4.0 RESULTS

4.1 Molecular cloning and characterization of GhSPL5D gene in cotton

We previously reported global expression profiling of five contrasting genotypes (JKC725, JKC777, JKC703, JKC737, JKC783) of *G. hirsutum* at their six developmental stages (0, 6, 9, 12, 19 and 25 DPA) (Nigam et al., 2013). These five contrasting genotypes includes two superior (JKC725 and JKC777) and three inferior (JKC703, JKC737 and JKC783) genotypes based on their fiber quality parameters. We reported 37 potential transcription factors expressing contrastingly among superior genotypes compared to inferior and one of such transcription factor family was SPL. There were nine SPL gene family members present on cotton affymetrix chip expressing differentially in all five contrasting genotypes from 0 to 25DPA stages (Figure 4.1).

![Figure 4.1 - Hierarchical clustering of G.hirsutum SPL genes present on cotton affymetrix chip](image)

The nine SPL genes expressed differentially at 6 developmental time point in five (two superior and three inferior) contrasting genotypes of *G. hirsutum*. Color scale represents log2 transformed expression values. Green shows that the expression levels of GhSPLs are low; and red shows that the expression levels are high.

The heat map expression profile of SPL genes showed three distinct expression clusters. In cluster I, the expression of SPL genes (Ghi.8127.1.S1_s_at (SPL5) and GhiAffx.17946.1.A1_at (SPL3)) was higher than cluster II and III. Cluster II had two
Results

*SPL* genes (Gra.1588.1.A1_s_at (*SPL6*) and Ghi.3540.1.A1_at (*SPL12*)) having lower expression compared to cluster I and higher expression compared to cluster III. Cluster III contained four *SPL* genes (Gra.1300.1.A1_at (*SPL14*), GraAffx.24703.1.A1_s_at (*SPL9*), Gra.36.1.A1_at (*SPL2*), Gra.1201.1.A1_at (*SPL1*) and GraAffx.19401.1.A1_at (*SPL13*) with having least expression. The expression of two Cluster I *SPL* genes viz., *SPL5* and *SPL3* was predominantly higher in all genotypes but showed specificity towards developing fiber stages. We selected total of six genes viz *SPL1*, *SPL3*, *SPL9*, *SPL11*, *SPL13* and *SPL14* for validation of their expression using RT-PCR in all six selected fiber developmental stages and also selected leaf, root and bud as other tissue to check their fiber specificity (Figure 4.2).

![Figure 4.2 - Analyses of six GhSPLs gene expression in various tissues of cotton.](image)

Real-time PCR analysis of (A) *GhSPL1* (B) *GhSPL3* (C) *GhSPL9* (D) *GhSPL5* (E) *GhSPL13* and (F) *GhSPL14* expression was conducted in root, leaf, bud and 0, 6, 9, 12, 19, and 25DPA fibers. The tissue at 0 DPA was complete ovule and the others were fiber samples.

The RT-PCR revealed that *SPL3* and *SPL5* showed true fiber specific expression, the expression of *SPL5* was significantly high in 0 to 6 DPA and then gradually decline and abolishes at 25 DPA (Figure 4.2 B&D), thus validating microarray result. Further,
the expression of SPL3 was significantly higher at 6 and 9 DPA and gradually decline and abolishes at 25 DPA. RT-PCR results thus indicate that fiber initiation specificity of SPL5 and fiber elongation specificity of SPL3. In the present study we decided to explore the role of GhSPL5D in fiber initiation.

Further, we were interested to validate the expression of GhSPL5D in superior and inferior genotypes of cotton. Importantly, the expression of GhSPL5D was upregulated in the fiber of superior (JKC725) as compared to the inferior genotype (JKC703) throughout all (0 to 25DPA) fiber developmental stages (Figure 4.3).

![Figure 4.3 - The GhSPL5D expression level in superior and inferior genotype.](image)

The expression analysis of GhSPL5D in fibers of superior (JKC725) and inferior (JKC703) genotypes at 0, 6, 9, 12, 19 and 25DPA developmental stages.

The higher expression of GhSPL5D in superior genotype suggests that it may play role in cotton fiber quality.

### 4.2 Cloning and characterization of GhSPL5D promoter in cotton

To further validate the initiation specific expression of GhSPL5D, the 663 bp promoter region of GhSPL5D gene was cloned by genome walking and was fused upstream of GUSA reporter gene in a plant expression vector. Several cotton transgenic lines expressing GUSA was developed and bolls from the transgenic lines were subjected to histochemical staining. The histochemical staining of bolls from transgenic lines expressing pGhSPL5D::gusa construct revealed that the GUS
expression was significantly higher during initiation (0DPA) but subsequently in the elongation (6-12DPA) stage it was declined and eventually abolishes at 25 DPA stage (Figure 4.4-A). In non-transgenic control, GUS activity was not detected in any of the developmental stage of cotton ball. Further, fluorimetric assay of GUS in the bolls of independent transgenic lines also revealed that the expression of pGhSPL5D::GUSA was higher during the initiation stages (0-6 DPA) and declined subsequently in later stages (Figure 4.4-B). Thus expression analysis of *GhSPL5D* promoter confirmed its fiber specific expression.

