Genome-wide identification and expression profiling of microRNAs genes during SE and in different organs/tissues in Dimocarpus longan Lour.

Yuling Lin
Fujian Agriculture and Forestry University
Yan Chen
Fujian Agriculture and Forestry University
Youjing Zeng
Fujian Agriculture and Forestry University
Shuting Zhang
Fujian Agriculture and Forestry University
Zihao Zhang
Fujian Agriculture and Forestry University
Jiawei Gong
Fujian Agriculture and Forestry University
Zhongxiong Lai (laizx01@163.com)

Research article

Keywords: Dimocarpus longan, somatic embryogenesis, pri-miRNA, pre-miRNA, qRT-PCR

Posted Date: November 22nd, 2019

DOI: https://doi.org/10.21203/rs.2.17616/v1

Abstract

Background MicroRNAs (miRNAs) is a conserved class of single-stranded noncoding RNAs, which regulate growth and development, hormone response, and stress responses in plant. Longan (Dimocarpus longan Lour.) is an important fruit tree with medicinal value and considerable economic value. However, there is no research on genome-wide identification of miRNAs genes in longan.

Results Here, we conducted a genome-wide survey of longan miRNA genes, characterizing their structure and expression. A total of 88 miRNA precursors comprised of 30 known miRNA families and three undefined miRNA families were obtained. Of the 30 known longan miRNA families, 16 were represented by multiple loci in longan genome, which were also found more diverse in genome loci numbers in other plants; and four of them, involving miR395, -399, -477, and -3627 families, were identified as miRNA clusters genes; Among the 16 miRNA families, the conserved region as mature miRNA, as would be expected, and the loop region sequences of miRNA genes were divergent. Further, amplification and characterization of the longan pri-miRNA genes were carried out, and the upstream transcription 5 'region of 19 miRNAs and the downstream transcription 3' region of 12 miRNAs with typical poly (a) tails were obtained, showing that longan miRNAs genes possess multiple transcription start sites (TSSs) or polymorphism trait, suggesting pri-miRNA has similar transcription specialty as protein-encoding genes. Subsequently, promoter cis-element analysis on longan miRNA genes showed that the top three most abundant cis-elements were light-responsive elements, stress- responsive elements, and hormone responsive elements. Quantitative real-time PCR showed that pre-miRNAs were differentially expressed during the early stages of longan somatic embryogenesis, and the most number of highly expressed pre-miRNAs were detected in anther samples in different organs/tissues types of longan,

Conclusion This is the first report concerning genome-wide identification and expression profiling of microRNAs genes, and will serve as a foundation for future research into the functional roles of miRNA genes in longan.

Background

MicroRNAs (miRNAs) is a conserved class of single-stranded noncoding RNAs of 20–24 nt in length, derived from primary miRNA that form characteristic stem-loop structures. These hairpin RNAs, which are known as pre-miRNAs. In plant, pre-miRNAs were transcribed by RNA polymerase II, and processed by DICER-LIKE1 to produce miRNAs[1, 2]. The mature miRNAs negatively regulate mRNA expression by cleavage of targeted mRNA by perfectly complementary recognition of target mRNA sites[3], or near perfect complementarity to recognize target mRNA sites to inhibit translation[4]. miRNAs have been found to regulate growth and development[5], hormone response[6] and stress responses[7] in plant. The first MiRNA was discovered in *Caenorhabditis elegans* in 1993[8]. In plants, the first miRNA was cloned in Arabidopsis until 2002[9]. With the development of plant miRNA research, more and more studies have identified miRNA genes. Lifang et al. performed the first comprehensive characterization of maize pri-miRNA transcripts, and amplified 68 miRNA transcript precursors from 18 conserved families by PCR-

RACE [10]; Barvkar et al. analyzed the length distribution of pri-miRNA in flax and the expression of pre-miRNA in different tissue types[11]. The promoter is an important entry site for the regulation of gene transcription levels, controlling the time of gene expression and the intensity of expression. A crucial component in the analysis of a miRNA promoter region is the accurate identification of the transcription start site (TSS). In Arabidopsis, the core priming elements, TSS and TATA boxes of 52 miRNA genes (such as MIR156, 157, 159, 160, 166, 169, and 172) were identified by 5' RACE[12]. Han et al. systematically predicted the core promoter elements (TSS and TATA boxes) of 440 miRNAs in soybean and analyzed the *cis*-acting elements of 369 miRNAs[13].

Somatic embryogenesis (SE) is a process in which plant somatic cells are not subjected to fertilization in vitro culture, and maintain cell pluripotency even under conditions of highly differentiated cells[14]. In recent years, there were several reports have been published on the role of miRNAs in controlling SE such as Arabidopsis[15], citrus[16], cotton[17], lilium[18], maize[19], longan[20], etc. The same miRNA is involved in different stages of SE in different plants. For example, miR166, a highly conserved miRNA family, played a major role in controlling SE induction in Arabidopsis[21] and might play regulatory roles during cotyledon embryo (CE) development in larch[22] and citrus[23]. Furthermore, multiple miRNAs collectively regulate the same stage of SE. As reported in citrus, miR159, 164, 390 and 397 were related to globular embryo formation, and miR166, 167 and 398 were required for cotyledon embryo morphogenesis[23].

