

## RESEARCH ARTICLE

# Genome-wide identification and expression analyses of *Sm* genes reveal their involvement in early somatic embryogenesis in *Dimocarpus longan* Lour

Xue Li<sup>1,2</sup>, Yan Chen<sup>1,2</sup>, Shuting Zhang<sup>1,2</sup>, Liyao Su<sup>1,2</sup>, Xiaoping Xu<sup>1,2</sup>, Xiaohui Chen<sup>1,2</sup>, Zhongxiang Lai<sup>1,2</sup>, Yuling Lin<sup>1,2\*</sup>

**1** College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China, **2** Institute of Horticultural Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

\* [buliang84@163.com](mailto:buliang84@163.com)



## OPEN ACCESS

**Citation:** Li X, Chen Y, Zhang S, Su L, Xu X, Chen X, et al. (2020) Genome-wide identification and expression analyses of *Sm* genes reveal their involvement in early somatic embryogenesis in *Dimocarpus longan* Lour. PLoS ONE 15(4): e0230795. <https://doi.org/10.1371/journal.pone.0230795>

**Editor:** Meng-xiang Sun, Wuhan University, CHINA

**Received:** January 5, 2020

**Accepted:** March 8, 2020

**Published:** April 3, 2020

**Copyright:** © 2020 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The genome database of *Dimocarpus longan* has been upload at NCBI (accession number: BioProject PRJNA305337). The transcriptome datasets of four developmental stages (EC, ICpEC, GE and NEC) have been upload to NCBI (SRA accession: PRJNA565345). The transcriptome datasets of nine organs of 'Sijimi' was deposited at NCBI under GSE84467. Other unpublished datasets presented in this study are provided either in the manuscript or supporting information.

## Abstract

The *Sm* proteins are a conserved protein family with *Sm* motifs. The family includes *Sm* and *Sm*-like proteins, which play important roles in pre-mRNA splicing. Most research on the *Sm* proteins have been conducted in herbaceous plants, and less in woody plants such as *Dimocarpus longan* (longan). And the embryo development status significantly affects the quality and yield of longan. In this study, we conducted a genome-wide analysis of longan *Sm* genes (*DISm*) to clarify their roles during somatic embryogenesis (SE) and identified 29 *Sm* genes. Phylogenetic analysis deduced longan *Sm* proteins clustered into 17 phylogenetic groups with the homologous *Sm* proteins of *Arabidopsis thaliana*. We also analyzed the gene structures, motif compositions, and conserved domains of the longan *Sm* proteins. The promoter sequences of the *DISm* genes contained many light, endosperm development, hormone, and temperature response elements, which suggested their possible functions. In the non-embryogenic callus(NEC) and during early SE in longan, the alternative splicing(AS) events of *DISm* genes indicated that these genes may influence SE development by changing gene structures and sequences. The kinetin(KT) hormone, and blue and white light treatments affected the differentiation and growth of longan embryonic callus(EC) probably by affecting the AS events of *DISm* genes. Expression profiles showed the possible functional divergence among *Sm* genes, and different hormones and light qualities affected their expression levels. The expression trends of the *DISm* genes determined by RNA sequencing as fragments per kilobase of exon model per million mapped reads (FPKM) and by real-time quantitative PCR(qRT-PCR) during early SE in longan showed that the expression of the *DISm* genes was affected by the growth and differentiation of longan SE, and decreased as the somatic embryo differentiation progressed. The results will contribute to understanding the longan *Sm* gene family and provide a basis for future functional validation studies.

**Funding:** The fees for high-throughput sequencing and data processing were supported by the National Natural Science Foundation of China (31672127), the fees for quantitative real time polymerase chain reaction detection were supported by the Constructions of Plateau Discipline of Fujian Province (102/712201801101), the publication fees were supported by the Technology Innovation Fund of Fujian agriculture and forestry university (CXZX2017187 and CXZX2018078). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

*Sm* family proteins are a family of RNA-binding proteins found in bacteria, archaea, and eukaryotes, including *Sm* and *Sm*-like (*LSm*) proteins [1]. Structurally, *Sm* proteins have a protein-protein interaction site, the *Sm* motif, consisting of two parts, *Sm*1 and *Sm*2 [2, 3]. The *Sm* motif contains an amino-terminal  $\alpha$ -helix and an five-stranded antiparallel  $\beta$ -sheet. The amino-terminal  $\alpha$ -helix and the neighboring  $\beta$ 1 $\beta$ 2 $\beta$ 3 sheet fold to form the *Sm*1 motif, and the  $\beta$ 4 $\beta$ 5 sheet folds to form the *Sm*2 motif, and these motifs are involved in the protein-protein interactions that enable the *Sm* proteins to assemble into RNA-interacting heptameric ring complexes [4, 5]. The *Sm* core proteins include mainly *Sm*B, *Sm*B', *Sm*N, *Sm*D1, *Sm*D2, *Sm*D3, *Sm*E, *Sm*F and *Sm*G, which stably bind with U1, U2, U4 and U5 small nuclear RNAs (snRNAs) and other proteins to form particles called small nuclear ribonucleoproteins or snRNPs in the nucleus of cells [6, 7]. These snRNPs can form a larger complex, the spliceosome, which can cleave introns not encoded by the pre-mRNAs, splicing the exons together, and then undergoing a series of modifications to become mRNAs [6], and the mRNAs are transported through the nuclear pores into the cytosol and eventually translated into proteins by ribosomes eventually [6]. The central component of the spliceosome, U6 snRNA, is very specific, it binds only to the *LSm* proteins and not to *Sm* proteins [7, 8]. The *LSm* proteins belong to the protein family that has a conserved *Sm*-like motif and are homologous with the *Sm* proteins. The *LSm* proteins comprises mainly nine proteins (*LSm*1-9), most of which (*LSm*1-8) are involved in forming heptameric ring complexes. *LSm*1-7 form a cytoplasmic complex that can bind to oligoadenylated mRNA and promote mRNA capping. *LSm*2-8 form a complex that can stabilize U6 snRNA and play a role in pre-mRNA splicing. *LSm*2-7 have been shown to be closely related to their *Sm* counterparts, *Sm*D1 to *Sm*G, whereas *LSm*1 and *LSm*8 may be closely related to the *Sm*B subfamily [8-10]. Larger proteins that not only have an *LSm* domain but also other protein domains, (e.g., *LSm*10-16, ataxin-2, and archaeal *Sm*3) have been found recently [11].

Until now, most of the studies on the *Sm* protein family have focused mainly on protein structure and splicing principle, and mainly in bacteria, yeast, and humans. Stark et al. [12] identified seven *LSm* proteins (*LSm*1-7) in red algae, indicating that red algae do not contain *LSm*8. Subsequent studies by Reimer et al. [13] demonstrated that the *LSm*1-7 complex was involved in splicing and cytoplasmic mRNA degradation. In higher plants, the *Sm* protein family has been identified only in *Arabidopsis*, rice, and maize. The genetic structure, phylogeny, chromosomal location, selection stress and expression analysis of the *Sm* gene family have been reported in *Arabidopsis* [14]. In rice and corn, the systematic development, chromosome mapping, gene organization, adaptive evolution, expression profile, and functional network of the *Sm* gene family have been studied [14, 15]. The functions of some of the *Arabidopsis* *Sm* proteins have been verified. For example, the expression of *LSm*5 was found to regulate the circadian clock and the mutation of *LSM5* or *LSM4* prolonged the circadian rhythm in *Arabidopsis* [16]. The function of the *LSm*2-8 complex in pre-mRNA splicing is controlled by external signals, which differentially regulated the adaptability of *Arabidopsis* to abiotic stress [17]. Further, overexpression of the *SAD1* gene, which encodes the *LSm*5 protein enhanced the salt tolerance of *Arabidopsis* [18]. Thus, current knowledge about the *Sm* proteins indicates they can promote snRNA cap modification and target snRNPs to the appropriate cell locations, but it is unclear whether they have other functions. Until now, most of the research on the *Sm* protein family has been conducted in herbaceous plants, so not much is known about their roles in woody plants, especially in *Dimocarpus longan* (longan).

Longan is an important tropical and subtropical woody fruit tree in China. The fruit is highly nutritious and the embryogenesis status of the plant significantly affects fruit quality

and yield. Cell biology and molecular biology studies of somatic embryogenesis (SE) in longan have greatly progressed the understanding of the SE process. A large number of molecular and gene regulation mechanisms inherent in the development of longan SE have been studied in our laboratory. Until now, the focus has been on the regulation of microRNAs[19, 20], regulation of DNA methylation-related genes[21, 22], secondary metabolism [23, 24], and stress[25, 26], as well as genes related to the regulation of growth and development[27–30] related genes in longan SE. However no studies have investigated alternative splicing factors associated with longan SE, and the function of most plant *Sm* proteins is still unclear.

In this study, we conducted a the genome-wide analysis, alternative splicing(AS) analysis, and investigated the expression partterns of *Sm* genes in longan. Bioinformatics analysis of the longan *Sm* family was carried out, and 29 *Sm* genes were identified. The physicochemical properties, conserved motifs and protein domains and phylogenetic tree of the *Sm* protein sequences, and the *cis*-acting elements in the promoters of the *Sm* genes were analyzed. The AS events of the *Sm* genes in the non-embryogenic callus(NEC) stage and the early SE, longan embryonic callus(EC) stages, were also analyzed under different hormone and light quality treatments. Finally, the FPKM values of the *Sm* genes in NEC, early SE stages and different longan tissues were also analyzed. The expression of *DLSm* genes in longan SE stages were verified by real-time quantitative PCR(qRT-PCR). The results of this study will provide a basic resource for future research on the function of the longan *Sm* gene family.

## Materials and methods

### Plant materials

The original experimental material was 'Honghezi' longan EC, which was provided by the Institute of Horticultural Biotechnology, Fujian Agriculture and Forestry University[31]. We used the culture method of Lai Zhongxiong as reference to obtain the longan EC, incomplete embryogenic compact structure(ICpEC), globular embryo(GE) and NEC by adding different concentrations of 2,4-Dichlorophenoxyacetic acid(2,4-D) to the culture medium[32, 33]. The original experimental materials used in qRT-PCR and the original experimental materials used in transcriptome sequencing were taken from different single plant of 'Honghezi' cultivar. This study did not involve any near threatened species and did not require a statement of ethical approval.

### Identification of *DLSm* gene family members

We retrieved *Arabidopsis* *Sm* protein sequences[15] and then used the *Arabidopsis* genomic *Sm* amino acid sequence as a probe sequence and local blasting with the longan genome database(NCBI accession number: BioProject PRJNA305337)[34]. 29 longan *Sm* genes were obtained by screening annotations. According to the genome annotation, we used the NCBI-BLAST homologous alignment analysis to compare the genomic DNA(gDNA) sequences, coding sequences(CDS), and amino acid sequences of *Sm* with DNAMAN6.0 software. According to the annotations of the *Sm* family of *Arabidopsis thaliana* in the longan genome and their chromosomal location, the longan *Sm* genes were named and identified by reference to the common nomenclature of most species in the NCBI-BLAST homologous alignment. Among them, *DLSm14b*, *DLSmD3* and *DLSm7b* had no complete amino acid sequences, and subsequent analysis of protein physicochemical properties, gene structure, conserved motif, and protein domain will not be carried out.

## Analysis of the phylogenetic tree and the conservative domain of the DLSm protein family system

The amino acid sequences of longan and *Arabidopsis Sm* genes were compared by DNAMAN software. Then based on neighbor-joining(NJ) method, 61 *Sm* genes amino acids from longan and *Arabidopsis thaliana* were constructed by MEGA7 software. Next, the sequences with large homology differences were manually deleted and 1000 boot strap samples were carried out. Combined with CDD(Conserved Domain Database)(<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), the conserved domains of protein sequences in longan and *Arabidopsis thaliana* were analyzed by ITOL (<https://itol.embl.de/>).

## DLSm gene structure and conservative domain analysis

We analyzed the introns and exons of 61 *Sm* genes in longan and *Arabidopsis thaliana* by GSDS2.0 (<http://gsds.cbi.pku.edu.cn/>). We used ExPASy (<https://www.expasy.org/>) to predict the basic physical and chemical properties of 26 DLSm amino acid sequences, including the number of amino acids(aa), molecular weight (Mw), and other electric point (pI), N-terminal signal peptides. After MEME online analysis (<http://meme-suite.org/>) of conservative motif of longan *Sm* proteins, ttools was used to draw the obtained motif.

## Analysis of cis-acting elements of DLSm genes promoter

We extracted the promoter sequences of 3000bp upstream of ATG in *DLSm* genes. PlantCARE online website was used to predict its *cis*-acting elements, in which the *DLSm1a* promoter sequence contained many N, and the gene promoter was not analyzed for the time being. The different types of *cis*-acting elements of each gene promoter were counted and plotted by Excel software.