**Figure 4.4** – The *GhSPL5D* promoter analysis. (A) GUS staining driven by *GhSPL5D* promoter in transgenic cotton plants: a, c, e, g, i and k are cotton boll with ovule and fiber at 0, 6, 9, 12, 19 and 25DPA, respectively, and b, d, f, h, j, l are control of the same stages. (B) Quantitative assay of GUS activity in 0, 6, 9, 12, 19 and 25DPA fibers.
4.3 Identification and phylogenetic analysis of SPL genes in cotton

To identify the SPL genes in cotton, G. raimondii and G. arboreum genome database (http://www.cottongen.org/species/Gossypium_raimondii/bgi575cgp_genome_v1.0) with a profile Hidden Markov Model (pHMM) of the SBP domain was searched. We identified 29 putative SPL genes in G. raimondii and G. arboreum genome. These 29 SPL genes were subjected to protein BLAST. The BLAST result showed that most of the SPLs were truncated part of the same SPL gene.

![Exon/Intron structure and phylogenetic analysis of SPLs.](image)

Figure 4.5 - The Exon/Intron structure and phylogenetic analysis of SPLs. Exons and introns structures of A and D genome SPLs are indicated with the black boxes and brown lines, respectively. The number indicates intron phase.

After removal of truncated SPLs, we found 13 SPL members in G. raimondii and 12 in G. arboreum and they were named according to their homology with Arabidopsis SPLs. Their ORF was predicted by ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and exon/intron structure was predicted by Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/chinese.php) (Figure 4.5).
The genomic structure of these SPLs revealed that GrSPL3, GrSPL4 and GrSPL5 had 1 intron and 2 exon whereas SPL1, SPL7, SPL12 and SPL14 had 10 exon and 9 intron. The G. raimondii, SPL gene sequences were used as query in BLAST programme to search their homologs in G. hirsutum. We were able to identify a total 13 SPL gene family members in G. hirsutum and their ORF was predicted by ORF finder. All SPL genes have a highly conserved DNA binding domain known as SBP domain (Figure 4.6). The SBP domain is consist of four motifs; zinc finger1 (Zn1), zinc finger2 (Zn2), joint peptide which connects Zn1 & Zn2 and a nuclear localization signal (NLS). Zn1 (CX₄CX₁₀₋₁₁HXC) and Zn2 (CX₂CX₁₃CX₆H) are CCHC and CCCH-type of zinc finger motifs present in SBP domain. Joint peptide is not highly conserved and play role in the protein-DNA interaction (Yamasaki et al., 2004).

**Figure 4.6 - The SBP domain LOGO view.** The SBP domain sequence LOGO view of SPLs. The amino acid height depicts the level of conservation. The motifs of the SBP domain are shown by rectangles.

The SBP domain protein sequences were used to study the evolutionary relationship of SPL genes in different plant species. In a recent report, 128 SBP domain proteins from nine plant species (green alga, moss, Arabidopsis, Vitis spp., Solanum spp., A. majus, poplar, rice and maize) were identified (Hou et al., 2013). We identified 26 SBP domain proteins (13 from each G. raimondii, G. arboreum and G. hirsutum) in cotton. Thus a total 167 SBP domain protein sequences from 12 plant species were used to construct a phylogenetic tree using the neighbor-joining algorithm. On the basis of un-rooted phylogenetic tree, the 167 plant SPLs were classified into 12 subgroups (Figure 4.7). The result indicates that SPL gene family is evolutionary diverse and cotton SPLs exhibited closer relationship with dicot angiosperms as compare to monocot, algae and moss. The G. hirsutum SPL5 (GhSPL5D) grouped in
subgroup 2 along with *G. raimondii* (GrSPL5), *Arabidopsis* (AtSPL5) and *Populus* (PtSPL5).

**Figure 4.7 - Un-rooted phylogenetic tree of various plant species SPLs.** A total 167 SBP domain protein sequences from 12 species were used to construct this unrooted-phylogenetic tree using the neighbor-joining algorithm. The 12 groups of SBP domains are denoted by red bars. SBP domains taken from different plant species are denoted by circles of different colors.
Results

These results suggest that the plant SPL gene family may be evolved from a common ancestor and many of them may have undergone further differentiation separately in monocot and dicot plants.

4.4 Genomic organization of GhSPL5D gene

The genomic structure of GhSPL5D gene showed that it has 2exon, 1 intron and a miR156 target site (Figure 4.8). The miR156 target site is present in the 3’ UTR, 26 nucleotides away from the stop codon. The exon 1 of 302 bp long and exon 2 of 247bp long are linked together by a 144 bp long intron.

![Figure 4.8 - Genomic organization of GhSPL5D gene.](image)

The putative ORF of 549 bp was cloned which encode a predicted polypeptide of 182 amino acid (Figure 4.9). The deduced polypeptide sequence has a highly conserved SBP domain composed of a 76 amino acids which is involved in DNA binding. Within the SBP domain, 17 amino acid differences were found in between G. hirsutum (Gh) SPL5, G. raimondii (Gr) SPL5, G. arboreum (Ga) SPL5 and A. thaliana (At) SPL5 proteins. However, the amino acid sequences outside the SBP domain varies greatly.
### Results

The *GhSPL5D* gene was classified into two groups on the basis of their evolution from A genome and D genome which were named *GhSPL5A* and *GhSPL5D*, respectively. Between *GhSPL5A* and *GhSPL5D* coding sequences six SNPs were found (Figure 4.10 A-D).