Longan (*Dimocarpus longan* Lour.) is an important fruit tree with medicinal value and considerable economic value. Longan embryo development can affect seed size, fruit quality and yield. The longan SE system established by Lai et al.[24] is considered to be an excellent model system for woody plant SE, laying a foundation for longan SE research. Recently, our laboratory has conducted systematic research on miRNAs in longan SE. In 2013, Lin et al. first reported the study of miRNAs and their targets in longan SE. A total of 169 miRNA families were identified, including 643 conserved and 29 novel miRNAs (including star strands)[20]. Among them, the evolutionary characteristics and expression of miR397[25], miR171[26], miR166[27], miR172[28], miR159 [29] family were analyzed. Moreover, the precursor of miR398a[30] and miR398b[31] and the primary of miR166(Zhang et al., 2018), miR156a–1, miR156a–2, miR166a, and miR397a[32] have been cloned. Studies have shown that auxin can induce SE, while miR167[33], miR390[34] and miR160[35] affect the growth and development of longan somatic embryos by regulating auxin response factors. Previous studies on longan miRNAs focused mainly on molecular properties, evolutionary rules, and functional studies, without focusing on systematically identified miRNA genes.

In this study, we performed genome-wide identification of longan miRNAs based on longan genomic data[36] and miRNA databases[20]. A total of 88 miRNA precursors were screened by searching longan small RNA datasets, and their structures and sizes were analyzed. Then, the number distribution of 16 multiple locus known miRNA genes was compared with longan and other nine plants (Arabidopsis, grape, orange, apple, peach, tomato, papaya, populus and rice), and the precursor sequences and conservation of 16 miRNA genes were analyzed. Finally, the longan pri-miRNAs were amplified and identified by

experiments. We used RLM-RACE and SMART-RACE to determine the transcription start sites (TSSs) of miRNAs in longan and analyzed its promoter *cis*-elements. The expression of 16 pre-miRNA genes in early longan SE and various tissues was determined by quantitative real-time PCR(qRT-PCR). The results enhance our understanding of miRNAs and lay a good foundation for studying the functions of different miRNA members in longan.

Results

Identification of miRNA precursors in the longan tree

To identify the miRNAs precursors in the longan tree, the longan small RNA datasets were searched and potential pre-miRNAs were checked by using BLAST against the miRBase server (Release 21). A total of 102 miRNA precursors were obtained. After removing duplicated sequences, the remaining 88 miRNA precursors, were comprised of 30 known miRNA families and three undefined miRNA families (miR2621, –2864, and –3954). The length of longan pre-miRNAs varied from 77 to 372 nucleotides with an average of 148.37 nt. Of these pre-miRNAs, 21.6 % were between 77 and 99 nt in length, 56.8 % have 100–178 nucleotides, 14.8 % were 200 to 284 nucleotides, few (6.8 %) were 333 to 372 nt in length.

In addition, we observed that longan pre-miRNAs have high negative folding free energies (ΔG) of between –18.5 (miR1511) and –185 kcal/mol (miR773) with an average of –51.7 kcal/mol according to MFOLD. In this study, we also found that most of longan pre-miRNAs have this similar feature. However, for some longan pre-miRNAs, we found that their ΔG values also related their secondary structures. The more complicated stem-loop hairpin structures of pre-miRNAs, their ΔG values have much higher. Such as miR159a_scaffold101 (257 nt, –53.5 kcal/mol), miR1511_scaffold142 (97 nt, –18.5 kcal/mol), miR2621_scaffold75(363 nt, –70.4 kcal/mol), miR3954_scafffold338 (94 nt, –24.9 kcal/mol), and miR6140_scaffold597(171 nt, –34.5 kcal/mol). All of these mentioned above miRNAs has complicated stem-loop hairpin structures (Figure 1). Also, the simpler structures of pre-miRNAs have lower ΔG values, such as miR773(333nt, –185 kcal/mol).

Uneven distribution of 16 miRNA genes in ten selected plants

Of the 30 known longan miRNA families, 14 of these, such as miR162, –168, –169, –396, –397, –530, –535, –773, –827, –862, –2118, –5225, –6140 and –7493, were encoded by single copy miRNA genes, whereas other 16 were represented by multiple loci in longan genome. To further know the distribution of miRNAs in different family in plants, the comparison analysis of the number distribution of 16 multiple loci miRNA genes between longan and other nine plants (Arabidopsis, grape, orange, apple, peach, tomato, papaya, populous, and rice) was carried out. Comparison analysis showed that the 16 miRNA genes showed up and down variations in genome loci numbers among the ten plants tested (Table 1). Of 16 miRNA families, we found that miR156/157 family has the most genome loci numbers in apple (31 loci numbers), then followed by miR395 family in rice (25), miR166 family in populus (17), miR171 in apple (15), miR399 in peach (14), miR159/miR319 in populus (14), and miR167 in apple/rice (10); while

these miRNAs has six, five, seven, eight, three, 12 and four genome loci numbers in longan genome, respectively. These results further suggest that plant miRNA precursors were also more diverse in genome loci numbers, this may be caused by the different genome size of different plants. Moreover, eight of 16 miRNA genes have more genome loci numbers in apple than that of in the other selected fruits, possibly as a result of a whole-genome wide duplication event in apple[37, 38]. This uneven distribution of miRNAs among the various miRNA families in different plants further suggests that different miRNAs families and different members may have different evolutionary history[39].

MiRNA gene clusters in longan

To analyze the miRNA gene cluster in longan, we analyzed 33 miRNA familes. The result showed that miR162, –168, –169, –396, –397, –530, –535, –773, –827, –862, –2118, –2621, –2864.2, –3954, –5225, –6140, and –7493 familes were only one member. The remaining miR156, –159, –160, –166, –167, –171, –390, –393, –394, –395, –398, –399, –477, –482, –1511, and –3627 families exist multiple members. Previous study defined 10 kb as the maximum inter-miRNA distance for two miRNA genes to be considered as clustered[40]. According to this rule, in this study, four potential miRNA clusters genes, involving miR395, –399, –477, and –3627 families, were also identified in longan (Table 2). All of miR395 members all located on scaffold6; the closest distance of (scaffold6: 2146151:2146267:+) and miR395d–5p (scaffold6: 2145212:2145296:-) is 855 nt, and the fairest distance among them is 51,585 nt. Two of three miR399 members on scaffold 167 were 5,095 nt separating adjacent genes. Two members of the miR477 family (miR473 and –477c) on scaffold192 were separated by only 110 nt, and miR477 clusters. Another cluster contained three miR3627 members on scaffold24, the closest distance of them were separated by 4,485 nt. While in our study, miR166 clusters were not detected in longan genome. The two miR167 members on scaffold53, were separated by less than 14 kb, may be a potential cluster in longan genome.