## Analysis of AS of DLSm gene family

The longan transcriptome datasets were used to analyze the AS events of *DLSm* gene family in longan NEC and the early stage of SE(EC, ICpEC and GE)(NCBI SRA accession: PRJNA565345)[35]. The datasets of longan EC treated with different hormones and longan EC under two different light-quality treatments(blue and white light) have not been published. The hormone treatments were as follows: 2,4-D&KT: Murashige and Skoog(MS) medium + 2,4-D(1.0 mg/L)+ KT(0.5 mg/L), 2,4-D: MS medium+ 2,4-D(1.0 mg/L), KT: MS medium + KT(0.5 mg/L) and MS medium without any hormones. The materials were dark cultured in these liquid medium for 24 hours. The datasets of light quality treatment have been published, and the specific processing methods refer to the study of Li et al.[36]. Adding 20 g·L<sup>-1</sup> sucrose, 6 g·L<sup>-1</sup> agar and 4.5μM 2,4-D to MS medium for light treatment experiments of longan EC. The starting material for each bottle was 0.04 g. Longan ECs were incubated in a plant light chamber for 25 days and darkly treated as a control. The intensity and period of blue light (457 nm) and white light were fixed at 32μmol·m<sup>-2</sup>·s<sup>-1</sup> and 12 h·Day<sup>-1</sup>, respectively[36]. All experimental materials were stored at -80°C for subsequent analysis. We used rMATS[37] to detect the AS events between samples. AS events under different developmental stages and different hormone treatments of longan EC were collected from more than three samples among every dataset. Under different light quality treatments, the AS events were collected from more than two samples among every dataset.

## Analysis of specific expression of *DLSm* family genes

We introduced the FPKM value of nine organs (flower bud, flower, leaf, pericarp, pulp, root, seed, stem, young fruit) of 'Sijimi' longan cultivar (the datasets was deposited at NCBI under GSE84467) and NEC, EC, ICpEC, GE of 'Honghezi' longan cultivar [34, 35], and longan EC treated with different hormones (the datasets have not been published) and light quality [36] to the Omicshare online website to map the heat expression of gene expression (<http://www.omicshare.com/tools/index.php/>). The disclosure of the transcriptome datasets is the same as above.

## Real-time quantitative PCR

The total RNA of longan was extracted by TransZol Up Kit and complementary DNA (cDNA) was synthesized by PrimerScript RT Reagent Kit (TaKaRa). The specific primers for *DLSm* genes were designed by dnaman software. qRT-PCR refers to the method of Lin et al. [38], and we used eukaryotic elongation factor 1-alpha (*EF-1 $\alpha$* ) as internal reference gene. We amplified qRT-PCR on a Roche LightCycler 480 instrument (Roche Applied Science, Switzerland). The qRT-PCR reaction system (20  $\mu$ l) consists of 10  $\mu$ l SYBR premix ex Taq<sup>TM</sup> II (TaKaRa), 7.4  $\mu$ l ddH<sub>2</sub>O, 0.8  $\mu$ l upstream and downstream specific primers (S1 Table) and 1  $\mu$ l cDNA template. The reaction procedure is 95°C pre-denaturation 30 s, 95°C denaturation 5 s, 60°C annealing 30 s, 72°C extension 15 s, 40 cycles. We used the  $2^{-\Delta\Delta C_t}$  method to calculate expression values and then used Graphpad prism 7.0 to analyze the data and make the related image.

## Results

### Identification of *DLSm* gene family members

To identify members of the longan *Sm* gene (*DLSm*) family, the amino acid sequences of *Arabidopsis* *Sm* proteins were listed as probe sequences, and we searched the sequences by using local blast and NCBI BLAST. The initial screening identified 30 potential members. We found that only one base in the coding sequences of Dlo\_022245.1 and Dlo\_032946.1 (LSm6) was different, the translated amino acid sequences were exactly the same, and the translated start sites were also the same, so we regarded them as the same sequence, and used the Dlo\_032946.1 for further analysis. Therefore, 29 *Sm* genes were identified in the longan genome. The longan *Sm* genes were named on the basis of their conserved domains, chromosomal locations, and annotations of the homologous *Sm* genes in *Arabidopsis thaliana*, as *DLSm1-7*, *DLSm14a~DLSm14e*, *DLSmD1*, *DLSmB*, *DLSmD*, *DLSmD1a*, *DLSmD1b*, *DLSmD2a*, *DLSmD2b*, *DLSmD3*, *DLSmEa*, *DLSmEb*, *DLSmF*, and *DLSmG* (Table 1). The physicochemical properties of deduced *DLSm* proteins were analyzed using the ExPASy server. We found the *DLSm* amino acid sequences were from 74 to 862 amino acids long, their molecular weights were from 8682.79ku to 92952.89ku, and their electric points were from 4.45 to 11.73. None of the 29 *Sm* sequences had a signal peptide, so they were not secretory proteins.

### Evolutionary relationship of the longan and *Arabidopsis* *Sm* proteins family and analysis of the protein domains

To examine the phylogenetic relationship of the *Sm* proteins between *Arabidopsis thaliana* and longan, we used MEGA7 to construct an evolutionary tree based on their amino acid sequences. We found the *Sm* proteins clustered into 17 groups I-XVII (Fig 1), and each group contained the *Sm* proteins of *Arabidopsis* and longan. Group IV contained 11 LSm14 proteins and was the largest group. Among them, the *Arabidopsis* proteins AtLSm14a, AtLSm14c, and AtLSm14d in Group IV shared a closer relationships than other members of this group. There

**Table 1. Information of longan *Sm* family genes.**

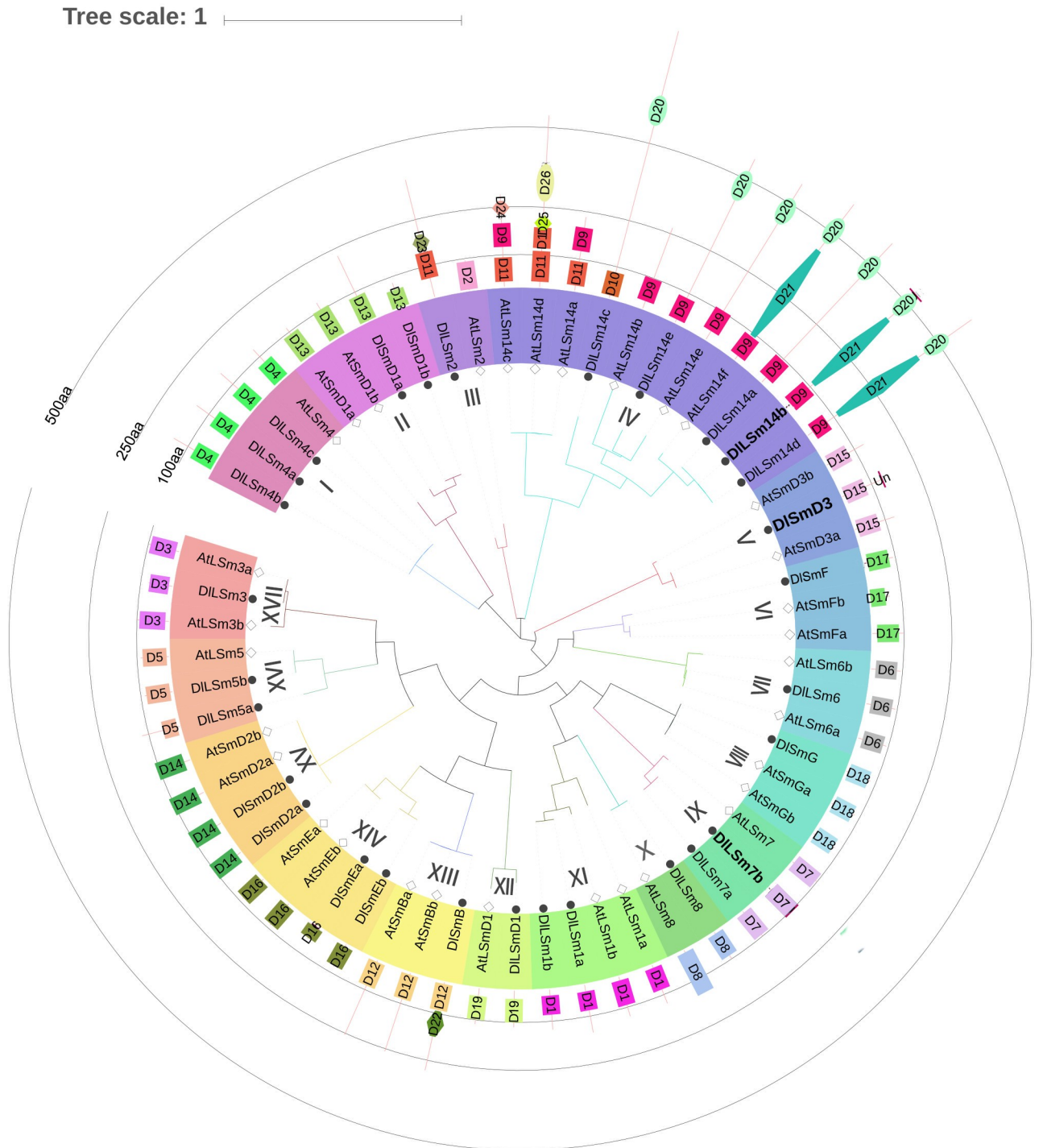
Gene ID	Gene name	Comments to members of <i>Arabidopsis</i>	Gene localization	aa number	Mw	pI	Signal peptide
Dlo_023124.1	<i>DLSm1a</i>	AT3G14080 <i>AtLSm1b</i>	scaffold504:56684:58077	151	17067.94	8.94	No
Dlo_023126.1	<i>DLSm1b</i>	AT3G14080 <i>AtLSm1b</i>	scaffold504:65206:66779	125	14179.59	5.26	No
Dlo_024845.1	<i>DLSm2</i>	AT1G03330 <i>AtLSm2</i>	scaffold556:163264:167575	377	42539.65	9.16	No
Dlo_009273.1	<i>DLSm3</i>	AT1G76860 <i>AtLSm3b</i>	scaffold185:484650:485838	74	8682.79	4.45	No
Dlo_014195.1	<i>DLSm4a</i>	AT5G27720 <i>AtLSm4</i>	scaffold263:327344:329240	146	15843.89	10.01	No
Dlo_021351.1	<i>DLSm4b</i>	AT5G27720 <i>AtLSm4</i>	scaffold449:612750:615216	150	16135.20	10.14	No
Dlo_031761.1	<i>DLSm4c</i>	AT5G27720 <i>AtLSm4</i>	scaffold845:311468:313948	150	16135.20	10.14	No
Dlo_023673.1	<i>DLSm5a</i>	AT5G48870 <i>AtLSm5</i>	scaffold52:850634:851892	120	13777.51	5.54	No
Dlo_012057.1	<i>DLSm5b</i>	AT5G48870 <i>AtLSm5</i>	scaffold224:406385:407477	87	9599.07	4.50	No
Dlo_032946.1	<i>DLSm6</i>	AT2G43810 <i>AtLSm6b</i>	scaffold908:10301:13180	91	9794.13	9.13	No
Dlo_028712.1	<i>DLSm7a</i>	AT2G03870 <i>AtLSm7</i>	Scaffold70:1068910:1071695	99	10746.27	4.79	No
Dlo_034364.1	<i>DLSm7b</i>	AT2G03870 <i>AtLSm7</i>	Scaffold1018:97362:100063	No termination codon			No
Dlo_036673.1	<i>DLSm8</i>	AT1G65700 <i>AtLSm8</i>	Scaffold4:715280:716076	99	10759.14	4.68	No
Dlo_026784.1	<i>DLSm14a</i>	AT1G26110 <i>AtDCP5/(LSm14g)</i>	scaffold62:1097537:1107145	596	63092.69	6.19	No
Dlo_038405.1	<i>DLSm14b</i>	AT1G26110 <i>AtDCP5/(LSm14g)</i>	scaffold62:1151845:1157828	No termination codon			No
Dlo_006168.1	<i>DLSm14c</i>	AT1G26110 <i>AtDCP5/(LSm14g)</i>	Scaffold145:358361:366094	862	92952.89	9.18	No
Dlo_011575.1	<i>DLSm14d</i>	AT1G26110 <i>AtDCP5/(LSm14g)</i>	scaffold217:3535:10764	627	67457.47	6.27	No
Dlo_029021.1	<i>DLSm14e</i>	AT5G45330 <i>AtDCP5-L/(LSm14f)</i>	Scaffold711:39163:43435	568	61111.37	8.62	No
Dlo_020209.1	<i>DLSmD1</i>	AT4G18372( <i>AtLSmD1</i> )	scaffold41:74616:74960	99	10876.35	8.61	No
Dlo_033912.1	<i>DIsmB</i>	AT4G20440 <i>AtSmBb</i>	scaffold98:392671:393516	281	29663.61	11.28	No
Dlo_024823.1	<i>DIsmD1a</i>	AT4G02840 <i>AtSmD1b</i>	scaffold556:554:6323	224	24725.29	10.24	No
Dlo_031697.1	<i>DIsmD1b</i>	AT4G02840 <i>AtSmD1b</i>	scaffold842:61514:61871	80	8856.32	11.73	No
Dlo_014464.1	<i>DIsmD2a</i>	AT3G62840 <i>AtSmD2b</i>	scaffold27:468436:469034	104	12077.18	9.74	No
Dlo_009060.1	<i>DIsmD2b</i>	AT3G62840 <i>AtSmD2b</i>	scaffold1806:24804:26735	105	12063.15	9.74	No
Dlo_038407.1	<i>DIsmD3</i>	AT1G20580 <i>AtSmD3b</i>	scaffold69:954276:957329	No termination codon			No
Dlo_015839.1	<i>DIsmEa</i>	AT2G18740 <i>AtSmEb</i>	Scaffold3:469665:469951	79	9194.89	9.89	No
Dlo_019122.2	<i>DIsmEb</i>	AT2G18740 <i>AtSmEb</i>	Scaffold38:1862876:1865058	79	9167.82	9.82	No
Dlo_021780.1	<i>DIsmF</i>	AT4G30220 <i>AtSmF(AtSmFa)</i>	scaffold46:632326:633944	88	9910.28	4.43	No
Dlo_009662.1	<i>DIsmG</i>	AT2G23930 <i>AtSmGa</i>	scaffold19:1245388:1247149	80	8906.33	9.22	No

For the convenience of drawing, the name of *Arabidopsis thaliana* gene in parentheses is named according to its protein domain and amino acid size, which is only applicable in this paper.

