of co-expression genes in a model plant. *Sci. Rep.* 6:38956. doi: 10.1038/srep38956
- Hu, H., Wang, M., Ding, Y., Zhu, S., Zhao, G., Tu, L., et al. (2018). Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 16, 1002–1012. doi: 10.1111/pbi.12844
- Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51, 739–748. doi: 10.1038/s41588-019-0371-5
- Huang, G., Huang, J.-Q., Chen, X.-Y., and Zhu, Y.-X. (2021). Recent advances and future perspectives in cotton research. *Annu. Rev. Plant Biol.* 72, 437–462. doi: 10.1146/annurev-arplant-080720-113241
- Huang, G., Wu, Z., Percy, R. G., Bai, M., Li, Y., Frelichowski, J. E., et al. (2020). Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat. Genet.* 52, 516–524. doi: 10.1038/s41588-020-0607-4
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* 19:189. doi: 10.1186/s12859-018-2203-5
- Kim, H. J., and Triplett, B. A. (2001). Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127, 1361–1366. doi: 10.1104/pp.010724
- Kumar, V., Singh, B., Singh, S. K., Rai, K. M., Singh, S. P., Sable, A., et al. (2018). Role of GhHDA5 in H3K9 deacetylation and fiber initiation in *Gossypium hirsutum*. *Plant J.* 95, 1069–1083. doi: 10.1111/tpj.14011
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., et al. (2015). Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 33, 524–530. doi: 10.1038/nbt.3208
- Li, L., Zhang, C., Huang, J., Liu, Q., Wei, H., Wang, H., et al. (2021). Genomic analyses reveal the genetic basis of early maturity and identification of loci and candidate genes in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 19, 109–123. doi: 10.1111/pbi.13446
- Li, N., Xu, C., Zhang, A., Lv, R., Meng, X., Lin, X., et al. (2019). DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytol.* 223, 979–992. doi: 10.1111/nph.15820
- Li, Z., Wang, P., You, C., Yu, J., Zhang, X., Yan, F., et al. (2020). Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. *New Phytol.* 226, 1738–1752. doi: 10.1111/nph.16468
- Liu, K., Sun, J., Yao, L., and Yuan, Y. (2012). Transcriptome analysis reveals critical genes and key pathways for early cotton fiber elongation in Ligon lintless-1 mutant. *Genomics* 100, 42–50. doi: 10.1016/j.ygeno.2012.04.007
- Liu, W., Lin, L., Zhang, Z., Liu, S., Gao, K., Lv, Y., et al. (2019). Gene co-expression network analysis identifies trait-related modules in *Arabidopsis thaliana*. *Planta* 249, 1487–1501. doi: 10.1007/s00425-019-03102-9
- Long, R. L., Bange, M. P., Delhom, C. D., Church, J. S., and Constable, G. A. (2013). An assessment of alternative cotton fibre quality attributes and their relationship with yarn strength. *Crop Pasture Sci.* 64, 750–762. doi: 10.1071/CP12382
- Ma, D., Hu, Y., Yang, C., Liu, B., Fang, L., Wan, Q., et al. (2016). Genetic basis for glandular trichome formation in cotton. *Nat. Commun.* 7:10456. doi: 10.1038/ncomms10456
- Ma, Q. F., Wu, C. H., Wu, M., Pei, W. F., Li, X. L., Wang, W. K., et al. (2016). Integrative transcriptome, proteome, phosphoproteome and genetic mapping reveals new aspects in a fiberless mutant of cotton. *Sci. Rep.* 6:24485. doi: 10.1038/srep24485
- Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., et al. (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50, 803–813. doi: 10.1038/s41588-018-0119-7
- MacHado, A., Wu, Y., Yang, Y., Llewellyn, D. J., and Dennis, E. S. (2009). The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J.* 59, 52–62. doi: 10.1111/j.1365-313X.2009.03847.x
- Mathangadeera, R. W., Hequet, E. F., Kelly, B., Dever, J. K., and Kelly, C. M. (2020). Importance of cotton fiber elongation in fiber processing. *Ind. Crops Prod.* 147:112217. doi: 10.1016/j.indcrop.2020.112217
- Mercatelli, D., Scalambra, L., Triboli, L., Ray, F., and Giorgi, F. M. (2020). Gene regulatory network inference resources: a practical overview. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863:194430. doi: 10.1016/j.bbagr.2019.194430