#### Figure 4.9 - The polypeptide sequence analysis of SPLs.**
Alignment of the amino acid sequences of *GhSPL5, GrSPL5, GaSPL5* and *AtSPL5*. Underscore (●) points to the different amino acid in the conserved SBP domain.

#### Figure 4.10 - Sequence analysis of *GhSPL5D* mRNA with A genome, D genome, superior and inferior genotypes. (A) Differences in *GhSPL5D* CDS sequences of *SPL5A* (A genome) and *SPL5D* (D genome). (B, C) Comparison of *SPL5A* and *SPL5D* sequences with superior and inferior genotypes *GhSPL5D* CDS sequences. (D) Alignment of 5’UTR of *SPL5A* and *SPL5D* mRNA with *SPL5AADD*.
Further, we analysed the *GhSPL5A* and *GhSPL5D* coding sequence with superior and inferior *GhSPL5* sequence. We found nine SNP variations in each *SPL5A* and *SPL5D* genome sequences. The 5’UTR of *SPL5A* was highly diverged while in *SPL5D* it was highly conserved with *GhSPL5D*.

### 4.5 The expression of miR156 and *GhSPLs* are inversely related during cotton fiber development

Previous studies, in *Arabidopsis*, have reported that the expression of *SPLs* was inversely related to miR156 during vegetative to reproductive phase transition (Wu and Poethig, 2006). The *SPL* genes positively regulate the expression of miR172, whose level increases during shoot development (Jung et al., 2007). To address this, we analysed the expression levels of miR156 and miR172 in the cotton fiber at different developmental stages. We isolated RNA from 0, 6, 9, 12, 19 and 25 DPA cotton fibers and performed small-RNA northern blotting of miR156, miR172 and U6 as control. The expression analysis of miR156 revealed that its expression was slightly higher at 0DPA but it subsequently increases till 25DPA (Figure 4.11A). Further, in complete contrast, the expression of miR172 showed highest expression at 0 DPA but it subsequently declined till 19DPA (Figure 4.11B).

![Figure 4.11](image_url)

**Figure 4.11 - Expression analyses of miR156 and miR172.** (A) Small RNA blot analysis of GhmiR156 and (B) GhmiR172 expression at 0, 6, 9, 12,19 and 25 DPA stages of fiber development; U6 was used as a loading control.
Thus the expression of miR156 showed completely inverse expression profiling as compared to expression of \emph{GhSPL5D} (Figure 4.11A & 4.2 D) while expression of miRNA 172 was directly related to the expression of \emph{GhSPL5D} (Figure 4.11B & 4.2D). Thus, our results are agrees to the previous results obtained in case of \emph{Arabidopsis} (Wu et al., 2009).

\subsection*{4.6 miR156 regulate expression of \textit{GhSPL5D} by targeting its 3’ UTR}

To identify the miR156 target sites in \textit{GhSPL} transcripts, we computationally searched the \textit{SPL} mRNA sequences in miRbase target search tool (http://www.mirbase.org/). The result showed that the \textit{GhSPLs} produces transcripts that are targeted by miR156. The 7 \textit{SPL} members (\textit{GhSPL2, GhSPL3, GhSPL5, GhSPL6, GhSPL9, GhSPL10} and \textit{GhSPL13}) out of 13 were identified in \textit{G. hirsutum} that contains the target site for miR156 (Figure 4.12).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mir156_target.png}
\caption{The SPLs transcript complementary to miR156 taget site. \textit{GhSPLs} transcript bearing the target site of miR156 shown in red color; black rectangle shows SBP domain. The arrow shows cotton miR156 members target sequences in \textit{G. hirsutum SPLs} transcript.}
\end{figure}
We were next interested to investigate whether miR156 indeed target *GhSPL5D* and if yes then map its targeted cleavage site. To evaluate, we performed uncapped 5′RNA ligase mediated rapid amplification of cDNA ends (5′RLM-RACE).

![Diagram of miR156 cleavage site](image)

**Figure 4.13 – The in-vivo identification of miR156 degradation site in *GhSPL5D* transcript.** Mapping of miR156 cleavage site in *GhSPL5D* transcript identified by 5′RLM-RACE.

The cleaved products were amplified, cloned and sequenced. We sequenced 18 clones and out of 18 clones 14 clones showed cleavage site present in the middle of miR156 target site. A total 4 clones also showed cleavage site present outside of miR156 target site which suggest any other type of cleavage of *GhSPL5D* transcript. The result showed that *GhSPL5D* is cleaved by GhmiR156 in the middle of the target site (Figure 4.13). This result confirmed that miR156 indeed regulate the expression of *GhSPL5D* by cleaving it during the fiber development.

### 4.7 Semi-quantitative PCR analysis to measure *GhSPL5D* cleaved product

Next to estimate degradation of *GhSPL5D* transcript by miRNA156, we performed semi-quantitative 5′ RACE RT-PCR. The cDNA was synthesized from RNA isolated at five (0, 6, 9, 12 and 19DPA) developmental stages of cotton fiber. The result showed gradual increase in the level of degraded product from 0 to 19 DPA stages (Figure 4.14).
Results

Figure 4.14 - Measurement of GhSPL5D cleaved product level. Semi-quantitative 5’ RACE-PCR at 0, 6, 9, 12, and 19 DPA stages of fiber development to detect the 3’ fragments produced by miR156-guided cleavage of GhSPL5D transcripts. UBQ10 was used as a control.