<https://doi.org/10.1371/journal.pone.0230795.t001>

were five *DLSm14* proteins in Group IV, and was also the largest branch of *DIsm* phylogeny, and one *DIsm* protein in each of nine groups.

Analysis of the protein domains in the longan and *Arabidopsis* of *Sm* proteins using the CDD and InterPro revealed that they all contained the *Sm*-like super family domain (Fig 1). The *DLSm2* protein contained the *LSm* superfamily domain and an HMA domain, whereas the *AtLSm2* protein only contained only the *LSm2* protein domain. The *DLSm14* proteins contained an FDF domain in addition to the *LSm* superfamily domain. The *DLSm14b* and *DLSm14d* proteins also contained an Atrophin-1 super family domain. The *DIsmB* protein contained the *LSm* superfamily domain and a DNA\_pol3\_delta2 superfamily domain. *AtLSm14a*, *AtLSm14c*, and *AtLSm14d*, which formed a cluster separate from the other members of Group IV, contained protein domains that were quite different from the domains those of the other members. Except for the *Sm* proteins in Groups III and Group IV, the proteins of other groups contained the same *LSm* superfamily domain, which suggested the functions of the *Sm* proteins in these two groups members may had changed during evolution. The *Sm*



**Fig 1. Longan and *Arabidopsis* Sm protein family evolutionary tree with their protein domains.** D1–D26 are LSm1, LSm2, LSm3, LSm4, LSm5, LSm6, LSm7, LSm8, LSm14-N, LSm14, LSm super family, SmB, SmD1, SmD2, SmD3, SmE, SmF, SmG, LSmD1, FDF, Atrophin -1, DNA-Pol3, HMA, Kelch1, F-box, BTB protein domains; labeled the CDS sequence of DILSm14b, DISmD3, DILSm7b without termination codon.

<https://doi.org/10.1371/journal.pone.0230795.g001>

protein sequences were generally short, and the characteristic LSm domain was located between positions 1 and 150aa, indicating that the domain was relatively conserved.

### ***DILSm* gene structure and conserved motif analysis**

Introns are an important part of eukaryotic genes and play key roles in genetic evolution. To investigate the mechanism of structural evolution of the *Sm* genes, we compared the exon–intron structures of the longan *Sm* genes (Fig 2). The results of GSDS analysis showed that the number of introns in the *DILSm* genes contained from zero to ten introns, with most of the genes having three to four introns, which was similar to what was found for the *Arabidopsis Sm* genes. The *Sm* genes that encoded members of Group IV contained six to ten introns, whereas the genes that encoded members of the other groups had one to five introns, which indicated they were quite different. In addition, the genes that encoded the members of each clustered usually had similar genetic structures and there were also large differences in the structures of the genes among the different groups. The genes encoding members of Groups XII and XIII had no introns. Overall, the numbers and locations of introns in the *Sm* genes varied widely among the different groups, probably because of differences in the intron gain and loss rates [39]. Besides, the differences in the lengths and structures of the longan *Sm* genes, the quantity and distribution of the exons and introns were different among the different groups, which may lead to differences in the differentiation of gene function.

A total of 20 different motifs were identified in the longan *Sm* proteins by MEME (Fig 2). The most closely related *Sm* proteins in the phylogenetic tree had similar motif compositions, reflecting the functional similarity between them. Of the 29 *DILSm* proteins, only *DILSmD1b* and members of Group IV members did not contain Motif 1, and 20 of all members contained Motif 6. The motif composition of the proteins in Group IV was very different from that of proteins in the other groups, which confirmed that Group IV was far away from the other groups. Together, these results provide a reference for the functional diversity of the longan *Sm* proteins.

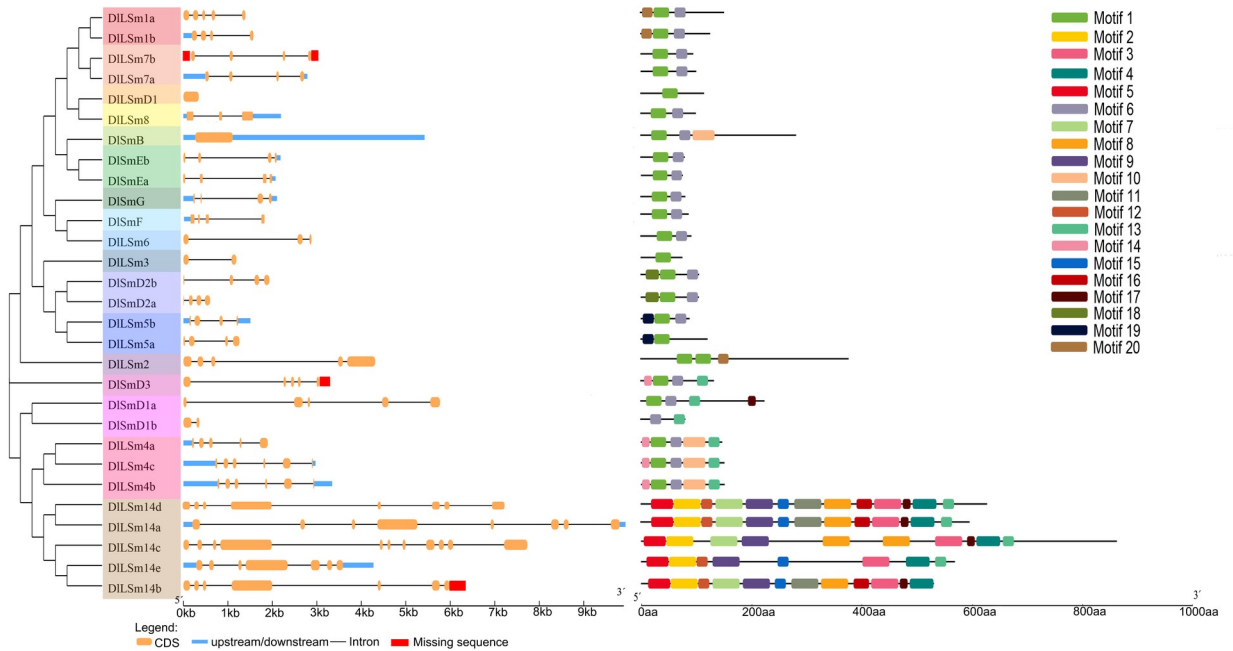
### **Analysis of *cis*-acting elements in the promoters of *DILSm* genes**

We used PlantCARE to analyze the promoters of 28 longan *Sm* genes and the results were shown in Fig 3. The promoters of all 28 genes contained the basic components such as CAAT-box and TATA-box. The identified *cis*-acting elements in these promoters were roughly classified into 5 categories: plant hormone response elements, stress response elements, tissue-specific related action elements, photoresponsive elements, and cell cycle response elements. The longan *Sm* genes contained a large number of photoresponsive elements, and 93% of the promoters sequences contained a large number of anaerobic induction response elements. More than 80% of the promoters contained abscisic acid (ABA) and methyl jasmonate (MeJA) response elements, about 75% of the members responded to wound, and 50% of the genes contained photoreactive MYB binding sites and protein metabolism regulatory elements. The *Sm* promoters also contained elements that responded to physiological circadian rhythm control and endosperm development. Among the 28 *Sm* genes, the *DILSm1b* and *DILSm5a* promoters contained MYB transcription factor action sites that regulate flavonoid biosynthesis; the *DILSm4c*, *DILSmEa*, *DILSmD2b* promoters respond to cell cycle regulation; and the *DILSmEa* promoter responds to seed-specific regulation. Seven of the *Sm* promoters (*DILSm4b*, *DILSm4c*, *DILSmD1a*, *DILSmB*, *DILSmEb*, *DILSm14c*, *DILSm14e*) contained *cis*-acting elements associated with meristem expression. These results can provide a basis for the regulatory network of longan *Sm* genes.

### **AS analysis of *DILSm* family in the NEC and early SE in longan**

We used our group's transcriptional databases and screened for four different types of splicing events in longan NEC and in three stages of early SE (EC, ICpEC and GE), namely alternative





**Fig 2. Structural analysis of *DLSm* genes and their encoded proteins.** The unknown sequences were located in the non-coding region or the non-conservative region, which did not affect the analysis results.

<https://doi.org/10.1371/journal.pone.0230795.g002>

3' splice site(A3SS) and alternative 5' splice site(A5SS), intron retention (IR), and exon skipping (ES). The results showed that, the number of AS events was more in the NEC(35) compared with the numbers in the other stages, and the number of AS events was lowest in the ICpEC(23) stage (Fig 4A). The number of IR events in the NEC stage was significantly higher than the numbers in the other three stages. There were up to nine A5SS events in GE stage, and the numbers of ES events were the same in all four stages. The results showed that the percentage of IR events was highest in NEC stage (49%) and IR events accounted for the largest proportion of AS events in the NEC stage (Fig 4B). There were no obvious differences in the percentages of AS events in EC, ICpEC and GE stages.

We also compared the distribution of AS events among the *DLSm* genes in longan NEC and early SE stages(Fig 5; S2 Table). The A3SS events had the highest frequency, occurring in nine, six, five and eight genes at the NEC, EC, ICpEC, and GE stages, respectively. In the NEC stage, 14 genes showed AS, and only three genes had more than two types of AS events(Fig 5A), among them, *DLSm14c* had four types of AS events occurred in *DLSm14c*. Only one type of AS events occurred in the other 11 genes(Fig 5A). Five genes had IR events, but the IR events in *DLSm1a* and *DLSm1b* occurred only in the NEC stage, where the number of IR events per gene was significantly higher than the numbers of the other types of AS events. The total number of IR events per genes also was the highest in the NECs and was higher than the other three AS types(Fig 5A). In the other three stages, although the number of IR events per gene was significantly higher than that of other types of AS events, the overall frequency of A3SS events was higher than that of IR events. The AS events in *DLSm1a* and *DLSm1b* occurred only in the NEC stage and in *DLSm6* they occurred only in the GE stage, whereas in *DLSm14c*, *DLSm14d*, *DLSm7a*, *DLSm7b*, *DLSmd2a*, and *DLSmEb* AS events occurred in all four stages(Fig 5). The AS events of *DLSm14e* were the highest in the NEC stage, and that of *DLSmD2a* were the highest in the ICpEC stage(Fig 5A and 5C). AS events of *DLSm14* were highest in the EC and GE stages(Fig 5B and 5D). The types of AS events types in *DLSm14c*