- Mittal, A., Jiang, Y., Ritchie, G. L., Burke, J. J., and Rock, C. D. (2015). AtRAV1 and AtRAV2 overexpression in cotton increases fiber length differentially under drought stress and delays flowering. *Plant Sci.* 241, 78–95. doi: 10.1016/j.plantsci.2015.09.013
- Nagrle, D. T., Gawande, S. P., Gokte-Narkhedkar, N., and Waghmare, V. N. (2020). Association of phytopathogenic *Pantoea dispersa* inner boll rot of cotton (*Gossypium hirsutum* L.) in Maharashtra state, India. *Eur. J. Plant Pathol.* 158, 251–260. doi: 10.1007/s10658-020-02071-0
- Naoumkina, M., Thyssen, G., Fang, D. D., Hinchliffe, D. J., Florane, C., Yeater, K. M., et al. (2014). The Li2 mutation results in reduced subgenome expression

- bias in elongating fibers of allotetraploid cotton (*Gossypium hirsutum* L.). *PLoS One* 9:e0090830. doi: 10.1371/journal.pone.0090830
- Nigam, D., Kavita, P., Tripathi, R. K., Ranjan, A., Goel, R., Asif, M., et al. (2014). Transcriptome dynamics during fibre development in contrasting genotypes of *Gossypium hirsutum* L. *Plant Biotechnol. J.* 12, 204–218. doi: 10.1111/pbi.12129
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29
- Pan, Y., Meng, F., and Wang, X. (2020). Sequencing multiple cotton genomes reveals complex structures and lays foundation for breeding. *Front. Plant Sci.* 11:e560096. doi: 10.3389/fpls.2020.560096
- Pang, M., Woodward, A. W., Agarwal, V., Guan, X., Ha, M., Ramachandran, V., et al. (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. doi: 10.1186/gb-2009-10-11-r122
- Parekh, M. J., Kumar, S., Fougat, R. S., Zala, H. N., and Pandit, R. J. (2018). Transcriptomic profiling of developing fiber in levant cotton (*Gossypium herbaceum* L.). *Funct. Integr. Genomics* 18, 211–223. doi: 10.1007/s10142-017-0586-4
- Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798
- Peng, Z., He, S., Gong, W., Sun, J., Pan, Z., Xu, F., et al. (2014). Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. *BMC Genomics* 15:760. doi: 10.1186/1471-2164-15-760
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). RNA-seq experiments with HISAT StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016-095
- Petráček, J., Mravec, J., Bouchard, R., Blakeslee, J. J., Abas, M., Seifertová, D., et al. (2006). PIN proteins perform a rate-limiting function in cellular auxin efflux. *Science* 312, 914–918. doi: 10.1126/science.1123542
- Prakash, P., Srivastava, R., Prasad, P., Kumar Tiwari, V., Kumar, A., Pandey, S., et al. (2020). Trajectories of cotton fiber initiation: a regulatory perspective. *Preprints* doi: 10.20944/preprints202011.0060.v1
- Qi, H., Jiang, Z., Zhang, K., Yang, S., He, F., and Zhang, Z. (2018). PlaD: a Transcriptomics database for plant defense responses to pathogens, providing new insights into plant immune system. *Genomics Proteomics Bioinform.* 16, 283–293. doi: 10.1016/j.gpb.2018.08.002
- Qin, Y., Sun, H., Hao, P., Wang, H., Wang, C., Ma, L., et al. (2019). Transcriptome analysis reveals differences in the mechanisms of fiber initiation and elongation between long- and short-fiber cotton (*Gossypium hirsutum* L.) lines. *BMC Genomics* 20:633. doi: 10.1186/s12864-019-5986-5
- Qin, Y. M., Hu, C. Y., Pang, Y., Kastaniotis, A. J., Hiltunen, J. K., and Zhu, Y. X. (2007a). Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. *Plant Cell* 19, 3692–3704. doi: 10.1105/tpc.107.054437
- Qin, Y. M., Pujol, F. M., Hu, C. Y., Feng, J. X., Kastaniotis, A. J., Hiltunen, J. K., et al. (2007b). Genetic and biochemical studies in yeast reveal that the cotton fibre-specific GhCER6 gene functions in fatty acid elongation. *J. Exp. Bot.* 58, 473–481. doi: 10.1093/jxb/erl218
- Qin, Y. M., and Zhu, Y. X. (2011). How cotton fibers elongate: a tale of linear cell-growth mode. *Curr. Opin. Plant Biol.* 14, 106–111. doi: 10.1016/j.pbi.2010.09.010
- R Core Team (2013). *The R Project for Statistical Computing*, Vol. 3. Vienna: R Core Team, 1–12.