MiRNA precursors sequence alignments and conservation analysis in longan

To further know the conservation and diversity of longan miRNA family members, their similarities were observed by aligning in this study. Here, the similarity of selected 16 miRNA families with multiple members in longan were aligned by DNAMAN6.0. In this study, alignment of the 16 miRNA families, revealed that the nucleotide sequences similarity was high from different members of the same miRNA family, such as miR1511 family (82.68 %), miR319 family (77.68 %), miR160 family (76.24 %), miR482 family (65.04 %), and miR167 family (64.76 %); whereas, less similarity was observed for miR170 (49.70 %), miR393 (41.03 %), miR477 (35.71 %), miR394 (30.34 %), miR159 (33.28 %), and miR166 (26.77 %) families; and the remaining similarities from different members of the same miRNA family varied from 50% (miR399 family) to 63.03 % (miR398 family)(Figure 2). Overall, different members of the same miRNA family has highly conserved in the region of mature miRNA, but less conserved in the remaining region.

The web-based tool RNALogo graph, which creates a new graphical representation of the patterns in a multiple RNA sequence alignment with a consensus structure, also indicates the conserved region as mature miRNA, as would be expected, and the loop region sequences of miRNA genes were divergent (Figure 3).

Amplification and characterization of the longan pri-miRNA genes

Similar to protein-encoding genes, pri-miRNA transcripts were known to have features typical of transcription by RNA polymerase II, contain cap structures, poly(A) tails, and intron splicing[41]. Previously studies we have obtained the full length of pri-miR160 transcript (GenBank accession KJ372214) [35], and the upstream transcribed regions (5’ region) of pri-miR166 (scaffold53)[42], pri-miR167 (KJ372215)[33] and pri-miR390 (KJ372216) [34]. In this study, to further comprehensive characterize the full length of pri-miRNA transcripts and determined their structures in the longan genome, we designed gene-specific primers and conducted 5’RACE and 3’RACE using a template containing mixed tissues of NEC, ICpEC, and GE according to these pre-miRNAs information extracted from longan small RNA data. With the exception of miR160, –167, and –390, in addition, the upstream transcribed 5’ regions of 19 miRNAs, and the downstream transcribed 3’ regions of 12 miRNAs with the typical poly(A) tails were obtained. Among these, the presences of the full-length transcripts of 13 miRNAs were confirmed. This result indicated that the levels of successful amplified miRNA genes were high enough to be detected, whereas the levels of the other un-amplified miRNAs were not high enough in the tested samples, suggesting they maybe expressed in specific tissue/cell types, developmental stages, or environmental conditions, which need further experimental test.

In our study, the pri-miRNAs varied in size with the longest transcripts found for miR156 (scaffold29) (881 nt) and –156 (Unigene5849) (828 nt) and the shortest pri-miRNA for miR156 (scaffold38) (309 nt) for the 24 analyzed longan miRNA genes. GC content of miRNAs varied drastically in a range of 20~80 %, resulting in a diverse surface property of the molecule (Effect of microRNA GC Content on RNA Purification Efficiency). In our study, the GC content for the full lengths of pri-miRNAs in longan ranged from 32.06 (miR170_scaffold1266) to 50.4 % (miR398a_FJ973472), and most of pri-miRNAs have low GC contents in a range of 30 ~39 %, suggesting that pri-miRNAs tend to be transcribed with GC-low and AU-rich RNAs features.

In longan, the 3’ region of pri-miRNAs (from the stem-loop to the transcriptional stop site) is generally longer than the 5’ region (from the transcriptional start site (TSS) to the stem-loop). Among them, eight of them possess longer 3’ regions than the 5’ regions, and four has longer 5’ regions than the 3’ regions, such as pri-miR156 (scaffold38), pri-miR160 (KJ372214), pri-miR390a (scaffold24), and pri-miR396 (scaffold38). This result implies that the mature miRNAs tend to locate nearly to the 5’ region of pri-miRNAs, and this characteristic maybe conserved in plants. The mature miRNA location and length distribution of 3’ regions and the 5’ regions in pri-miRNAs were showed in Table2. Moreover, we observed that the full lengths of longan pri-miRNAs usually possess higher ΔG values than those from pre-

miRNAs(Figure S1), which varied from −194.2 (pri-miR156_scaffold29) to 87.9 kcal/mol (pri-miR156_scaffold38), with an average of −125.05 kcal/mol. Overall, the ΔG values of the pri-miRNAs were positively corrected with their RNA/DNA sequence length.

Multiple transcription start sites (TSSs) and 3'UTR polymorphism of longan miRNAs genes

The accurate identification of the TSSs is a crucial component in the analysis of a miRNA promoter region[43]. However, most TSSs of miRNA in plants were predicted by bioinformatics method without verifying by experiments[44]. In previously studies, we also identified TSSs for miR160[45], miR166_scaffold 53(Zhang et al., 2018), miR167[33] and miR390[46] by RLM-RACE technology in longan. Here, to get more accurate information of miRNA TSSs in longan, TSSs of miRNA genes in longan were also determined using RLM-RACE and SMART-RACE analysis, which were sensitive and accurate methods for obtaining 5'/3' ends of the full-length cDNA[47]. Of the available 22 miRNA gene transcripts with 5' regions information in longan, 16 had one TSS (Figure S2), and the remaining six, such as miR156(scaffold29), miR156(Unigene5849), miR168 (scaffold2621), miR170(scaffold537), miR319(Unigene13717), and miR2118a(scaffold26), had at least two TSSs (Figure 4). Among the 16 miRNAs with one TSS, the first nucleotide bias (TSS base composition) analysis indicated that A was the most frequently in longan miRNA genes.