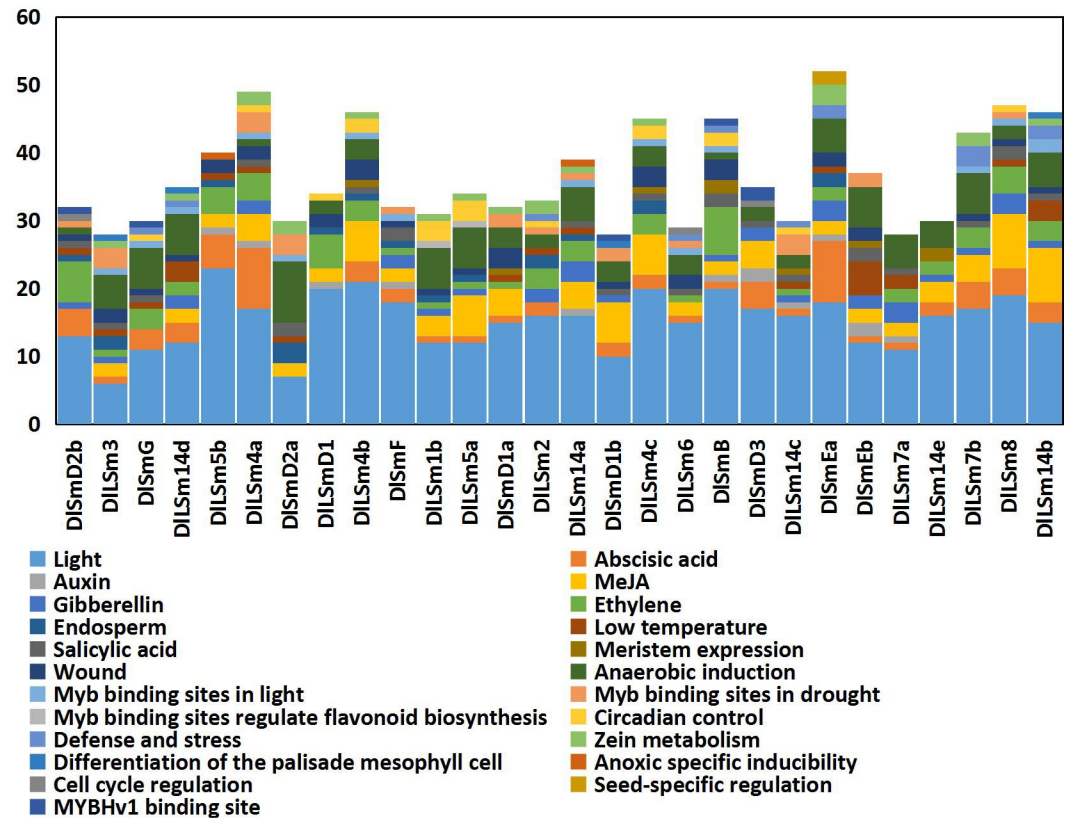


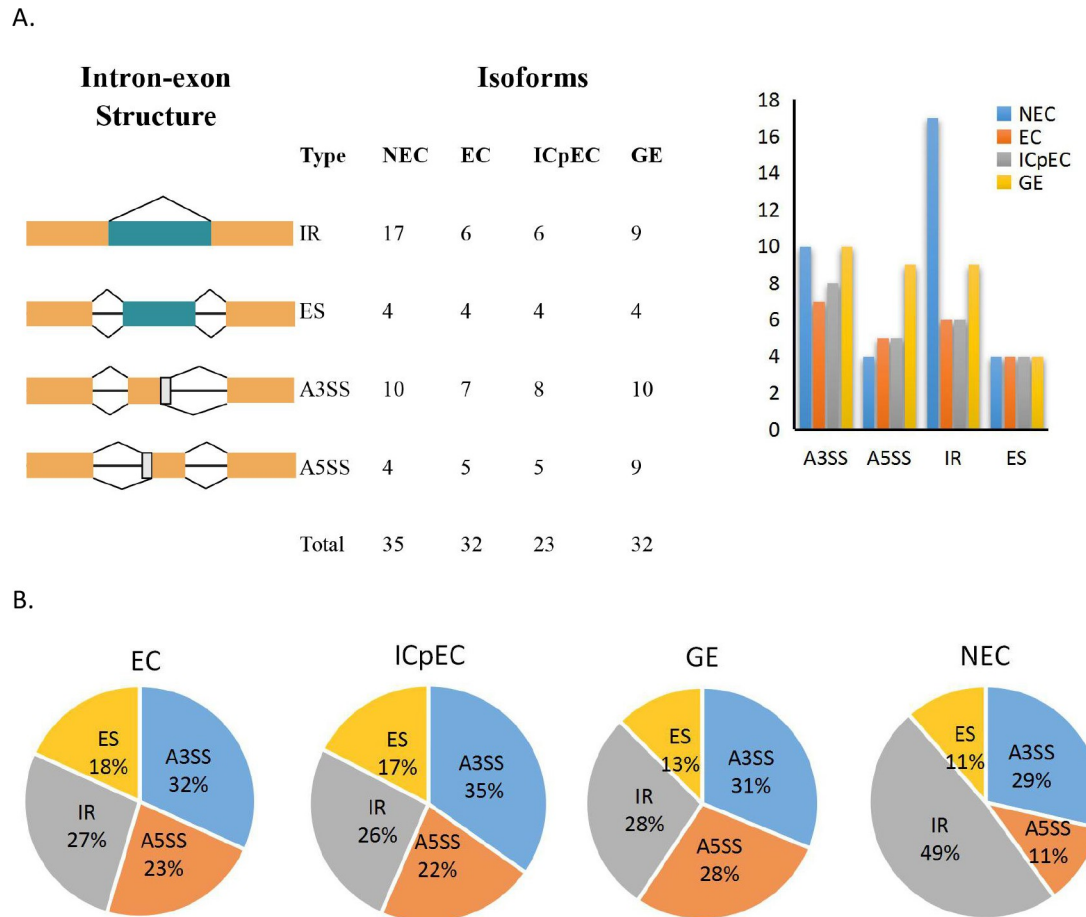
Fig 3. The distribution of *cis*-acting elements of the *DLSm* promoters. The 3.0 kb regulatory region upstream of ATG were analyzed by the PlantCARE software.

<https://doi.org/10.1371/journal.pone.0230795.g003>

decreased gradually as SE progressed; that is, four, two and one types of AS events occurred in the EC, ICpEC and GE stages, respectively (Fig 5B–5D). Our study indicated that a large number of AS events detected in the NEC stage may affect SE in longan in particular, the high occurrence of IR events. The observed decrease in the types of AS events in *DLSm14c* as SE progressed may promote the differentiation of longan EC stage into GE stage.

### AS analysis of *DLSm* family in longan EC stage treated with different hormones

2,4-D and KT play important roles in the growth and differentiation of longan EC [31, 40, 41]. To further explore the effects of different hormones on AS events of *DLSm* genes, we used our group's transcription data to analyze the AS events of longan EC treated with 2,4-D&KT, 2,4-D, KT and MS medium as control. All three hormone treatments reduced the numbers of the AS events in the *DLSm* genes compared with numbers in the control (Fig 6A). AS events occurred in four genes *DLSm4a*, *DLSm3*, *DLSm5b*, *DLSm14c* (Fig 6B; S3 Table), and were highest in *DLSm14c*. Further the numbers of AS events that occurred under each on the hormone treatments were the same in all the genes, except *DLSm14c*, in which the numbers of AS events were different under the different treatments. Under the 2,4-D&KT and 2,4-D treatments, the IR events of *DLSm14c* increased, and A3SS events occurred only in the control (Fig 6C). In addition, two, two, and one specific AS events occurred in *DLSm14c* under the 2,4-D&KT, 2,4-D and KT treatments, respectively, compared with the control (Fig 6D). These



**Fig 4. Distribution of AS events in NEC and early SE stages of longan.** A four splicing event patterns and their number of events in longan early SE. B distribution of AS events in longan NEC, EC, ICpEC, GE stages.

<https://doi.org/10.1371/journal.pone.0230795.g004>

results suggested that 2,4-D and KT may affect the growth and differentiation of longan EC by affecting the AS of *DILSm14c*.

### AS analysis of the *DISm* family in longan EC treated with different light qualities

The growth and physiological changes of longan EC are known to be affected by light, and different light qualities have different effects on their growth [36, 42–44]. We analyzed the transcriptional data to explore the relationship between light quality and AS of *DISm* genes. Longan ECs were treated with blue light and white light, and dark as the control (Fig 7A). We found the total numbers of AS events did not change after the blue light treatment, but decreased after the white light treatment compared with the control. We found that AS events occurred in ten *DISm* genes under the blue and white light treatments, and in 11 *DISm* genes under the dark treatment. The highest numbers of AS events occurred in *DILSm14c* under the three treatments, and IR events occurred only in *DILSm14c* under the three treatments (S4 Table). Under the blue light treatment, the A5SS and ES events of *DILSm14c* were one more than under the other two treatments (Fig 7B). Compared with the dark treatment, four specific AS events were detected in the *DISm* genes under the blue light treatment (Fig 7D) and one

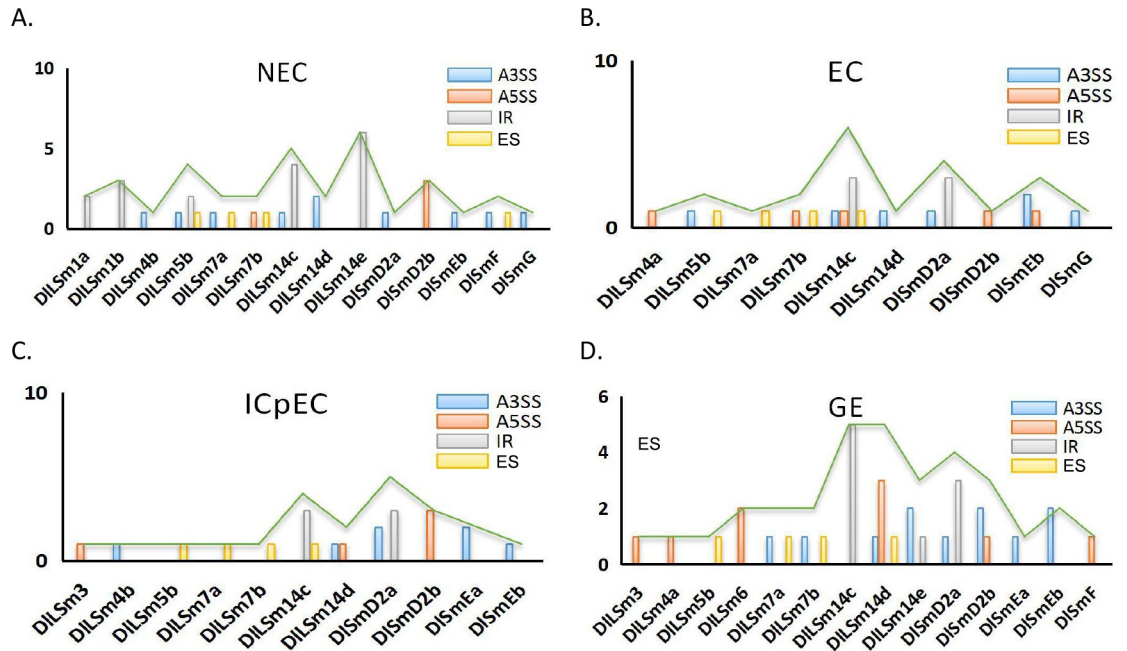


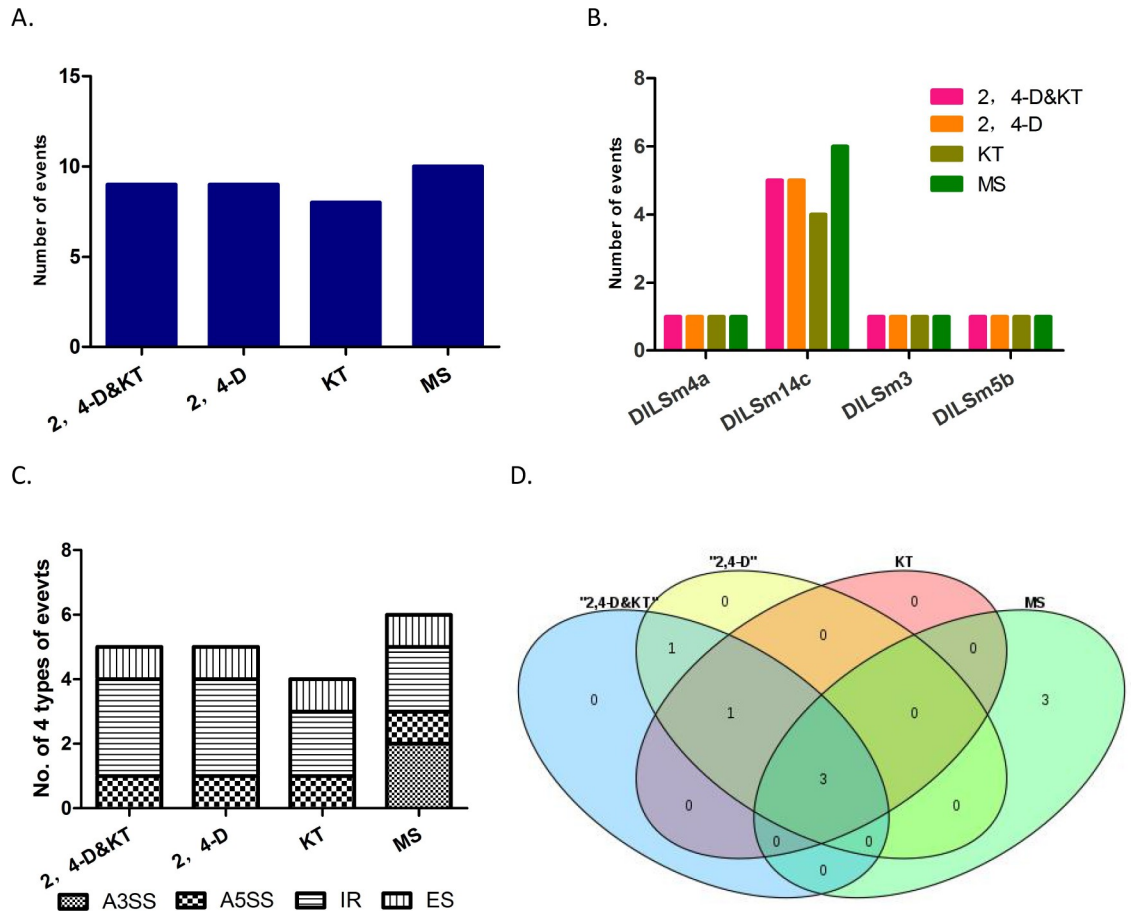
Fig 5. Distribution of four types of AS events in the *DLSm* family of longan NEC and early SE stages.

<https://doi.org/10.1371/journal.pone.0230795.g005>

specific AS event was detected under the white light treatment. All the AS events that occurred under the white light treatment also occurred under the blue light treatment. There was no change in the numbers of AS events detected in the *DLSm* genes under blue light; however, AS events were more under blue light, and less under white light. These results suggest that blue light may affect the diversity of *DLSm* splicing more strongly than white light.