Significant difference in the abundance of 3’ cleavage product was detected at 0 and 6 DPA stages. Interestingly, the level of 3’ cleaved products of GhSPL5D transcript was positively correlated with the expression level of miR156. Thus results confirmed the regulation of GhSPL5D by miRNA 156 during cotton fiber development.

4.8 GhSPL5D controls cotton fiber initiation and boll biomass

To investigate the biological role of cotton GhSPL5D gene, we cloned 549 bp ORF of GhSPL5D (GenBank accession number Ghi.8127.1.S1.s_at) isolated from JKC725 line in front of CaMV 35S promoter in pCAMBIA1301 plant transformation vector to obtained over-expression (O/E) construct. Similarly, 281 bp of GhSPL5D fragment was used to develop RNAi construct in PFGC1008 RNAi plant expression vector to knockdown (KD) GhSPL5D gene expression. To make GhSPL5D specific silencing the conserved SBP domain sequence was excluded. The GhSPL5D O/E and KD constructs were mobilized into cotton (G. hirsutum cv. Coker-310) via Agrobacterium mediated transformation. A total of 22 GhSPL5D O/E and 20 KD transgenic cotton lines were developed. Initial verification of the transgenics was done by performing PCR using HPTII gene specific primer (Figure 4.15A-D).
**Results**

**Figure 4.15** - Methods to screen the *GhSPL5D* cotton transgenic plants. Strategies of *GhSPL5D* O/E and knockdown (KD) transgenic analysis (A& C) and PCR screening for transgene confirmation at T1 generation plants derived from five T0 lines (B&D).

The transgenics were further confirmed by performing real time PCR. The qRT-PCR analysis of T1-transgenic cotton lines showed that expression level of *GhSPL5D* in 6DPA fibers of transgenic line expressing O/E construct was significantly higher than that of untransformed control (Figure 4.16A), whereas it’s expression was repressed in KD transgenic lines (Figure 4.16B).

**Figure 4.16** - Expression analysis of *GhSPL5D*. The qRT-PCR analysis of *GhSPL5D* expression in O/E (A) and KD (B) T1 generation transgenic and control plants. RNA was isolated from 6 DPA fibers.

**4.9 Scanning electron microscopy analysis of *GhSPL5D* O/E and KD lines ovules**

The ovules at 1 DPA from O/E line (*GhSPL5D O/E_8*), KD line (*GhSPL5D KD_10*) and control were subjected to scanning electron microscopy (SEM). The result showed that the number of fiber initials were significantly higher in O/E as compared...
Results

to untransformed control while that in KD lines was significantly lower. Further, the fiber initials were elongated at 1DPA in O/E lines as compared to both control and KD lines (Figure 4.17A). Further, to investigate exact developmental stage when these differences become apparent, we evaluated ovules of 0 DPA from control, O/E and KD lines using SEM analysis. The bulging of fiber initials in overexpressed ovules was higher and synchronized; whereas it was significantly retarded in KD ovules as compared to the control (Figure 4.17B). The 0DPA stage is a key stage for fiber cell fate determination and this result showed that the regulation of fiber cell initiation by *GhSPL5D* occurred at this stage.

**Figure 4.17 - Morphology of fiber initials in O/E, KD and control plants.** (A) Fiber cells bulging out from control (a, b, c), *GhSPL5D* O/E_8 (d,e,f) and *GhSPL5D* KD_10 (g, h, i) ovules at 0DPA and 1DPA (j, k, l, m, n, o and p, q, r) stages. *GhSPL5D* O/E_8 ovule showed higher fiber initials and elongation whereas *GhSPL5D* KD_10 ovule showed sunken fiber cells with reduced fiber initials. The fiber cells were viewed at 80X (a, d, g, j, m and p), 600X (b, e, h, k, n and q), and 1200X (c, f, i, l, o and r) magnification. (B) Fibre initial cell numbers were counted on SEM images taken from the middle of ovules at 1 dpa of 3600 mm² area and averaged of three randomly selected ovules.

The number of fiber initials was counted on scanning electron microscopy images at several grids of the selected 1 DPA ovules. The result showed that the number of fiber initials in O/E ovules was significantly higher as compared to the control and KD ovules both (Figure 4.17B). Similarly, the number of fiber initials in KD ovules was significantly lower as compared with the control and O/E ovules.

**4.10 Measurement of cotton boll biomass and fiber quality of *GhSPL5D* O/E and KD lines**

Next, we evaluated fully matured bolls of control, O/E and KD lines. The O/E transgenic lines showed bigger boll size (4.36 g) compared with control (3.37 g), in
Results

contrast, boll size of KD lines (3.25 g) was slightly smaller as compared to control (Figure 4.18A-B). Interestingly, GhSPL5D O/E transgenic lines had higher number of boll (5.6) compared to control (3.37) and KD lines (2.73) (Figure 4.18C). Further, significant differences were also observed among O/E, control and KD transgenic plants for seed cotton weight per plant. The maximum seed cotton weight per plant was obtained from O/E lines (16.92 g) whereas control lines produced only 13.25 g and the least seed cotton weight per plant (9.57g) was produced by KD lines (Figure 4.18D). Similarly, cotton lint yield per plant of the O/E lines and control was 5.67g and 4.17 g whereas KD lines produced only 3.42g lint per plant (Figure 4.18E). The lint percentage [(weight of lint fibers/weight of seed cotton) × 100] in GhSPL5D O/E lines was higher (33.50%) than control (31.40%), resulting in a 35.97% increase in lint yield. We also observed significant differences among O/E, KD and control plants for hundred seed weight.