Among the multiple TSSs of longan miRNA genes, miR156(scaffold29) has five TSSs that were 123 nt, 11nt, 5nt, and 1 nt apart, and its TSSs base compositions were G, C, C, G, and A; miR156(Unigene5849) has two TSSs with G and A, which were 5 nt apart; miR168 (scaffold2621) also has two TSSs with A, which were 44 nt apart; miR170 (scaffold537) contains two TSSs with G and C that were 314 nt apart; miR319(Unigene13717) possess three TSSs with A, G, and G, which were 35 nt and 220 nt apart; miR2118a(scaffold26) has three TSSs with A, A, and G, which were 4 nt and 118 nt apart. Apparently, the alternative TSS is a common phenomenon for pri-miRNAs in plants, and the multiple TSSs base compositions of miRNAs tend to be A, then followed by G, and C.

In addition, Of the available 12 miRNA gene transcripts with 3' regions information in longan, five of them have 3'UTR polymorphism trait, such as miR395(scaffold6), miR396(scaffold38), miR398a (FJ973472), miR398b(scaffold58), and miR408(Unigene40210). These miRNAs all contain two stop sites, and the distance between two stop sites varied from 22 nt to 123 nt. In maize, the distance between various stop sites can be up to several hundred nucleotides[10]. 3'UTR polymorphism trait of longan miRNA suggests a coordinated mechanism for regulation of gene expression. Actually, those miRNAs with one TSSs or one stop site in our study might possess more TSSs or polymorphism trait, which limited by the 5' or 3' RACE products.

Cis-acting element analysis of longan miRNA promoters

To understand of transcriptional regulation, and the function of miRNAs in response to environmental stress, the investigation of miRNA promoters is necessarily. Here, the identified TSSs of miRNA genes in

longan enabled us to precisely find out the proximal promoter region for each of the 22 miRNA genes. In previously studies, the 2000-bp sequence upstream region[48], 1000-bp sequence upstream region[49], or 800-bp sequence upstream region[43] from the TSSs as approximation for miRNA promoters to comprehensively identify putative promoter motifs. Here, to get more information of the putative promoter motifs of longan miRNA, we analyzed 2.5 kb upstream region from the TSSs of the 19 miRNA genes using the PlantCARE database. The sequences of 3 miRNAs could not be used for the promoter analysis because they lacked the corresponding genomic sequences (lus-MIR319b, lus-MIR396band lus-MIR828a). In this study, the presence of conserved TATA boxes was revealed in all 19 selected longan miRNA genes(data no shown), indicating that the selected longan miRNA genes have the same promoter characteristics as the protein-coding genes.

More than 1, 252 promoter motifs were identified in the 19 miRNA genes except TATA and CAAT boxes. Among these *cis*-acting motifs listed in figure 5, the light-responsive elements were the most abundant and occurred 587 times among the 19 sequences. Of them, miR156(scaffold29) contained the most abundant light-responsive elements (41), followed by miR393h(scaffold978) and miR156 (scaffold38). The second abundant elements were stress-responsive elements (221), including MYB binding site involved in drought-inducibility (MBS, 36 elements), AT-rich DNA binding protein (ATBP–1) binding site (AT-rich element, 11), elicitor-responsive elements (Box-W1/EIRE/ELI-box, 34), anaerobic induction elements (ARE, 52), heat stress responsiveness elements (HSE, 37), low-temperature responsiveness elements (LTR, 11), defense and stress responsiveness elements (TC-rich repeats, 36), wound-responsive elements (WUN-motif, 4). It is worth noting that the anaerobic induction elements were present in 18 miRNA promoters, but not existed in miR398b (scaffold58); in addition, wound-responsive element is only present in miR168(scaffold2621), miR170(scaffold1266), miR395(scaffold6), and miR397a (scaffold21). The third abundant elements were hormone responsive elements (207 elements), involved in auxin responsiveness (AuxRR-core, 5), ABA responsiveness (ABRE /CE1, 29), gibberellin-responsive elements (GARE-motif/TATC-box, 36), salicylic acid responsiveness elements (TCA-element, 32), MeJA-responsiveness elements (CGTCA-motif / TGACG-motif/ TGA-elements, 94), and ethylene-responsive elements (ERE, 11). Among of them, MeJA-responsiveness elements were the most abundant, detected in 18 MIR genes for 94 times except miR482(scaffold26), and appeared 10 times in miR156(scaffold38); gibberellin- responsive elements were the second most abundant (36), detected in each of the 19 MIR genes, occurred 4 times in miR166(scaffold338) and miR398b (scaffold58), respectively; the auxin responsiveness elements were the least detected, only occurred 1 time in each of miR156(scaffold29), miR390((KJ372216), miR393h (scaffold978), miR397a (scaffold21), and miR482 (scaffold26), suggesting these miRNAs might be responsive to auxin. The fourth abundant elements that were involved in the growth-specific and regulation of response to circadian control (50), reached 183 times in 19 miRNA genes, such as seed-specific regulation element (RY-element, 3), M-phase-specific element (MSA-like, 1), root-specific element (as1, 1), and elements required for the vascular expression of the *PAL2 (*AC-I, 3), zein metabolism regulation (O2-site, 18), endosperm expression (Skn–1_motif, 99), meristem expression (CAT-box, 8). Of these, elements involved in circadian control and endosperm expression were dectected in each miRNA genes; the M-phase-specific element in miR397a(scaffold21) and root-specific

element in miR167(KJ372215) only appeared one time, suggesting that the two miRNAs have specific role in longan; in addition, seed-specific element was only detected one time in each of miR156 (scaffold38), miR160 (KJ372214), and miR170(scaffold1266), implying these miRNAs might be have special roles in longan seed development.