### The expression profiles of the *DLSm* genes in longan NECs, early SE, different tissues and organs of longan and ECs under different hormones and light quality treatments

Expression profiling is used widely to understand gene function[45]. To investigate the expression profiles of the *DLSm* genes, we analyzed their FPKM values of *DLSm* genes in NECs, early SE (EC, ICpEC, GE) stages of 'Honghezi' cultivar SE and different tissues and organs of the 'Sijimi' cultivar were analyzed by cluster analysis. The results showed that in the NEC stage and the early SE stages (Fig 8; S5 Table), 12 *DLSm* genes (*DLSmF*, *DLSmD2b*, *DLSm1b*, *DLSmEa*, *DLSm2*, *DLSmD1*, *DLSm5b*, *DLSm5a*, *DLSm14b*, *DLSmG*, *DLSmD2a*, *DLSmD3*) were most highly expressed in the NEC, seven (*DLSmD1a*, *DLSm1a*, *DLSmB*, *DLSm14a*, *DLSm7b*, *DLSmEb*, *DLSmD1b*) were most highly expressed in the EC, *DLSm3* was most highly in the NEC and EC stages, *DLSm8* was most highly in the ICpEC, and *DLSm4c* was most highly expressed in the GE stage. Remarkably, some members of same phylogenetic group had similar expression patterns, such as *DLSm5a* and *DLSm5b*, which were highly expressed in NECs and lowly expressed in the EC, ICpEC, GE stages. The expression patterns of some group members were complementary, such as high expression of *DLSm14a* in ECs and low expression in the other three stages, high expression of *DLSm14b* in NECs and low expression in the other three stages, and high expression of *DLSm14d* and *DLSm14e* in the EC, ICpEC and GE stages, low expression in the NEC stage. The equally complementary groups were Group I (*DLSm4*) and the Group XII (*DLSmD1*). The above members, which were highly expressed in the EC, ICpEC and GE stages, may play an important role in maintaining of SE in longan.

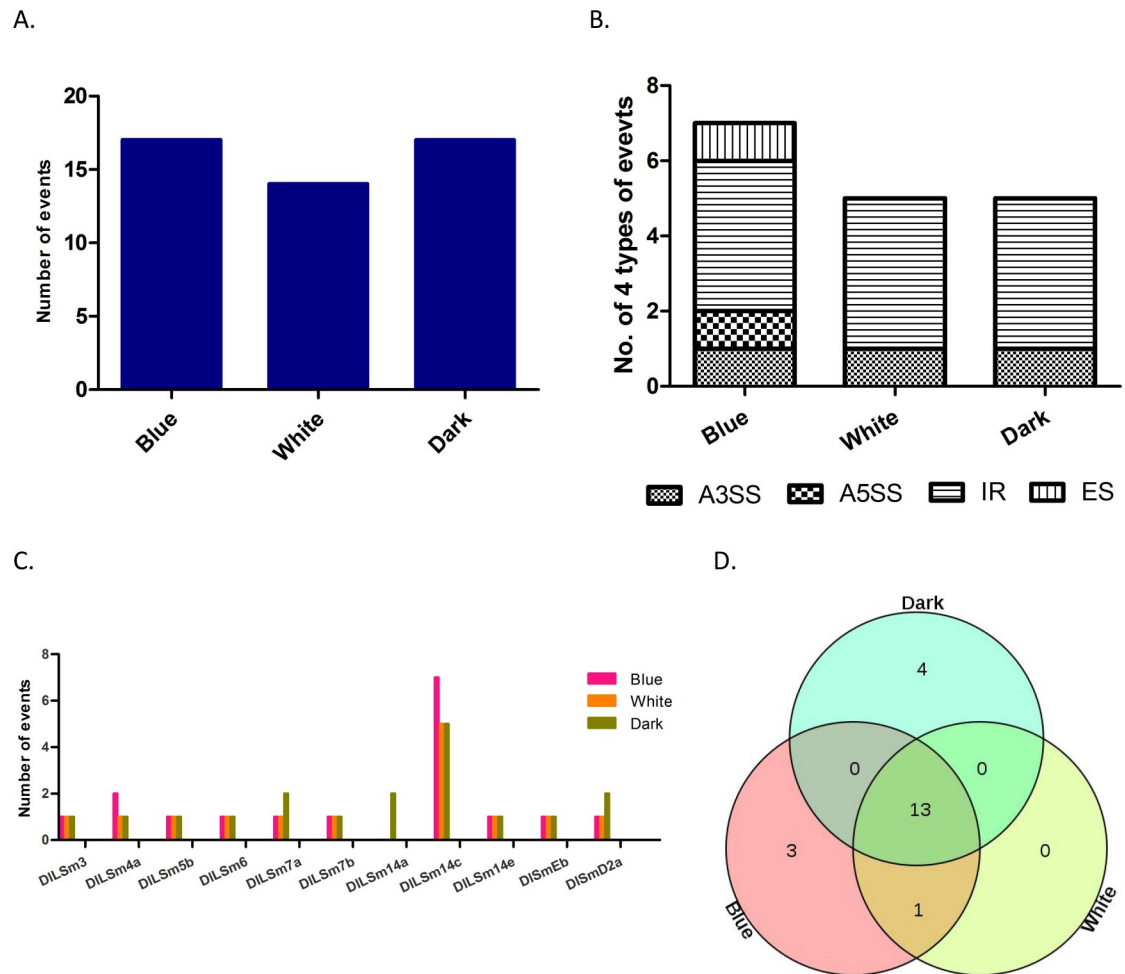


**Fig 6. AS events of *DISm* genes in longan EC stage treated with different hormones.** **A** Quantities of AS events in *DISm* family under different hormone treatments. **B** Distribution of AS events in *DISm* family under different hormone treatments. **C** The quantity and distribution of four types of AS events in *DILSm14c* under different hormone treatments. **D** Specific splicing of *DILSm14c* under different hormone treatments.

<https://doi.org/10.1371/journal.pone.0230795.g006>

Analysis of the FPKM values of the *DISm* genes in different tissues and organs of longan (Fig 9; S6 Table) showed that *DISm* genes were expressed in all the tissues tested, but their expression patterns were different. For example, *DILSm7b* was most highly expressed in flower, and *DILSm14e*, *DILSm5a*, *DISmB*, *DILSmD1* and *DILSm14c* were highly expressed in pulp, and *DILSm3*, *DILSm4b*, *DISmD1a*, *DISmD2b*, *DILSm8*, and *DISmD3* were significantly higher in roots than that in other tissues. In stem, *DILSm2* and *DILSm1b* were expressed higher than in other tissues, and *DILSm14a*, *DILSm4a*, and *DISmEa* were highly expressed in seeds. The expression patterns of the *DISm* genes in the nine tissues tested did not show any obvious similarity or complementary relationship.

*Sm* proteins are RNA binding proteins that play an important role in RNA metabolism, while plant growth regulators have an important affect the growth and induction of SE in longan. Until now, the effects of growth regulators on the *DISm* genes involved in longan SE have not been reported. Therefore, we analyzed the FPKM values of the *DISm* genes in longan EC treated with three different hormones (Fig 10; S7 Table). Under the 2,4-D, 2,4-D&KT, and KT treatments, the numbers of *DISm* genes that were highly expressed decreased. Most of the *DISm* genes had low expression under the 2,4-D, and 2,4-D&KT treatments, and *DISmD3* was highly expressed only under the 2,4-D treatments and *DILSm4b* and *DILSm14c* were highly

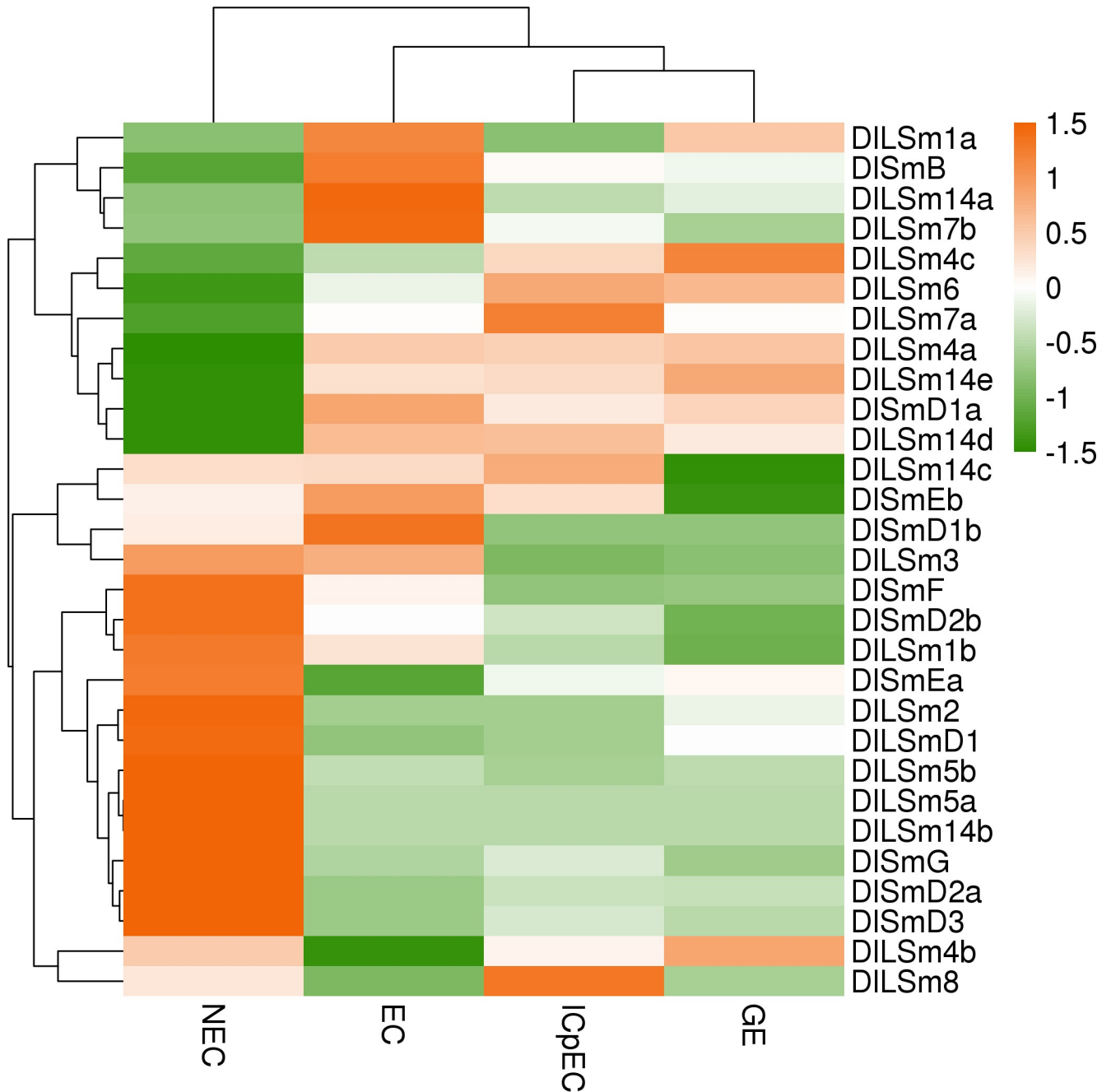


**Fig 7. AS events of *DLSm* genes in longan EC treated with different light qualities.** **A** The number of AS events in Longan EC under different light quality treatments. **B** Distribution of AS events in *DILSm14c* under different light quality treatments. **C** Number of four types AS events of *DLSm* family members under different light quality treatments. **D** Specific splicing of *DILSm14c* under different light quality treatments.

<https://doi.org/10.1371/journal.pone.0230795.g007>

expressed only under the 2,4-D&KT treatment. However, more than 50% of the genes were highly expressed under the KT treatment, and *DILSm14e* and *DILSm8* were highly expressed only under this treatment. KT can induce the differentiation of ECs, so the highly expressed *DLSm* genes that were highly expressed under this treatment may promote the occurrence of RNA events related to differentiation.

We also analyzed the FPKM values of the *DLSm* genes in longan EC stage treated with different light qualities. We found that more *DLSm* genes were highly expressed under the white light treatment than under the dark treatment, and the genes with high expression under the white light treatment are usually low expressed under the dark treatment, whereas most of the *DLSm* genes showed low expression under the blue light treatment (Fig 11; S8 Table). Among them, *DILSm14c* was highly expressed only under the blue light treatment, and *DILSm1a*, *DILSm1b*, and *DILSm14e* were highly expressed only under the white light treatment, and *DISmB*, *DILSm4b*, and *DILSm14a* showed low expression under both the white and blue light treatments. The above results showed that the expression of the *DLSm* genes were different under the different light quality treatments. In general, white light promotes the expression of