Figure 4.18 - Analysis of cotton yield parameters. (A) The 20 DPA old cotton boll size phenotype of independent T1 transgenic and wild type plants. (B) Boll size value in transgenic lines compared with the wild type; the value is average weight of 40 mature bolls ± SD. (C) Boll number in transgenic lines compared with the wild type; the value is average weight of 54 mature bolls ± SD (D) Seed cotton weight in transgenic lines compared with the wild type; the value is average weight of 14 plants of each lines ± SD. (E) Lint cotton weight without seed in transgenic lines used to quantify the seed cotton weight (F) Effect of altered expression of GhSPL5D gene on 100-seed weight; ±SD was calculated from three replicates of per 100 seeds weight in transgenic line and control (n=14); *P-value ≤0.05 ; ** P-value ≤0.01 (Student’s t test).
Results

Hundred seed weight values of *GhSPL5D* O/E and KD lines obtained was 8g and 7.06g whereas control 100-seed weight value was 7.08g (Figure 4.18F).

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Span Length (2.5%:mm)</th>
<th>Strength (3.2mm; g/Tex)</th>
<th>Micronaire</th>
<th>UR (%)</th>
<th>SFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>O/E</td>
<td>30.6-31.0</td>
<td>24.5-24.9</td>
<td>2.5-2.7</td>
<td>50-51</td>
<td>5.3-6.2</td>
</tr>
<tr>
<td>Control</td>
<td>30.4-30.8</td>
<td>22.6-23.4</td>
<td>2.3-2.5</td>
<td>51-52</td>
<td>5.3-6.2</td>
</tr>
<tr>
<td>KD</td>
<td>30.0-30.3</td>
<td>22.0-22.3</td>
<td>2.5-2.6</td>
<td>51-52</td>
<td>4.0-4.6</td>
</tr>
</tbody>
</table>

About 20g fiber sample from *GhSPL5D* O/E, KD and control plants were used for the analysis of fiber quality parameters. The result showed that the fiber quality parameters like fiber length, fiber strength, micronaire, uniformity ratio and short fiber index remains largely unaltered in *GhSPL5D* O/E and KD lines as compared to the control (Table 4.1). These results indicate that the *GhSPL5D* transcription factor controls overall fiber initiation and biomass of the boll without affecting the fiber quality parameters in cotton.

4.11 Cotton transgenic plants overexpressing *GhmiR156* and target mimicry (*MIM156*) validate interaction of *GhSPL5D* and *miR156*

The GhmiR156 precursor (Pre-GhmiR156) having a 20 nucleotide long mature sequence was used to overexpress mature miR156 in cotton. The Pre-GhmiR156 sequence was amplified from *G. hirsutum* genomic DNA and cloned downstream of CaMV35S promoter, in pCambia 1301 plant expression vector. Similarly, target mimicry *MIM156* vector (Franco-Zorrilla et al., 2007) was constructed by insertion of seven tandemly arrayed miRNA156 binding site into pEarleyGate 103 (ABRC stock number CD3-685) plant expression vector.
Results

The 156O/E and MIM156 constructs were mobilized into cotton (G. hirsutum cv. Coker-310) via agrobacterium mediated transformation. A total 15 156O/E and 12 MIM156 transformants were generated at T1 generation. Based on the molecular analysis, total of ten MIM156 and six 156O/E lines at T1 generation were used for detailed analysis. Initially, the miR156 O/E lines were verified by performing PCR with HPTII gene specific primer and the MIM156 plants were by BASTA gene specific primer (Figure 4.19A-D).

Further, the expression of miR156 in 156O/E transgenic lines and control plants was measured by small RNA blot. As expected, higher abundance of the miR156 level was detected in 156O/E lines as compared to the control plants (Figure 4.20A). We next investigated the expression level of GhSPL5D gene in 156O/E and MIM156 transgenic plants. As expected, the GhSPL5D transcript level was higher in MIM156 and lower in 156O/E transgenic plants as compared to control (Figure 4.20B). This result further confirmed regulation of GhSPL5D by miR156.

Figure 4.19 - Methods to screen the miR156 cotton transgenic plants. Strategies of miR156O/E and MIM156 transgenic analysis (A-C) and PCR screening for transgene confirmation at T1 generation plants derived from 4 & 5 T0 lines (B-D).
Figure 4.20 - Expression analysis of miR156 and GhSPL5D. (A) Small RNA blot analysis of GhmiR156 expression in miR156 overexpression (O/E) transgenic and control cotton plants. (B) Relative expression levels of GhSPL5D mRNA in control and T1 transgenic plants which were determined by qRT-PCR. From left to right: O/E miR156 1, 2 and 5, control; 156MIM 2, 5 and 6 transgenic lines ± SD.