Validation of distinct expression patterns of pre-miRNAs during longan early SE

there is little verification of the expression level of miRNA precursors, and there is no systematic verification of the expression of miRNA precursors in longan. To validate the expression levels and roles of pre-miRNA genes during longan early SE, we selected the pre-miRNAs corresponding to the 18 pri-miRNAs cloned, of which a total of 19 pre-miRNAs were verified by qRT-PCR. Except for pre-miR156-scaffold29, pre-miR156e*-scaffold3, and pre-miR167d-scaffold53, the remaining 16 pre-miRNAs were detected in early SE of longan.

The expression patterns of the 16 pre-miRNAs were divided into three categories according to the timing of their highest expression level. In category I, five pre-miRNAs exhibited high levels in the EC stage, but low levels in the other three stages (Figure 6a). Pre-miR2118a-scaffold26 and pre-miR168a-scaffold2621 showed an increase in expression from the ICpEC to GE stage, while the expression levels of pre-miR396a-scaffold38 decreased. Furthermore, the expression levels of pre-miR170-scaffold1266 and pre-miR394a-scaffold3884 increased at the ICpEC to CpECGE stage, but decreased at the CpECGE to GE stage. The five pre-miRNAs could be related to maintaining the pre-embryonic status of the EC.

In category II, six pre-miRNAs had the highest expression in the ICpEC stage or CpECGE stage. As shown in Figure 6b, pre-miR166-scaffold338, pre-miR171f-scaffold537, pre-miR393h-scaffold978, pre-miR390a-scaffold73, and pre-miR397-scaffold21 were highly expressed in the ICpEC stage, while pre-miR395a-scaffold6 was in the CpECGE stage. Among them, the expression levels of pre-miR393h-scaffold978 and pre-miR397-scaffold21 increased from CpECGE to GE, and the remaining four pre-miRNAs decreased.

In category III, the expression of the 5 pre-miRNAs was highest in the GE stage (Figure 6c), and the expression level was rapidly increased in the CpECGE to GE stage. The expressions of pre-miR398b-scaffold58 and pre-miR482a-scaffold26 decreased gradually from the EC to CpECGE stage, while that of pre-miR160a-scaffold209 did not change much at the EC and CpECGE stages. In summary, there were differences in the expression of pre-miRNAs in the longan early SE.

Expression levels of pre-miRNAs across tissue types

To better understand the expression of pre-miRNAs in longan tissue types, we performed qRT-PCR validation on 16 pre-miRNAs. The expression patterns of 16 pre-miRNAs were classified into four categories based on the tissue sites at the highest expression levels. Of the 13 tissues surveyed, Anther

samples showed the most number of highly expressed pre-miRNAs, while root, stem, alabastrum, flower bud, filament and ripe fruit samples did not have high expression levels of pre-miRNAs.

In category I, we noted that pre-miR482a-scaffold26, pre-miR396a-scaffold38, and pre-miR390a-scaffold24, which were highly expressed in the leaf stage (Figure 7a). In category Ⅱ, five of the 16 pre-miRNAs were highly expressed in the anther stage (pre-miR319a-scaffold517, pre-miR171f-scaffold537, pre-miR170-scaffold1266, pre-miR166-scaffold338, and pre-miR397-scaffold21) (Figure 7b). In category Ⅲ, pre-miR394a-scaffold3884, pre-miR395a-scaffold6, pre-miR393h- scaffold978, and pre-miR398b-scaffold6 were highly expressed in longan pulp (Figure 7c). In category Ⅳ, the remaining 4 pre-miRNAs were highly expressed in the female flowers (pre-miR160a-scaffold209), leaf buds (pre-miR390a-scaffold73), male flower (pre-miR2118a-scaffold26) and young fruit (pre-miR168a-scaffold2621), respectively (Figure 7d). In summary, pre-miRNA genes regulate different tissue parts of the longan. Among them, pre-miR398b-scaffold6, pre-miR168a-scaffold2621, and pre-miR390a- scaffold73 were relatively stable in various tissues, and we speculated that they did not play a major role in each tissue.

Discussion

Longan (*Dimocarpus longan* Lour.) is an important fruit tree with medicinal value and considerable economic value. However, there is no research on genome-wide identification of miRNAs genes in longan. In earlier studies, longan genomic data[36] and miRNA databases[20] have been published. A total of 88 miRNA precursors were screened by searching longan small RNA datasets. By analyzing the length and structure of longan precursor miRNAs, it is found that the structure and size of plant miRNA precursors are more diverse[50]. The length of longan pre-miRNAs varied from 77 to 372 nucleotides with an average of 148.37 nt, this distribution of longan pre-miRNAs lengths is similar to previously reported in Arabidopsis, rice, cotton, and maize[51]. In addition, Longan pre-miRNAs have high negative folding free energies (ΔG) of between −18.5 (miR1511) and −185 kcal/mol (miR773) with an average of −51.7 kcal/mol. These values were close to those pre-miRNAs from soybean (average = −39.60 ±17.26 kcal/mol)[39], and Arabidopsis (average = −58.8 kcal/mol), but lower than pre-miRNAs from flax (average = −62.41 kcal/mol), and maize (an average of 126 ± 51 kcal/mol) [51]. Previous study has showed that the ΔG values of the pre-miRNAs were strongly and positively corrected with their RNA/DNA sequence length[39].