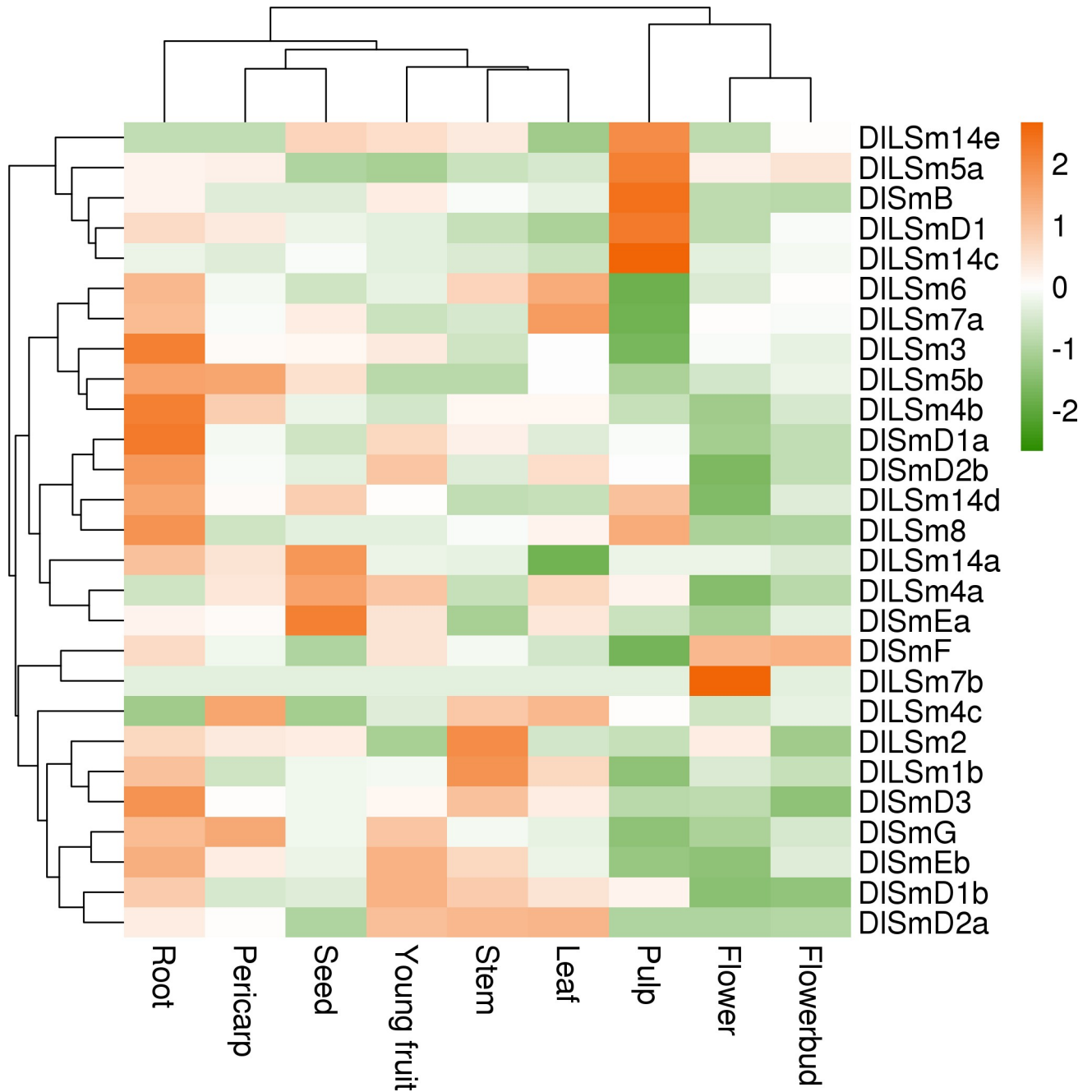
**Fig 8. The FPKM of the *DISm* family in NEC and the early SE stages of 'Honghezi' longan.**

<https://doi.org/10.1371/journal.pone.0230795.g008>

the *DISm* genes in longan ECs, whereas the blue light inhibited the expression of the *DISm* genes in the ECs.

### Verification of the expression of the *DISm* genes by qRT-PCR

In order to identify the expression of *DISm* genes in the early stage of longan SE, we selected the 28 *DISm* genes that were expressed in the early stage of SE in longan for verification by

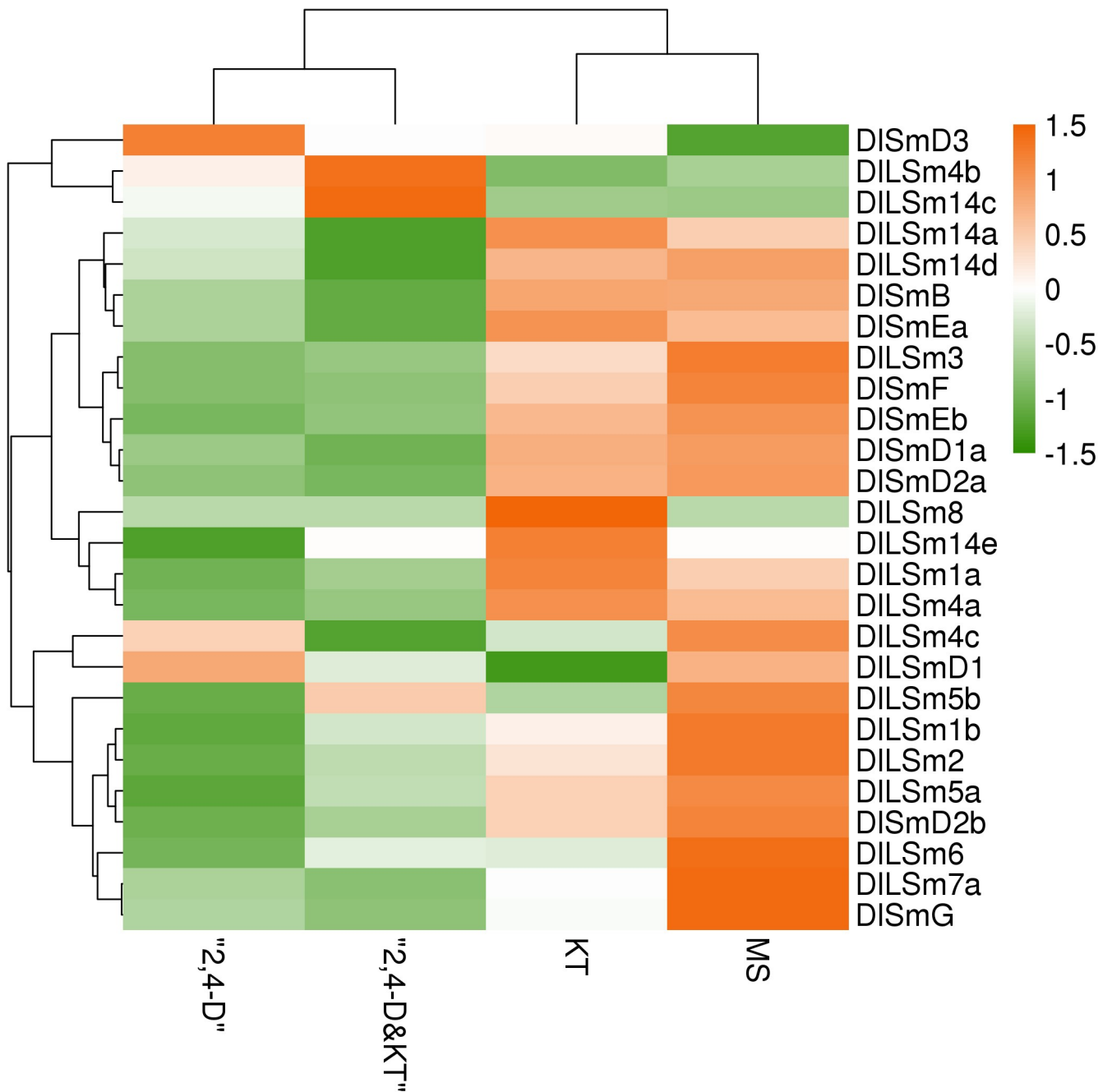


**Fig 9.** The FPKM of the *DISm* family in different tissues and organs of 'Sijimi' longan.

<https://doi.org/10.1371/journal.pone.0230795.g009>

qRT-PCR (Fig 12). The results showed that the expression trends obtained from the RNA-Seq data (FPKM values) and the qRT-PCR data at the three SE stages were the same for eight of the selected *DISm* genes; however, the expression trends were opposite for some genes including *DILSm1a*, *DILSm2*, *DILSm3*, and *DISmEa*. These discrepancies may be related to differences in the batch of the material and the growth state that were used. The expression trends obtained from the RNA-Seq data and qRT-PCR data was consistent for 15 of the *DISm* genes

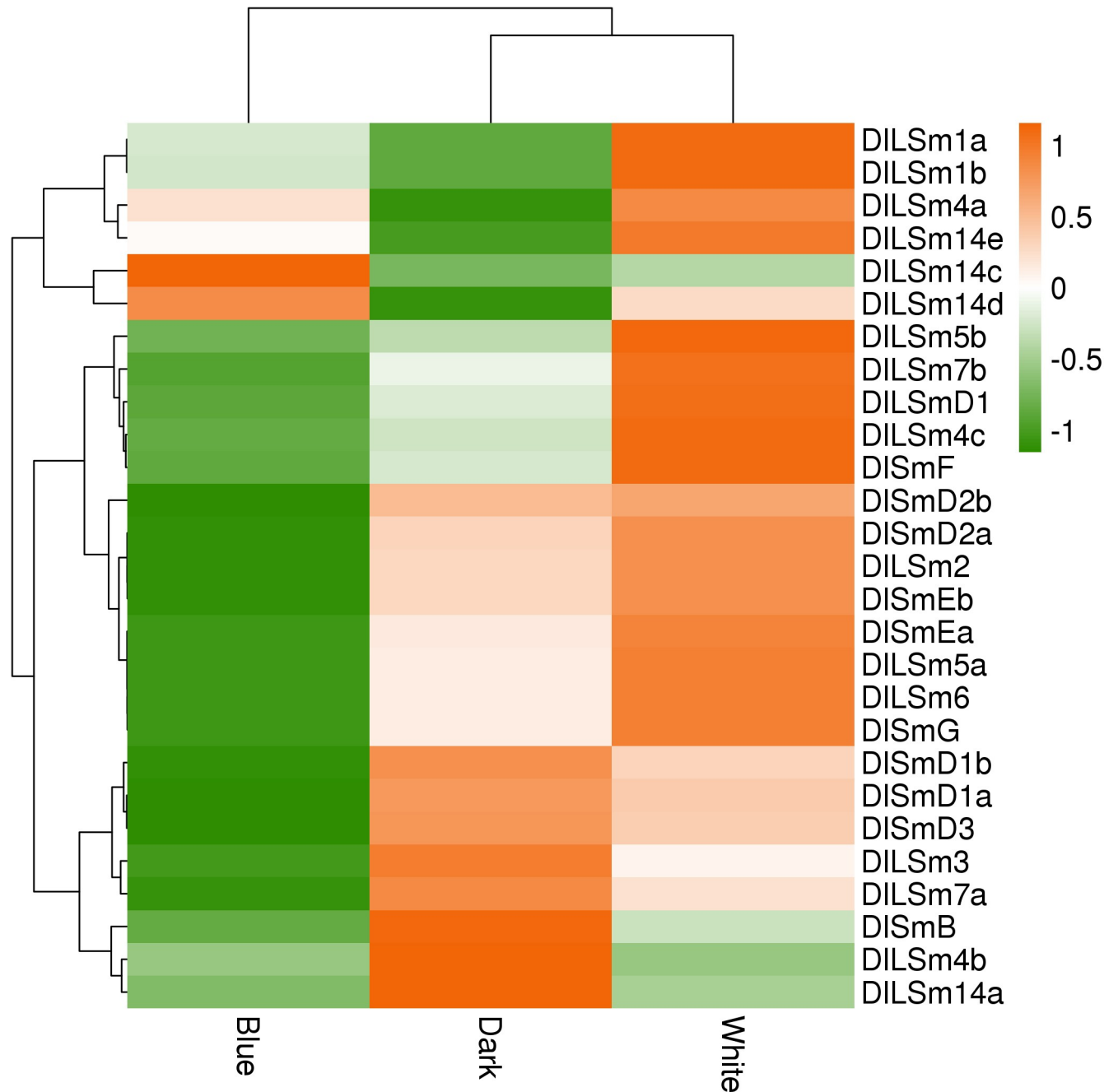




**Fig 10.** The FPKM values of the *DISm* family in longan EC stage treated with different hormones.

<https://doi.org/10.1371/journal.pone.0230795.g010>

from the EC to ICpEC stages and from the ICpEC to GE stages, respectively. However, the expression trends of *DILSm1a*, *DILSm2*, *DILSm4b*, *DILSm5a*, *DILSm14a*, *DILSm14c*, *DISmD2b*, *DISmEa*, and *DISmG* varied greatly from the EC to GE stages. The expression of 25 genes decreased from the EC to GE stages. These results showed that the expression of *DISm* genes was affected by the growth and differentiation of somatic embryos, and their expression decreased as the increase of somatic embryo differentiation progressed.