4.12 Measurement of boll biomass and fiber quality of GhmiR156 O/E and MIM156 lines

To study the impact of miR156 expression on plant development detailed phenotypic evaluation of miR156 O/E and MIM156 lines was performed. Interestingly, we observed 156O/E transgenic lines showed shorter plant height (75cm) compared to control (91.83cm), in contrast, plant height of MIM156 lines (102cm) was larger compared to control (Figure 4.21A, B). The 156 O/E transgenic lines had lower boll number (4.5) compared to control (7.0) while we do not see any change in boll number MIM156 (7.0) lines (Figure 4.21C). The expression of GhSPL5D was significantly higher in case of MIM156 transgenic lines, thus as expected we also observed bigger boll size (4.073g) as compared to control (2.702 g) (Figure 4.21D). The expression of GhSPL5D was suppressed in case of 156 O/E transgenic lines but in complete contrast we observed higher boll size (4.263g) as compare to control (Figure 4.21D).
Results

Figure 4.21 - Phenotype and boll density in 156O/E and MIM156 lines. (A) Phenotype of 156O/E, control and MIM156 plants (B) Plant height value of transgenic and control plants (C) Boll number in transgenic lines compared with the wild type; the value is average of 10 MIM156 and 6 miR156 O/E plants ± SD. (D) Boll weight value in transgenic lines compared with the wild type; the value is average weight of mature bolls ± SD.

The flowering time was delayed in miR156 O/E lines compared to control; while in complete contrast, flowering of MIM156 lines was slightly earlier compared to control (Figure 4.22A). The maximum seed cotton weight per plant was obtained from MIM156 lines (16.85 g); whereas control and 156 O/E lines produced only 11.66 g and 11.16g (Figure 4.22B). Similarly, the higher amount of lint per plant was obtained from MIM156 lines (7.1g) whereas control and 156O/E lines produced only 5.0 g and 4.75g (Figure 4.22C). We observed no significant differences among MIM156, 156 O/E and control plants for hundred seed weight. Hundred seed weight values of MIM156 and 156O/E lines obtained was 9.166g and 10.083g whereas wild type 100-seed weight value was 9.166g (Figure 4.22D). We also studied fiber quality parameters of lint cotton obtained after ginning from 156 O/E, MIM156 and control plants.
Figure 4.22 - Flowering and lint yield parameters. (A) Delayed flowering period of 156O/E, MIM156 and control plants (B) Seed cotton weight in transgenic lines compared with the wild type; the value is average weight of 10 MIM156 and 6 miR156 O/E plants ± SD (C) Lint cotton weight in transgenic lines compared with the wild type; the value is average weight of 10 MIM156 and 6 miR156 O/E plants ± SD (D) 100-seed weight ±SD was calculated from three replicates of per 100 seeds weight in transgenic line and control (n=10 & 6). *P-value < 0.05; **P-value < 0.01 (Student’s t test).

The result indicates that the fiber quality parameters like fiber length, strength, micronaire,

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Span Length (2.5%;mm)</th>
<th>Strength (3.2mm; g/Tex)</th>
<th>Micronaire</th>
<th>UR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OE156</td>
<td>29.4-30.0</td>
<td>22.0-22.4</td>
<td>4.0-4.4</td>
<td>49-50</td>
</tr>
<tr>
<td>Control</td>
<td>30.4-30.8</td>
<td>23.2-23.8</td>
<td>3.5-3.8</td>
<td>50-51</td>
</tr>
<tr>
<td>MIM156</td>
<td>30.0-30.5</td>
<td>25.0-25.8</td>
<td>3.3-3.5</td>
<td>50-51</td>
</tr>
</tbody>
</table>
uniformity ratio and short fiber index remains unchanged in 156 O/E and MIM156 transgenic plants as compared with control (Table 4.2). The results on 156 O/E and MIM156 transgenic lines suggests that although results matches well in many instances with results obtained with O/E and RNAi lines of GhSPL5D. However, in certain instances 156 O/E and MIM156 over expressing lines showed completely unexpected phenotypes indicating much complicated interplay between miRNA156 and GhSPL5D.

4.13 RNA-seq analysis of 0DPA initiating fibers

To further understand the regulatory mechanism by which GhSPL5D regulate cotton fiber initiation, we performed RNA-seq analysis to identify genes differentially expressed in 0DPA fiber cells in between control and KD lines. The RNA extracted from 0 DPA ovules of control and KD lines was sequenced on Illumina platform. A total 174438416 and 104219940 high quality reads were obtained after quality filtering in control and KD-lines respectively, approximately 63.33% of the reads mapped to the cotton D genome.
Table No. 4.3 - Top 15 up-regulated transcripts in SPL5 knockdown line:

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>TAIR ID</th>
<th>Short Description</th>
<th>FPKM_Control</th>
<th>FPKM_KD</th>
<th>Log2 FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gorai.006G119900</td>
<td>AT5G60490.1</td>
<td>FASCICLIN-like arabinogalactan 12</td>
<td>0.699771</td>
<td>4296.22</td>
<td>12.5839</td>
</tr>
<tr>
<td>Gorai.001G084300</td>
<td>AT1G71140.1</td>
<td>MATE efflux family protein</td>
<td>0.0316749</td>
<td>143.955</td>
<td>12.15</td>
</tr>
<tr>
<td>Gorai.006G240600</td>
<td>AT3G22620.1</td>
<td>lipid-transfer protein</td>
<td>0.166298</td>
<td>636.529</td>
<td>11.9022</td>
</tr>
<tr>
<td>Gorai.007G057400</td>
<td>AT4G37800.1</td>
<td>XTH7</td>
<td>0.707324</td>
<td>2240.11</td>
<td>11.6289</td>
</tr>
<tr>
<td>Gorai.002G174600</td>
<td>AT1G14870.1</td>
<td>PLANT CADMIUM RESISTANCE 2</td>
<td>0.0301867</td>
<td>85.0259</td>
<td>11.4598</td>
</tr>
<tr>
<td>Gorai.010G223200</td>
<td>AT1G78970.1</td>
<td>lupeol synthase 1</td>
<td>0.0561738</td>
<td>142.417</td>
<td>11.3079</td>
</tr>
<tr>
<td>Gorai.009G028500</td>
<td>AT2G19770.1</td>
<td>profilin 5</td>
<td>0.290847</td>
<td>736.437</td>
<td>11.3061</td>
</tr>
<tr>
<td>Gorai.011G103900</td>
<td>AT2G23540.1</td>
<td>GDSL-like Lipase/Acylhydrolase protein</td>
<td>0.0185034</td>
<td>40.6668</td>
<td>11.1018</td>
</tr>
<tr>
<td>Gorai.013G194500</td>
<td>AT5G54160.1</td>
<td>O-methyltransferase</td>
<td>0.0201471</td>
<td>43.4951</td>
<td>11.0761</td>
</tr>
<tr>
<td>Gorai.010G022200</td>
<td>AT4G06536.1</td>
<td>SPLa/Ryanodine receptor (SPRY) protein</td>
<td>0.0087026</td>
<td>17.4056</td>
<td>10.9658</td>
</tr>
<tr>
<td>Gorai.003G165800</td>
<td>AT4G12320.1</td>
<td>cytochrome P450, family protein</td>
<td>0.0535535</td>
<td>75.6096</td>
<td>10.4634</td>
</tr>
<tr>
<td>Gorai.005G233500</td>
<td>AT4G37450.1</td>
<td>arabinogalactan protein 18</td>
<td>6.21934</td>
<td>8744.06</td>
<td>10.4573</td>
</tr>
<tr>
<td>Gorai.009G035800</td>
<td>AT5G20630.1</td>
<td>germin</td>
<td>3 2.32393</td>
<td>3002.76</td>
<td>10.3355</td>
</tr>
<tr>
<td>Gorai.006G068400</td>
<td>AT2G33380.1</td>
<td>Caleosin-related family protein</td>
<td>0.392901</td>
<td>493.294</td>
<td>10.2941</td>
</tr>
<tr>
<td>Gorai.008G245300</td>
<td>AT2G45800.1</td>
<td>GATA type zinc finger transcription factor</td>
<td>0.111735</td>
<td>131.978</td>
<td>10.206</td>
</tr>
</tbody>
</table>
## Results

Table No. 4.4 - Top 15 down-regulated transcripts in SPL5O/E line:

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene ID</th>
<th>Description</th>
<th>FDR</th>
<th>log2FC</th>
<th>FDR-adjusted log2FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gorai.006G030800</td>
<td>AT3G27810.1</td>
<td>myb domain protein 21</td>
<td>7.5671</td>
<td>0.0458202</td>
<td>-7.36735</td>
</tr>
<tr>
<td>Gorai.011G016000</td>
<td>AT1G11600.1</td>
<td>cytochrome P450 family</td>
<td>4.18636</td>
<td>0.0280732</td>
<td>-7.22036</td>
</tr>
<tr>
<td>Gorai.005G222400</td>
<td>AT1G17840.1</td>
<td>white-brown protein 11</td>
<td>2.72994</td>
<td>0.0223456</td>
<td>-6.93273</td>
</tr>
<tr>
<td>Gorai.002G248100</td>
<td>AT2G35030.1</td>
<td>Pentatricopeptide repeat protein</td>
<td>2.46106</td>
<td>0.0204317</td>
<td>-6.91233</td>
</tr>
<tr>
<td>Gorai.007G163500</td>
<td>AT1G14770.1</td>
<td>RING/FYVE/PHD zinc finger protein</td>
<td>5.17528</td>
<td>0.0443005</td>
<td>-6.86817</td>
</tr>
<tr>
<td>Gorai.008G092000</td>
<td>AT4G24660.1</td>
<td>homeobox protein</td>
<td>2218.7513</td>
<td>0.181086</td>
<td>-6.69417</td>
</tr>
<tr>
<td>Gorai.002G185900</td>
<td>AT1G23420.1</td>
<td>Plant-specific YABBY transcription factor</td>
<td>35.5271</td>
<td>0.359628</td>
<td>-6.62627</td>
</tr>
<tr>
<td>Gorai.003G160400</td>
<td>AT4G11650.1</td>
<td>osmotin 34</td>
<td>4.69053</td>
<td>0.0483381</td>
<td>-6.60045</td>
</tr>
<tr>
<td>Gorai.005G078100</td>
<td>AT5G61890.1</td>
<td>Integrase-type DNA-binding protein</td>
<td>5.00978</td>
<td>0.0590687</td>
<td>-6.40621</td>
</tr>
<tr>
<td>Gorai.009G095600</td>
<td>AT5G59120.1</td>
<td>subtilase 4.13</td>
<td>7.30898</td>
<td>0.0903747</td>
<td>-6.33761</td>
</tr>
<tr>
<td>Gorai.002G005800</td>
<td>AT1G75250.1</td>
<td>RAD-like 6</td>
<td>21.1958</td>
<td>0.28266</td>
<td>-6.22857</td>
</tr>
<tr>
<td>Gorai.001G086800</td>
<td>AT5G37800.1</td>
<td>RHD SIX-LIKE 1</td>
<td>2.48974</td>
<td>0.0381597</td>
<td>-6.0278</td>
</tr>
<tr>
<td>Gorai.005G234700</td>
<td>AT2G23260.1</td>
<td>UDP-glucosyl transferase 84B1</td>
<td>5.61839</td>
<td>0.0881493</td>
<td>-5.99406</td>
</tr>
<tr>
<td>Gorai.010G028100</td>
<td>AT4G37750.1</td>
<td>Integrase-type DNA-binding protein</td>
<td>98.9135</td>
<td>1.6663</td>
<td>-5.89137</td>
</tr>
<tr>
<td>Gorai.013G005900</td>
<td>AT2G40470.1</td>
<td>LOB domain-containing protein 15</td>
<td>4.44509</td>
<td>0.0798599</td>
<td>-5.7986</td>
</tr>
<tr>
<td>Gorai.002G068300</td>
<td>AT3G15270.1</td>
<td>squamosa promoter binding protein-like 5</td>
<td>14.6259</td>
<td>2.27414</td>
<td>-2.68513</td>
</tr>
</tbody>
</table>
The estimated gene-expression distribution was identified in FPKM (Fragment per Kilo per Million) unit using cufflink tool (Trapnell et al., 2012). Next, we checked expression of \textit{GhSPL5D} in control and KD-line, as expected the expression of \textit{GhSPL5D} was found to be about 8 fold down in KD-line which was significantly lower than that of its normal expression in control plants (Figure 4.23A). The differentially expressed genes (DEGs) were identified using the cuffdiff program with a false discovery threshold < 5\% and a fold change ≥ 2. We identified, 2993 significantly up-regulated genes (FDR<5\%, FC ≥ 2), and 1462 down-regulated genes (FDR<5\%, FC≤-2) in KD as compared to control. The top 15 up and down regulated genes are given in table 4.3 and 4.4. The genes which were down regulated in KD-lines were expected to be positively regulated by \textit{SPL5}, thus it was interesting to note that the most affected gene in KD line was \textit{myb21}. The role of MYB in fiber initiation is already well established (Walford et al., 2011). The other important down-regulated genes were cytochrome P450, white brown protein11, homeobox, YABBY transcription factor and osmotin 34 etc. (Table 4.3). The top up-regulated genes expected to be negatively regulated by \textit{SPL5} includes several Arabinogalactan protein genes, these proteins reported to play role in cell adhesion during initial stages of fiber development (Huang et al., 2013). Some other upregulated genes were lipid transfer protein, profilin5 and GDSL lipase etc. (Table 4.3).

4.14 Gene ontology analysis of differentially expressed genes

We next fetch the Gene ontology (GO) of DEGs identified in KD-lines, GO analysis showed that various important pathways were down-regulated including nucleus, membrane bound organelle, cell part, vegetative to reproductive phase transition of meristem, trichome morphogenesis, trichome differentiation, seed development, ovule development, fruit development, regulation of flower development and ATP binding (Figure 4.23B). Whereas up-regulated genes includes processes like secondary metabolic process, response to oxidative stress, response to osmotic stress, response to Abscisic acid (ABA), programmed cell death and flavonoid biosynthetic process, indicating \textit{GhSPL5D} negatively regulate these processes.
Figure 4.23 - Pathway analysis of *GhSPL5D* KD initiating fibers. (A) Bar graph of *GhSPL5D* FPKM values in control and *GhSPL5D* KD line. P-value < 0.01 (B) Gene ontology analysis of up and down genes in *GhSPL5D* KD lines; F denotes molecular function, P denotes biological processes and C denotes cellular component.
4.15 Metabolic pathway analysis of differentially expressed genes

We next also mapped DEGs onto MapMan (Usadel et al., 2009) to further understand the various metabolic and molecular pathways that might have affected in KD-lines. Interestingly we observed metabolic pathways belonging to plant hormones viz. auxin, ethylene and gibberellin, were down-regulated in KD line while ABA, jasmonic acid and cytokinin were up-regulated in KD initiating fibers as compared to control (Figure 4.24).

Figure 4.24 – The metabolic pathway analysis. MapMan analysis of up (Blue) and down (Red) genes in GhSPL5D KD lines as compared to control
Further, transcription factors such as MYB, TUB and ARF were down-regulated and C2H2 Zn finger family and RNA regulation of transcription were upregulated in \textit{GhSPL5D} KD fibers. The carbohydrate (CHO) synthesis was down-regulated and degradation was up-regulated in KD fibers.

The flavonoids and stress (biotic & abiotic) metabolism were up-regulated whereas lipid and glycolysis metabolism were down-regulated in KD fiber during initiation.