- Rai, K. M., Singh, S. K., Bhardwaj, A., Kumar, V., Lakhwani, D., Srivastava, A., et al. (2013). Large-scale resource development in *Gossypium hirsutum* L. by 454 sequencing of genic-enriched libraries from six diverse genotypes. *Plant Biotechnol. J.* 11, 953–963. doi: 10.1111/pbi.12088
- Rosloski, S. M., Singh, A., Jali, S. S., Balasubramanian, S., Weigel, D., and Grbic, V. (2013). Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of *Arabidopsis thaliana*. *Plant Mol. Biol.* 11, 953–963. doi: 10.1007/s11103-012-9982-2
- Rung, J., and Brazma, A. (2013). Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.* 14, 89–99. doi: 10.1038/nrg3394
- Salih, H., Leng, X., He, S. P., Jia, Y. H., Gong, W. F., and Du, X. M. (2016). Characterization of the early fiber development gene, Ligon-lintless 1 (L1), using microarray. *Plant Gene* 6, 59–66. doi: 10.1016/j.plgene.2016.03.006
- Samuel Yang, S., Cheung, F., Lee, J. J., Ha, M., Wei, N. E., Sze, S. H., et al. (2006). Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J.* 47, 761–775. doi: 10.1111/j.1365-313X.2006.02829.x
- Schmittgen, T. D., and Livak, K. J. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-Delta Delta C(T)} method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Schmitz, R. J., Grotewold, E., and Stam, M. (2021). Cis-regulatory sequences in plants: their importance, discovery, and future challenges. *Plant Cell* 34, 718–741. doi: 10.1093/plcell/koab281
- Shang, H., Wang, Z., Zou, C., Zhang, Z., Li, W., Li, J., et al. (2016). Comprehensive analysis of NAC transcription factors in diploid *Gossypium*: sequence conservation and expression analysis uncover their roles during fiber development. *Sci. China Life Sci.* 59, 142–153. doi: 10.1007/s11427-016-5001-1
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (1971). Cytoscape: a software environment for integrated models. *Genome Res.* 13:426. doi: 10.1101/gr.1239303.metabolite
- Shi, Y. H., Zhu, S. W., Mao, X. Z., Feng, J. X., Qin, Y. M., Zhang, L., et al. (2006). Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18, 651–664. doi: 10.1105/tpc.105.040303
- Song, Q., Guan, X., and Chen, Z. J. (2015). Dynamic roles for small RNAs and DNA methylation during ovule and fiber development in allotetraploid Cotton. *PLoS Genet.* 11:e1005724. doi: 10.1371/journal.pgen.1005724
- Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18:99. doi: 10.1186/s13059-017-1229-8
- Srivastava, A. K., Lu, Y., Zinta, G., Lang, Z., and Zhu, J. K. (2018). UTR-Dependent control of gene expression in plants. *Trends Plant Sci.* 23, 248–259. doi: 10.1016/j.tplants.2017.11.003
- Sun, Q., Huang, J., Guo, Y., Yang, M., Guo, Y., Li, J., et al. (2020). A cotton NAC domain transcription factor, GhFSN5, negatively regulates secondary cell wall biosynthesis and anther development in transgenic *Arabidopsis*. *Plant Physiol. Biochem.* 146, 303–314. doi: 10.1016/j.plaphy.2019.11.030
- Taliercio, E. W., and Boykin, D. (2007). Analysis of gene expression in cotton fiber initials. *BMC Plant Biol.* 7:22. doi: 10.1186/1471-2229-7-22
- Tian, Y., and Zhang, T. (2021). MIXTAs and phytohormones orchestrate cotton fiber development. *Curr. Opin. Plant Biol.* 59:101975. doi: 10.1016/j.pbi.2020.10.007
- Udall, J. A., Flagel, L. E., Cheung, F., Woodward, A. W., Hovav, R., Rapp, R. A., et al. (2007). Spotted cotton oligonucleotide microarrays for gene expression analysis. *BMC Genomics* 8:81. doi: 10.1186/1471-2164-8-81
- Udall, J. A., Swanson, J. M., Haller, K., Rapp, R. A., Sparks, M. E., Hatfield, J., et al. (2006). A global assembly of cotton ESTs. *Genome Res.* 16, 441–450. doi: 10.1101/gr.4602906
- Waggott, D., Chu, K., Yin, S., Wouters, B. G., Liu, F. F., and Boutros, P. C. (2012). NanoStringNorm: an extensible R package for the pre-processing of nanostring mRNA and miRNA data. *Bioinformatics* 28, 1546–1548. doi: 10.1093/bioinformatics/bts188
- Walford, S. A., Wu, Y., Llewellyn, D. J., and Dennis, E. S. (2011). GhMYB25-like: a key factor in early cotton fibre development. *Plant J.* 65, 785–797. doi: 10.1111/j.1365-313X.2010.04464.x
- Walford, S. A., Wu, Y., Llewellyn, D. J., and Dennis, E. S. (2012). Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. *Plant J.* 71, 464–478. doi: 10.1111/j.1365-313X.2012.05003.x