Multiple miRNA genes were generally transcribed as polycistronic primary transcripts and found in clusters [52], such as miR156[10, 53], −166 [54], −167[53], −169[37, 55], and −399[40]; these miRNA gene clusters may be functionally related [52]. In this study, All of miR395 members all located on scaffold6; the closest distance of (scaffold6: 2146151:2146267:+) and miR395d−5p (scaffold6: 2145212:2145296:-) is 855 nt, and the fairest distance among them is 51,585 nt. These results implied that miR395 cluster is not characterized only in case of monocots[10, 56], but also presented in dicotyledons [57]. Two of three miR399 members on scaffold 167 were 5,095 nt separating adjacent genes. This two mentioned above clusters of miR395 and −399 were also detected in apple [57]. Two members of the miR477 family (miR473 and −477c) on scaffold192 were separated by only 110 nt, and

miR477 clusters were also observed in Zinna, Helianthus, Parthenium, and Barnadesia [58]. Another cluster contained three miR3627 members on scaffold24, the closest distance of them were separated by 4,485 nt. Previous study showed that miR166 cluster is conserved across plants[59]. While in our study, miR166 clusters were not detected in longan genome. In pear study, 66 miRNAs within a 100 kb of inter-miRNA distance, were also considered as clustered[60]. Therefore, two miR167 members on scaffold53, were separated by less than 14 kb, may be a potential cluster in longan genome. As we known, 25.35% miRNAs in Arabidopsis, 17.09% in Populus, 22.29 % in rice and 21.62 % in sorghum were clustered within a 10-kb region[40]. So, clustered miRNA is a prevalent phenomenon in plants, and these fluctuation lengths of miRNA clusters suggests they may be differentially transcribed, which might be caused by gene duplication or segment duplication[37, 61]. These identified miRNAs clusters will lay a good foundation for detecting function of different miRNA members in longan.

In Arabidopsis, the 20 characterized pri-miRNAs varied in size from 378 nt (miR160b) to 3,108 nt (miR156a)[62], while the FLcDNA lengths for maize pri-miRNA transcripts ranged from 642 nt (miR169) to 2,780 nt (miR167d)[10]. In our study, the pri-miRNAs varied in size with the longest transcripts found for miR156 (scaffold29) (881 nt) and –156 (Unigene5849) (828 nt) and the shortest pri-miRNA for miR156 (scaffold38) (309 nt) for the 24 analyzed longan miRNA genes. It was obviously found that the longest miRNAs were miR156 genes both in Arabidopsis and longan, while is miR167d in maize, implying that the fluctuations lengths of miRNA genes may be not conserved between dicotyledons and monocotyledons. miRNA genes from the same miRNA family have significantly different lengths and structures in Arabidopsis[62] and maize[10], such as miR156, –160, and –172 families. Here, we also found the same result in longan miRNA genes. For example, the miR156 (scaffold29) gene is 881nt long while the other member for miR156 (scaffold38) is only 309 nt long.

The accurate identification of the TSSs is a crucial component in the analysis of a miRNA promoter region[43]. However, most TSSs of miRNA in plants were predicted by bioinformatics method without verifying by experiments[44]. TSSs for 63, 76, and 11 miRNA genes in Arabidopsis[12], maize[10], and soybean [63] were identified by conducting 5’ RACE, respectively. And results showed that pri-miRNAs contain multiple TSSs and have a high prevalence of an adenine (A) at the TSS[64]. In previously studies, we also identified TSSs for miR160[45], miR166_scaffold 53(Zhang et al., 2018), miR167[33] and miR390[46] by RLM-RACE technology in longan. Among the 16 miRNAs with one TSS, the first nucleotide bias (TSS base composition) analysis indicated that A was the most frequently in longan miRNA genes, which is simillar to that protein-encoding genes of longan[47] and pri-miRNAs of Arabidopsis, rice and maize[64], suggesting pri-miRNA has similartranscription speciality as protein-encoding genes, and prevalence of an A at the TSS in plants. In addition, studies of Arabidopsis, rice, and maize showing two TSSs per locus, with an average distance of 4 to 9 nt apart[64]. However, in our study, the average distance between two TSSs per locus of pri-miRNAs in longan is much farer than that those miRNAs in Arabidopsis, rice, and maize, which can be up to several hundred nucleotides, suggesting the complexity of miRNA process in plants.

Until now, almost all miRNA promoters were previously discovered using bioinformatics methods in Arabidopsis[65], *Oryza sativa*[66], soybean (*Glycine max)*[13], *Sorghum bicolor*[40], flax (*Linum usitatissimum)* [11], and Populus [48]. These studies indicate that plant *MIRNA* genes were usually transcribed by RNA polymerase II (Pol II), and contain TATA box and CAAT box like protein-coding genes. In contrast to the above-mentioned reports that attest to the importance of the regulation of *MIRNA* genes themselves, in longan, we just have cloned the partial 5′-flanking regions (less than 1000 bp) of upstream of the TSS in miR160, miR167, and miR390[67] by Tail-PCR. Thus, our knowledge on the regulatory role of longan miRNA genes at the transcriptional level is still largely unknown. In longan, the presence of conserved TATA boxes was revealed in all 19 selected miRNA genes, indicating that the selected longan miRNA genes have the same promoter characteristics as the protein-coding genes. However, the minority of TATA-less miRNA promoters were also found in Arabidopsis[68], rice[69], maize[10], flax[70] and soybean[53], indicating that TATA-less miRNA promoters were popular in plants. While in our study, the TATA-less promoters were not detected, it may be due to the limited analyzed number of miRNA genes in this study. Previously study showed that the MYB-binding domain was the most abundant in flax miRNA genes[70]. However, in our study, the elements of the MYB binding site were rare detected in the 19 miRNA genes just for 45 times in total. Moreover, compared with these miRNA promoters, the miR156(scaffold29) promoter contained the greatest number of elements, reach more than 82 elements, then followed by miR156(scaffold38) and miR170(scaffold1266), suggesting that their extensive roles in longan; and miR390((KJ372216) promoter contained the least number of elements (52). Significant differences in the number and distribution of *cis*-acting elements among miRNAs were also found in*Arabidopsis thaliana, Populus trichocarpa, Oryza sativa*[40]. The results provide new insights into the transcriptional regulation of longan miRNAs.