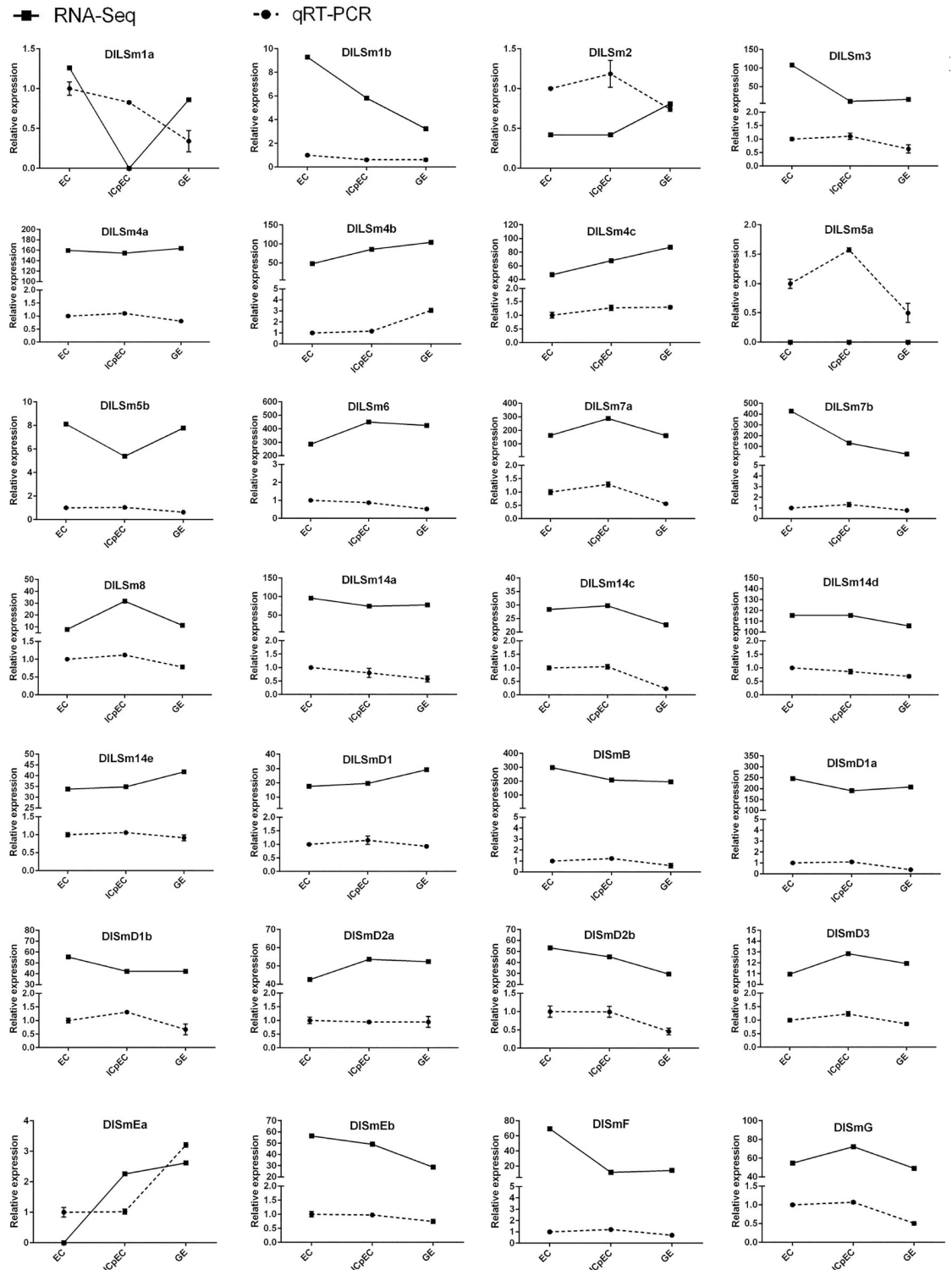
**Fig 11.** The FPKM of the *DISm* family in longan EC stage treated with different light qualities.

<https://doi.org/10.1371/journal.pone.0230795.g011>

## Discussion

### Longan *Sm* proteins are highly conserved and DILSm14 protein may have functional diversity

We identified 29 *DISm* genes of longan were identified in this study. Comparison of the deduced DISm protein sequences with homologous Sm proteins from *Arabidopsis* showed that they contained the conserved LSm domain that is typical of the LSm protein superfamily,



**Fig 12. Comparison of expression trend of the FPKM and qRT-PCR.** *EF-1α* was used as a reference gene to normalize the FPKM value and qRT-PCR expression data of *DILSm* genes. Active line section represented FPKM value; dotted line graph represented qRT-PCR results and vertical bars showed standard deviation.

<https://doi.org/10.1371/journal.pone.0230795.g012>

indicating that Sm proteins are highly conserved among different species. However, some domains varied among different members of the same phylogenetic group. For instance, DLSmB in Group XIII had one more DNA\_pol3\_delta2 superfamily domain than the two members of *Arabidopsis* proteins in this group, and the LSm14 proteins in Group IV had an extended C-terminal, and most had an FDF protein domain in the N-terminal. The FDF domain is a complete  $\alpha$ -helical domain with multiple exposed hydrophilic rings rich in glycine and charged residues, a feature that is unique to LSm 13–16[46–48]. LSm14 protein has been shown to contain three highly conserved DFDF/FFD/TFG motif boxes, which not only are involved make it participate in RNA metabolism like the other LSm proteins, but are also involved in the regulation of the mitotic G2/M phase[48–50]. We also found the DFDF/FFD/TFG motif boxes in longan LSm14 protein(S1 Fig). These results suggested that the structures of the Sm proteins were highly conserved in structure, but the specific domains in some of the proteins may contribute to their functional diversity. However, in the present study we focused on the function and structure of the small Sm and LSm proteins with no more than about 150 residues. Little is known about the large LSm domain proteins that contain additional domains[48]. The results of FPKM analysis revealed a complementary relationship in the gene expression patterns among different members of the same phylogenetic subgroup, indicating that there may be complementarities in their functions. DLSm14 proteins also had showed different expression patterns in the NEC stage and early stages of SE. These results led us to speculated that DLSm14 protein may play an important role during SE of longan; however, this needs to be verified in future studies.

### The promoters of longan *Sm* genes contain a large number of *cis*-acting elements involved in response to abiotic stress, circadian rhythm, and organ development

The growth, development and metabolism of plants are affected by many factors. Cristian et al.[17] found that the function of the LSm2–8 complex in *Arabidopsis* was controlled by external signals, and the adaptability of plants to abiotic stress was regulated differently. The level of U6 snRNP, which is closely related to the complex, changed splicing activity according to the external environment. Cui et al. [18] through a comprehensive transcriptional group analysis of *Arabidopsis thaliana*, it was proved that the variable splicing event of stress response gene was induced by SAD1 gene encoding lsm5 protein under salt stress. Seo et al.[51] overexpressed *CCA1 $\alpha$*  and *CCA1 $\beta$*  genes in *Arabidopsis thaliana* at a low temperature, which showed that the cold regulation of *CCA1* selective splicing increased the freezing tolerance of *Arabidopsis* and regulated the biological clock. AS of Sm genes can regulate plants' biological clocks and their resistance to abiotic stress by splicing. However, until now the promoters of *Sm* genes have not been analyzed. We found that the promoters of *DLSm* genes contained a large number of physiological circadian rhythm control response elements as well as abiotic stress response elements such as those related to mechanical injury, drought, low temperature, and anaerobic induction. In addition, 50% of the *DLSm* promoters contained protein metabolic response elements, which may play a role in maintaining the stability of protein metabolism.

The *DLSm* promoters also contained a large number of endosperm development response elements, but only *DLSm14a*, *DLSmEa*, and *DLSm4a* were found to be up-regulated in the seeds according to the FPKM values for the different longan tissues. SmB has been shown to be essential for stem cell regeneration, and LSm14 is related to mitosis of animal oocytes, and different mutations in SmD1a, SmD1b, SmD2a, SmD2b, and SmF were shown to kill yeast cells[49,52–54]. We found that promoters of *DLSmD2b* and *DLSmEa* responded to cell cycle regulation. The suitable concentration of 2,4-D was found to be beneficial for callus

production and inhibition of proembryo formation[55], we found the promoters of *DlSmD1a*, *DlSmB*, *DlSm14c*, and *DlSm14e* contained auxin response elements and these genes showed different expression patterns at different somatic embryo growth stages. Further, *DlSmD1a*, *DlSmB*, *DlSm14c*, and *DlSm14e* contained elements related to meristem expression. Whether these proteins also have the unique functional roles remains to be further investigated.

### **KT may affect the AS of *DlSm* genes and thus affect the differentiation of longan EC**

Alternating the culture medium containing 2, 4-D as well as that containing 2,4-D, KT and AgNO<sub>3</sub> was shown to be beneficial to the long-term maintenance of longan EC[31]. A suitable concentration of 2,4-D could maintain the embryogenic calli of longan, and also promoted ethylene and accelerated the senescence of longan EC, whereas KT promoted the growth of EC, resulting in somatic embryo differentiation, so alternating the cultures could control the negative effects of the two hormones[31, 41]. We found that compared with the MS treatment, the AS events in the longan EC decreased under the three hormone treatments, and the AS events in *DlSm14c* were different under the different treatments. We also found that the number of AS events first decreased in the early stage of SE; therefore, we speculated that the production of the specific gene variants may be related to the differentiation of longan somatic embryos induced by KT and 2,4-D.

The comparative analysis of the expression profiles of *DlSm* genes in longan SE obtained from the RNA-Seq data (FPKM values) and qRT-PCR analysis showed that their expression was affected by the growth and differentiation of longan somatic embryos. The expression of most of the genes decreased from the EC to GE stages, which we speculated may be related to cell differentiation. We found that the expression of *DlSm8* and *DlSm14e* was specific under the KT treatment, that of *DlSmD3* was specific under the 2,4-D treatment, and the expression of *DlSm4b* and *DlSm14c* were specific under the 2,4-D&KT treatment. These results suggest that the genes that were specifically expressed under the different treatments may be related to the different effects of the hormones on longan SE. The interactions among genes are very complex and are influenced by many factors, so whether the AS events produced under different hormonal treatments affect the translation of the mRNAs needs to be further verified.

### **Blue and white light may affect the growth of longan EC by affecting AS events of *DlSm* genes**

All kinds of light matter have been shown to influence the morphology, organ differentiation, and biochemical index of plant cells[56]. Different photoquality can have different effects on different plant embryos[43,57–59]. We found that AS events of the *DlSm* genes decreased under the white light treatment. Compared with the dark treatment, the blue light treatment produced five specific AS events and the white light treatment produced one specific AS event. The white light promoted the expression levels of the *DlSm* genes, whereas the blue light inhibited their expression levels. Further, and the high expression of the *DlSm* genes under the white light treatment was significantly higher than their expression under the blue light treatment. *DlSm14c* was expressed only in blue light and eight genes were expressed only in white light. Thus, the expression levels of the *DlSm* genes were promoted under the white light treatment and the AS events were reduced, whereas the expression levels of most of the genes were suppressed under the blue light treatment. These results, led us to speculate that blue light may increase the AS events in the *DlSm* genes, so their normal transcripts were not produced and their expression levels were suppressed. These two situations may trigger different

resistance mechanisms in longan because of light quality stress. Our hypothesis about the effects of the two kinds of light quality on longan EC is consistent with that of Lin Xiulian[60], who found that blue, white, red, and green light were not conducive to the growth of longan EC. However, Zeng Lilan[61] showed that blue light and green light were more beneficial to the growth of longan EC than white and red light. Our results are different from those of the previous studies, which may be related to the composition of culture medium and the wavelength used for the light quality treatments.

## Conclusions

In this study, we identified 29 *Sm* genes in 17 phylogenetic groups in longan. Most *DlSm* genes are respond to various plant hormones, such as methyl jasmonate, salicylic acid, abscisic acid, ethylene, and auxin, and some genes may regulate the development of the early SE and the tissues(seed, flower, pulp and root) of longan. The AS events were different in the four development stages, and may be related to the differentiation of longan somatic embryos. The expression profiles of the *DlSm* genes indicated their possible functional divergence. Notably, the numbers of AS events of *DlSm14* were significantly more than in the other genes both under the different hormone treatment and different photoquality treatments. The results show that the *DlSm* genes may play an important roles in SE of longan. The analysis and results reported here will help to identify candidate genes for further functional studies.

## Supporting information

**S1 Fig. Domains of DL<sub>Sm</sub>14 proteins.**

(PNG)

**S1 Table. The primers of qRT-PCR.**

(DOCX)

**S2 Table. Statistic of total AS events called from all samples in longan NEC stage and early SE stages.**

(XLSX)

**S3 Table. Statistic of total AS events called from all samples in longan EC treated with different hormones.**

(XLSX)

**S4 Table. Statistic of total AS events called from all samples in longan EC treated with different light quality.**

(XLSX)

**S5 Table. The FPKM value of NEC and early SE of 'Honghezi' cultivar.**

(XLSX)

**S6 Table. The FPKM value of the nine organs of 'Sijimi' cultivar.**

(XLSX)

**S7 Table. The FPKM value of EC under the different hormone treatments.**

(XLSX)

**S8 Table. The FPKM value of EC under the different light quality treatments.**

(XLSX)

## Acknowledgments

Thanks a lot to all the research workers in Institute of Horticultural Biotechnology for their contribution to our study.