- Wang, L., Kartika, D., and Ruan, Y. L. (2021). Looking into 'hair tonics' for cotton fiber initiation. *New Phytol.* 229, 1844–1851. doi: 10.1111/nph.16898
- Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., et al. (2019). Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* 51, 224–229. doi: 10.1038/s41588-018-0282-x

- Wang, N.-N., Li, Y., Chen, Y.-H., Lu, R., Zhou, L., Wang, Y., et al. (2021). Phosphorylation of WRKY16 by MPK3-1 is essential for its transcriptional activity during fiber initiation and elongation in cotton (*Gossypium hirsutum*). *Plant Cell* 33, 2736–2752. doi: 10.1093/plcell/koab153
- Wang, Q. Q., Liu, F., Chen, X. S., Ma, X. J., Zeng, H. Q., and Yang, Z. M. (2010). Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. *Genomics* 96, 369–376. doi: 10.1016/j.ygeno.2010.08.009
- Wang, X. C., Li, Q., Jin, X., Xiao, G. H., Liu, G. J., Liu, N. J., et al. (2015). Quantitative proteomics and transcriptomics reveal key metabolic processes associated with cotton fiber initiation. *J. Proteomics* 114, 16–27. doi: 10.1016/j.jprot.2014.10.022
- Wilkins, T. A., and Arpat, A. B. (2005). The cotton fiber transcriptome. *Physiol. Plant* 124, 295–300. doi: 10.1111/j.1399-3054.2005.00514.x
- Wu, H., Ren, Z., Zheng, L., Guo, M., Yang, J., Hou, L., et al. (2021). The bHLH transcription factor GhPAS1 mediates BR signaling to regulate plant development and architecture in cotton. *Crop J.* 9, 1049–1059. doi: 10.1016/j.cj.2020.10.014
- Xu, J., Yang, X., Li, B., Chen, L., Min, L., and Zhang, X. (2019). GhL1L1 affects cell fate specification by regulating GhPIN1-mediated auxin distribution. *Plant Biotechnol. J.* 17, 63–74. doi: 10.1111/pbi.12947
- Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., et al. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* 10:2989. doi: 10.1038/s41467-019-10820-x
- Yang, Z., Qanmber, G., Wang, Z., Yang, Z., and Li, F. (2020). *Gossypium* genomics: trends, scope, and utilization for cotton improvement. *Trends Plant Sci.* 25, 488–500. doi: 10.1016/j.tplants.2019.12.011
- Yuan, D., Grover, C. E., Hu, G., Pan, M., Miller, E. R., Conover, J. L., et al. (2021). Parallel and intertwining threads of domestication in allopolyploid cotton. *Adv. Sci.* 8:2003634. doi: 10.1002/adv.202003634
- Zeng, J., Zhang, M., Hou, L., Bai, W., Yan, X., Hou, N., et al. (2019). Cytokinin inhibits cotton fiber initiation by disrupting PIN3a-mediated asymmetric accumulation of auxin in the ovule epidermis. *J. Exp. Bot.* 70, 3139–3151. doi: 10.1093/jxb/erz162
- Zhang, F., Jin, X., Wang, L., Li, S., Wu, S., Cheng, C., et al. (2016). A cotton annexin affects fiber elongation and secondary cell wall biosynthesis associated with Ca²⁺ influx, ROS homeostasis, and actin filament reorganization. *Plant Physiol.* 171, 1750–1770. doi: 10.1104/pp.16.00597
- Zhang, J., Huang, G. Q., Zou, D., Yan, J. Q., Li, Y., Hu, S., et al. (2018). The cotton (*Gossypium hirsutum*) NAC transcription factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. *New Phytol.* 217, 625–640. doi: 10.1111/nph.14864
- Zhang, M., Zeng, J. Y., Long, H., Xiao, Y. H., Yan, X. Y., and Pei, Y. (2017). Auxin regulates cotton fiber initiation via GHPIN-mediated auxin transport. *Plant Cell Physiol.* 58, 385–397. doi: 10.1093/pcp/pcw203
- Zhang, M., Zheng, X., Song, S., Zeng, Q., Hou, L., Li, D., et al. (2011). Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat. Biotechnol.* 29, 453–458. doi: 10.1038/nbt.1843
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. Acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207
- Zhao, B., Cao, J. F., Hu, G. J., Chen, Z. W., Wang, L. Y., Shangguan, X. X., et al. (2018). Core cis-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. *New Phytol.* 218, 1061–1075. doi: 10.1111/nph.15063
- Zou, C., Wang, Q., Lu, C., Yang, W., Zhang, Y., Cheng, H., et al. (2016). Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). *Sci. China Life Sci.* 59, 164–171. doi: 10.1007/s11427-016-5000-2

Conflict of Interest: The CSIR-NBRI has an official collaboration with Tierra Seed Science Pvt. Ltd., for cotton research where DM and PB are employed.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Prasad, Khatoon, Verma, Aalam, Kumar, Mohapatra, Bhattacharya, Bag and Sawant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.