The expression level of mature miRNA in plant SE has been extensively verified. In Lilium SE, the expression levels of 12 differentially expressed miRNAs(miR156b.2, 319b.1, 390b, 399i, 482b, 529c, 2118, 167b, 394a, 395k, 396l, and 528a) were validated using qRT-PCR[71]. Ten conserved miRNAs(miR156, 168, 171, 159, 164, 390, 397, 166, 167, and 398) with critical functions in the development of the SE system of Valencia sweet orange were detected by stem-loop qRT-PCR[23]. Yang et al. verified relatively abundant miRNAs in cotton EC (miR156, 164, 167, 390, and 3476)[17]. However, there is little verification of the expression level of miRNA precursors, and there is no systematic verification of the expression of miRNA precursors in longan. In this study, the expression of 16 pre-miRNAs in the early SE stage of longan was verified. We compared the expression of corresponding miRNAs in plants. In category I, five pre-miRNAs exhibited high levels in the EC stage. The mature miRNA of pre-miR394a was down-regulated in EC, and its expression levels peaked in GE[72]. In longan, the mature miRNA of pre-miR168a expressed relatively high levels from stages heart-shaped embryos to mature embryos[20]. MiR394a and miR168a inconsistent with the expression of its precursor. Studies have shown that miR168[73], miR394(Han-Hua et al., 2008; Wei and Zhi, 2010), and miR396[74] were involved in development and abiotic stress regulation in Arabidopsis. MiR396a/b is able to target 7 auxin regulatory factors in Arabidopsis[74, 75], and overexpression of miR396a results in a decrease in cell number, cell enlargement, and inhibition of cell proliferation[74, 76]. Plant SE requires cell proliferation and differentiation. In this experiment, the

expression level of pre-miR396a-Scaffold38 decreases with SE and may be related to cell proliferation in longan SE. The five pre-miRNAs could be related to maintaining the pre-embryonic status of the EC. In category II, six pre-miRNAs had the highest expression in the ICpEC stage or CpECGE stage. In longan, miR393 was low expressed in the early stage of ICpEC, and is opposite to the expression of pre-miR393h-scaffold978[77], which may be due to different members of the same family. In contrast, pre-miR166 is consistent with mature miR166 expression in longan[56]. In addition, inducing longan EC to GE without exogenous 2,4-D conditions, both miR171b[78] and miR397a[25] were relatively highly expressed in the ICpEC stage, similar to the expression trends of pre-miR171f-scaffold537 and pre-miR397-scaffold21. Plant hormones play important roles in plant growth and SE in plants[79]. Studies have shown that miR166, miR171f, miR393 and miR390 were all related to plant response hormones. Among them, miR166 mediates the biosynthesis of auxin during embryogenesis[80, 81], miR393 is involved in auxin signal transduction[82], miR390 regulates auxin response factor[83, 84], and miR171f responds to IAA[85]. It is speculated that more miRNA precursors respond to hormones were involved in regulation. In category III, the expression of the 5 pre-miRNAs was highest in the GE stage. During longan somatic embryogenesis, the mature miRNA of pre-miR398b is highly expressed in heart-shaped embryos, torpedo-shaped embryos and cotyledon embryos [31], and miR160a is highly expressed in heart-shaped embryos and torpedo-shaped embryos[35]. The miR319 was involved in a variety of biological pathways in plants such as hormone synthesis, signal transduction, and cell proliferation and differentiation[86]. In addition, miR160a regulates auxin response factors[87] and miR398b responds to oxidative stress[88].

At present, a lot of research has been done on the function of miRNA in different tissue parts, including soybean (root and seed)[89, 90], *Spartina alterniflora* (leaf)[91], Upland cotton(flower)[92], and olive(fruit) [93]. OsmiR396a expression was detectable in all the tissues examined, but predominated in leaves[94]. In addition, miR390a has been previously reported to be involved in dark-induced leaf senescence in Arabidopsis[95]. This shows that pre-miR396a and pre-miR390a have similar expression trends to their mature miRNAs, and were highly expressed in leaves. In Arabidopsis, miR319a is expressed most broadly in developing petals and stamens[96]. In longan, miR166-scafflod338 has the highest expression in anthers and the lowest expression in leaf buds and flower buds[27]. miR171f maintain apical dominance in pear[85]. miR319a and miR171f do not have the same expression trend in tissue type as their precursors, while miR166 has the same expression trend as its precursor. Among them miR319a[86], miR171f(Jiang et al., 2018) and miR166[80, 81] associated with hormone response, miR397[97] associated with plant resistance. The study suggests that proper regulation by miR319a of *TCP4* is critical in these floral organs[100, 101]. The above analysis shows that a large number of pre-miRNAs were regulated in many aspects in the longan anther stage. Previous studies have shown that MiR395[98, 99] and miR394 [100] play important roles in fruit development and ripening. Xu et al. suggest that post-transcriptional regulation of CsTIR1 and CsAFB2 mediated by miR393 is essential for cucumber fruit set initiation[101]. Damodharan et al. demonstrated that sly-miR160a promotes the development of floral organs by quantitatively regulating its main target, SlARF10A[102]. miR168 showed significant differential expression during pre-ripening and post-ripening stages of the pineapple fruit.