## Author Contributions

**Conceptualization:** Zhongxiong Lai, Yuling Lin.

**Data curation:** Xue Li, Yan Chen.

**Formal analysis:** Xue Li, Yan Chen.

**Funding acquisition:** Zhongxiong Lai, Yuling Lin.

**Software:** Xue Li, Shuting Zhang, Liyao Su, Xiaoping Xu, Xiaohui Chen.

**Supervision:** Zhongxiong Lai, Yuling Lin.

**Writing – original draft:** Xue Li.

**Writing – review & editing:** Zhongxiong Lai, Yuling Lin.

## References

1. Khusial P, Plaag R, Zieve G W. LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends in Biochemical Sciences*. 2005, 30(9): 522–528. <https://doi.org/10.1016/j.tibs.2005.07.006> PMID: 16051491
2. Séraphin B. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *Embo Journal*. 1995, 14(9): 2089–2098. PMID: 7744014
3. H H, P F, V A R, K F, H H, H B, et al. snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *Embo Journal*. 1995, 14(9): 2076. PMID: 7744013
4. He W, Parker R. Functions of Lsm proteins in mRNA degradation and splicing. *Current Opinion in Cell Biology*. 2000, 12(3): 346–350. [https://doi.org/10.1016/s0955-0674\(00\)00098-3](https://doi.org/10.1016/s0955-0674(00)00098-3) PMID: 10801455
5. Kambach C, Walke S, Young R, Avis J M, Fortelle E D L, Raker V A, et al. Crystal Structures of Two Sm Protein Complexes and Their Implications for the Assembly of the Spliceosomal snRNPs. *Cell*. 1999, 96(3): 375–387. [https://doi.org/10.1016/s0092-8674\(00\)80550-4](https://doi.org/10.1016/s0092-8674(00)80550-4) PMID: 10025403
6. Will C L, Luhrmann R. Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology*. 2001, 13(3): 290–301. [https://doi.org/10.1016/s0955-0674\(00\)00211-8](https://doi.org/10.1016/s0955-0674(00)00211-8) PMID: 11343899
7. Lührmann R, Kastner B, Bach M. Structure of spliceosomal RNPs and their role in pre-mRNA splicing. *Biochimica Et Biophysica Acta*. 1990, 1087(3): 265–292. [https://doi.org/10.1016/0167-4781\(90\)90001-i](https://doi.org/10.1016/0167-4781(90)90001-i) PMID: 2147394
8. Salgado Garrido J, Bragado Nilsson E, Kandels Lewis S, Séraphin B. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *Embo Journal*. 2014, 18(12): 3451–3462.
9. Carlos P, Tamara H, Rosa L, María D M C, Julio S. LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal Arabidopsis development. *The Plant Cell*. 2012, 24(12): 4930–4947. <https://doi.org/10.1105/tpc.112.103697> PMID: 23221597
10. M F, J C R, P L. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genetics*. 1997, 16(3): 277–282. <https://doi.org/10.1038/ng0797-277> PMID: 9207794
11. Imre Törö S T C M. RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *Embo Journal*. 2014, 20(9): 2293–2303.
12. Stark M R D E A D. Dramatically reduced spliceosome in Cyanidioschyzon merolae. *Proc Natl Acad Sci U S A*. 2015, 112(11): 1191–1200.
13. Reimer K A, Stark M R, Aguilar L C, Stark S R, Burke R D, Moore J, et al. The sole LSm complex in Cyanidioschyzon merolae associates with pre-mRNA splicing and mRNA degradation factors. *Rna-a Publication of the Rna Society*. 2017, 23(6): 58116–58487.
14. Chen Y, Cao J. Comparative genomic analysis of the Sm gene family in rice and maize. *Gene*. 2014, 539(2): 238–249. <https://doi.org/10.1016/j.gene.2014.02.006> PMID: 24525402

15. Cao J, Shi F, Liu X, Jia J, Zeng J, Huang G. Genome-Wide Identification and Evolutionary Analysis of Arabidopsis Sm Genes Family. *Journal of Biomolecular Structure & Dynamics*. 2011, 28(4): 535–544.
16. Soledad P, Estefanía M, Lauren J F, Ruben Gustavo S, Ariel C, John B H, et al. Role for LSM genes in the regulation of circadian rhythms. *Proc Natl Acad Sci U S A*. 2014, 111(42): 15166–15171. <https://doi.org/10.1073/pnas.1409791111> PMID: 25288739
17. Carrasco-López C, Hernández-Verdeja T, Perea-Resca C, Abia D, Catalá R, Salinas J. Environment-dependent regulation of spliceosome activity by the LSM2-8 complex in Arabidopsis. *Nucleic Acids Research*. 2017, 45(12).
18. Cui P, Zhang S, Ding F, Ali S, Xiong L. Dynamic regulation of genome-wide pre-mRNA splicing and stress tolerance by the Sm-like protein LSM5 in Arabidopsis. *Genome Biology*. 2014, 15(1): R1. <https://doi.org/10.1186/gb-2014-15-1-r1> PMID: 24393432
19. Lin Y. Cloning of miR398a Precursor Gene from Embryogenic Callus and Its Expression Analysis During Somatic Embryogenesis in Longan. *Chinese Journal of Tropical Crops*. 2011.
20. Qinglin Z, Liyao S U, Xue L I, Shuting Z, Xiaoping X U, Xiaohui C, et al. Cloning and Expression Analysis of miR166 Primary Gene during the Early Stage of Somatic Embryogenesis of Longan. *Acta Horticulturae Sinica*. 2018.
21. Chen X, Bai Y, Chen X, Hansheng L I, Lin Y, Lai Z. Molecular Characterization and Expression Patterns of the DCL Gene Family during Somatic Embryogenesis in *Dimocarpus longan* Lour. *Acta Botanica Boreali-Occidentalia Sinica*. 2017.
22. Yang M M, Lin Y L, Wang T C, Lai Z X. Cloning of AGO6 from Embryogenic Callus and Its Expression Analysis by qPCR during Somatic Embryogenesis in *Dimocarpus longan* Lour. *Acta Agriculturae Universitatis Jiangxiensis*. 2015.
23. Hansheng L I, Yao D, Chen X, Liu W, Chen Y, Lin Y, et al. Effects of Blue Light on the Cultivation of Longan Cells and Flavonoid Metabolism. *Chinese Journal of Tropical Crops*. 2018.
24. Dong H, Zhou Y, Tian Q, Lai G, Lin Y, Lai Z. Effects of Different Light Qualities on the Contents of Flavonoids of Embryonic Callus in *Dimocarpus longan* Lour. *Chinese Journal of Tropical Crops*. 2014.
25. Wei Y E, Yuling L. Cloning and Analysis of NBS-LRR Resistance Genes from Embryogenic Callus in *Dimocarpus longan* Lour. *Chinese Journal of Tropical Crops*. 2012.
26. Lin Y L, Lai Z. Superoxide dismutase multigene family in longan somatic embryos: a comparison of CuZn-SOD, Fe-SOD, and Mn-SOD gene structure, splicing, phylogeny, and expression. *Molecular Breeding*. 2013, 32(3): 595–615.
27. Lin L, Ying Q U, Yang X U, Lin Y, Lai Z. Cloning and Expression Analysis of DIARF5a in the Process of Somatic Embryogenesis in *Dimocarpus longan* Lour. *Acta Botanica Boreali-Occidentalia Sinica*. 2014.
28. Tian Q, Lin Y, Lai Z, Wang T. Cloning and Bioinformatic Analysis of the Promoters of DIRan3A and DIRan3B from Embryogenic Callus in *Dimocarpus longan*. *Chinese Journal of Tropical Crops*. 2014.
29. Ruilian L, Yuling L, Zhongxiong L. Cloning of auxin receptor gene TIR1 and its interaction with mi R393 in and its interaction with mi R393 in *Dimocarpus longan* Lour. *Chinese Journal of Applied & Environmental Biology*. 2016, 22(01): 95–102.
30. Zhang D, Tian Q, Yang M, Lin Y, Lai Z. Cloning of DIWUS from Embryogenic Callus and Its Expression Analysis during Somatic Embryogenesis in *Dimocarpus longan* Lour. *Acta Botanica Boreali-Occidentalia Sinica*. 2015.
31. Lai Z X, Pan L Z, Chen Z G. Establishment and maintenance of longan embryogenic cell lines. *Journal of Fujian Agricultural University*. 1997.
32. L Z, C C. Changes of Endogenous Phytohormones in the Process of Somatic Embryogenesis in Longan (*Dimocarpus longan* Lour.). *Chinese Journal of Tropical Crops*. 2002, 23(2): 41–47.
33. Lai Z. Study of Longan biotechnology. 1st ed. Fuzhou: Fujian Science and Technology Press, 2003.
34. Lin Y, Min J, Lai R, Wu Z, Chen Y, Yu L, et al. Genome-wide sequencing of longan (*Dimocarpus longan*-Lour.) provides insights into molecular basis of its polyphenol-rich characteristics. *Gigascience*. 2017, 6(5): 1–14.
35. Yukun Chen X X Z L. Global scale transcriptome analysis reveals differentially expressed genes involve in early somatic embryogenesis in *Dimocarpus longan* Lour. *BMC Genomics*. 2020, 21(4).
36. Li H, Lyu Y, Chen X, Wang C, Yao D, Ni S, et al. Exploration of the Effect of Blue Light on Functional Metabolite Accumulation in Longan Embryonic Calli via RNA Sequencing. *Int J Mol Sci*. 2019, 20(2).
37. Shen S, Park J W, Lu Z, Lin L, Henry M D, Wu Y N, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. 2014, 111(51): 5593–5601.
38. Lin Y I L Z. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Science*. 2010, 178(4): 359–365.



39. Roy S W, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Reviews Genetics*. 2006, 7(3): 211–221. <https://doi.org/10.1038/nrg1807> PMID: 16485020
40. Xue-Lin H, Xiao-Ju L. The Relations of Biosynthesis of Ethlene and Polyamines to Somatic Embrogenesis of Plants. *Plant Physiological Communications*. 1995, 2(31): 81–85.
41. Lai Z X, Chen C L, Chen Z G. PROGRESS IN BIOTECHNOLOGY RESEARCH IN LONGAN. *Acta Horticulturae*. 2001, 558(558): 137–141.
42. Lin X P, Lai Z X, Huang Q. Effect of light quality on the growth, cell membrane protection enzyme activities of embryogenic calli in longan. *Journal of Fujian Agriculture & Forestry University*. 2008.
43. Seibert M, Wetherbee P J, Job D D. The Effects of Light Intensity and Spectral Quality on Growth and Shoot Initiation in Tobacco Callus. *Plant Physiology*. 1975, 56(1): 130–139. <https://doi.org/10.1104/pp.56.1.130> PMID: 16659243
44. Xiao X E. Effects of LED Light on Growth and Polysaccharide Accumulation of Embryogenic Callus in *Dimocarpus longan* Lour. *Horticulture & Seed*. 2016.
45. Chen Y, Hao X, Cao J. Small auxin upregulated RNA (SAUR) gene family in maize: Identification, evolution, and its phylogenetic comparison with Arabidopsis, rice, and sorghum. *Journal of Integrative Plant Biology*. 2014, 56: 133–150. <https://doi.org/10.1111/jipb.12127> PMID: 24472286
46. Tritschler F, Braun J E, Eulalio A, Truffault V, Izaurralde E, Weichenrieder O. Structural Basis for the Mutually Exclusive Anchoring of P Body Components EDC3 and Tral to the DEAD Box Protein DDX6/Me31B. *Molecular Cell*. 2009, 33(5): 661–668. <https://doi.org/10.1016/j.molcel.2009.02.014> PMID: 19285948
47. Felix T, Ana E, Vincent T, Hartmann M D, Sigrun H, Steffen S, et al. A divergent Sm fold in EDC3 proteins mediates DCP1 binding and P-body targeting. *Molecular & Cellular Biology*. 2007, 27(24): 8600.
48. Albrecht M, Lengauer T. Novel Sm-like proteins with long C-terminal tails and associated methyltransferases ☆. *Febs Letters*. 2004, 569(1–3): 18–26. <https://doi.org/10.1016/j.febslet.2004.03.126> PMID: 15225602
49. Zhang T, Yuanyuan L I, Hui L I, Xue-Shan M A, Ouyang Y C, Hou Y, et al. RNA-associated protein LSM family member 14 controls oocyte meiotic maturation through regulating mRNA pools. *Journal of Reproduction & Development*. 2017, 63(4): 383–388.
50. Lodomery M, Sommerville J. The Scd6/Lsm14 protein xRAPB has properties different from RAP55 in selecting mRNA for early translation or intracellular distribution in *Xenopus* oocytes. *BBA—Gene Regulatory Mechanisms*. 2015, 1849(11): 1363–1373. <https://doi.org/10.1016/j.bbagr.2015.10.002> PMID: 26455898
51. Pil Joon S, Mi-Jeong P, Mi-Hye L, Sang-Gyu K, Minyoung L, Baldwin I T, et al. A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis. *Plant Cell*. 2012, 24(6): 2427–2442. <https://doi.org/10.1105/tpc.112.098723> PMID: 22715042
52. Schwer B, Roth A J, Shuman S. Will the circle be unbroken: specific mutations in the yeast Sm protein ring expose a requirement for assembly factor Brr1, a homolog of Gemin2. *Rna*. 2016, 23(3): 420–430.
53. Mili D, Georgesse D, Kenani A. Localization and role of RAP55/LSM14 in HeLa cells: a new finding on the mitotic spindle assembly. *Acta Biochimica Polonica*. 2015, 62(3): 613. [https://doi.org/10.18388/abp.2015\\_1107](https://doi.org/10.18388/abp.2015_1107) PMID: 26339800
54. Enrique F T, S Ren M, Dagmar Z, Martin S, Sch Ler H R, Emili S, et al. Smed-SmB, a member of the LSm protein superfamily, is essential for chromatoid body organization and planarian stem cell proliferation. *Development*. 2010, 137(7): 1055–1065. <https://doi.org/10.1242/dev.042564> PMID: 20215344
55. Lai Z, Chen Z. Somatic embryogenesis of high frequency from longan embryogenic calli. *Journal of Fujian Agricultural University*. 1997.
56. Wang W. Somatic embryogenesis in *Dendrobium candidum* wall ex Lindl. by different light. Chongqing: Southwest Jiaotong University, 2008.
57. Zhang J C. Studies on the Effects of the Induction and Proliferation of Callus from Ginkgo biloba L. *Journal of Chongqing Teachers College*. 1999.
58. Liang J X, Chen B. Influence of Light Quality on Seedling Differentiation from Sugarcane Callus. *Sugar Crops of China*. 2006.
59. Hong S R. Effect of Light Quality on the Callus Inducement of *Pueraria lobata* Leaf and its Content of Soluble Protein. *Journal of Anhui Agricultural Sciences*. 2007.
60. Xiulian L. Studies on the minimal growth conservation in vitro of embryonic calli of longan germplasm and its physiological and genetic mechanism. Fujian: Fujian Agriculture and Forestry University, 2009.
61. Lilan Z. Characterization of SOD promoter and expression analysis of SOD genes of the embryogenic callus in *Dimocarpus longan* Lour. Fujian: Fujian Agriculture and Forestry University, 2013.