Conclusions

This is the first report concerning genome-wide identification and expression profiling of miRNAs genes in longan. Totally, 88 miRNA precursors were obtained. The length of longan pre-miRNAs varied from 77 to 372 nucleotides with an average of 148.37 nt. Through longan miRNA precursor sequence alignment and conservation analysis found that different members of the same miRNA family has highly conserved in the region of mature miRNA, but less conserved in the remaining region. We obtained the upstream transcription 5 'region of 19 miRNAs and the downstream transcription 3' region of 12 miRNAs with typical poly (a) tails. TSSs of 22 miRNA genes in longan were also determined using RLM-RACE and SMART-RACE analysis. In addition, Of the available 12 miRNA gene transcripts with 3' regions information in longan, 5 of them have 3'UTR polymorphism trait and contain two stop sites. The top three most abundant cis-elements were light-responsive elements, stress-responsive elements, and hormone responsive elements. In longan early SE, pre-miRNAs were differentially expressed. In different tissue types of longan, the most number of highly expressed pre-miRNAs were detected in anther samples. Our results contribute to the foundation for future research into the functional roles of miRNA genes in longan.

Methods

Plant materials

Synchronized embryogenic cultures at different developmental stages from *D. longan* 'Honghezi' (Fujian, China), consisting of friable-embryogenic callus (EC), pro-embryogenic cultures (ICpEC), compact pro-embryogenic cultures(CpECGE) and globular embryos (GE), were obtained as detailed in Lai and Chen (1997). The above materials were induced to longan tree of "HHZ" cultivar from the Fujian Agriculture and Forestry University, China. Thirteen tissue samples of *Dimocarpus longan* Lour. cv. Honghezi, including roots, stems, leaves, leafbud, floral buds, anther, filament, male flowers, female flowers, young fruits, ripe fruits and pulp, were collected from Fujian Agriculture and Forestry University venture park from April to July 2015. The mentioned above samples were collected and stored at −80°C for subsequent analyses. The specific permissions or licences of plant collection was not required.

Database searching for pre-miRNAs of longan and sequences analysis

Previously published small RNA data (Bio Sample accession SAMN04120614, Bio-Project ID PRJNA297248) generated using various synchronized embryogenic cultures and transcriptome data (Accession n° SRA050205) generated using EC from the longan 'Honghezi' variety[20], were used to search for pre-miRNAs. The potential pre-miRNAs were checked by using BLAST against the miRBase server (Release 21). The number distribution statistical analysis of selected miRNA genes from longan and 9 other selected plants (*Arabidopsis thaliana, Vitis vinifera, Citrus sinensis, Malus domestica, Prunus*

persica, Solanum lycopersicum, Carica papaya, Populus trichocarpa, and *Oryza sativa)* was conducted; and the miRNA genes from 9 other selected plants were extracted from miRBase server. For creating a graphical representation of the consensus pattern in the RNA secondary structure, we then performed a multiple RNA sequence alignment analysis of selected precursors using the RNALogo tool (http://rnalogo.mbc.nctu.edu.tw/)[103].

Pri-miRNAs 5′ and 3′ RACE experiments in longan

The full-length cDNA sequences of pri-miRNAs were obtained by 5′/3′ RACE PCR and RT-PCR from longan mixed cDNA. Total RNAs were isolated from longan embryogenic cultures at different developmental stages using a TRIzol Reagent kit (Invitrogen, Carlsbad, CA). The mixed cDNAs for pri-miRNAs 5′/3′ RACE PCR were synthesized from mixed total RNA (EC, ICpEC, and GE) using a GeneRacer™ Kit (Invitrogen, Corporation, Carlsbad, CA, USA) or SMARTer™ RACE cDNA Amplification Kit (Clontech, Cat. Nos. 634923 & 634924), following the manufacturer's instructions.

Assembled full-length transcripts from longan mix cDNA were then obtained using DNAMAN ver. 6.0 (Lynnon Biosoft Corp, Pointe-Claire, QC, Canada), and RT-PCR using specifically designed primer pairs confirmed the presence of the full-length cDNAs of pri-miRNAs in longan. Genomic sequences of pri-miRNAs were isolated from longan EC gDNA. The sequences of all primers mentioned above were provided in Table S1.

Genome-wide computational analysis of cis-acting elements in promoter region of longan miRNAs

To study the molecular mechanism underlying the transcriptional regulation of longan miRNA, potential promoter regions (from TSS to 2500 bp upstream) were obtained from longan genome data[36] to predict the potential *cis*-acting elements. For miRNA genes with multiple TSSs, the most far from to the mature miRNA start position is taken. The PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plant- care/html)[104], was used to analyze the *cis*-acting elements of the selected pri-miRNAs.

Quantitative real-time PCR (qRT-PCR) profiling of miRNA transcripts

Total RNA was extracted as described above; cDNAs was generated with 5 mg RNA for miRNA transcripts quantification using PrimeScript™ Perfect Real Time RT Reagent Kit (TaKaRa Code, DR037A). All reactions included 3 biological replicates and three replicate experiments were performed in 96-well reaction plates using a LightCycler 480 machine (Roche Applied Science, Switzerland) and SYBR premix (Takara). After each qRT-PCR run, dissociation curve analyses were performed. Ct values for all miRNA transcripts were normalized using internal standards [105]. *FSD, EF–1α* and *EIF–4α* were used as reference genes to normalize the expression data of pre-miRNAs in longan early SE. *UBQ, EF–1α* and *ACTB* were used as reference genes to normalize the expression data of pre-miRNAs in longan tissue

types. Fold change values were calculated using the comparative △Ct method. Statistical analysis was performed using SPSS 19. Primer sequences for longan miRNA genes were provided in Table S2.

GenBank accession numbers

Longan pri-miRNAs sequences obtained in 5/3′ RACE and RT-PCR experiments were submitted to GenBank (accession numbers: KY576043- KY576062).

Tables

Due to technical limitations, tables are only available as a download in the supplemental files section

Figures

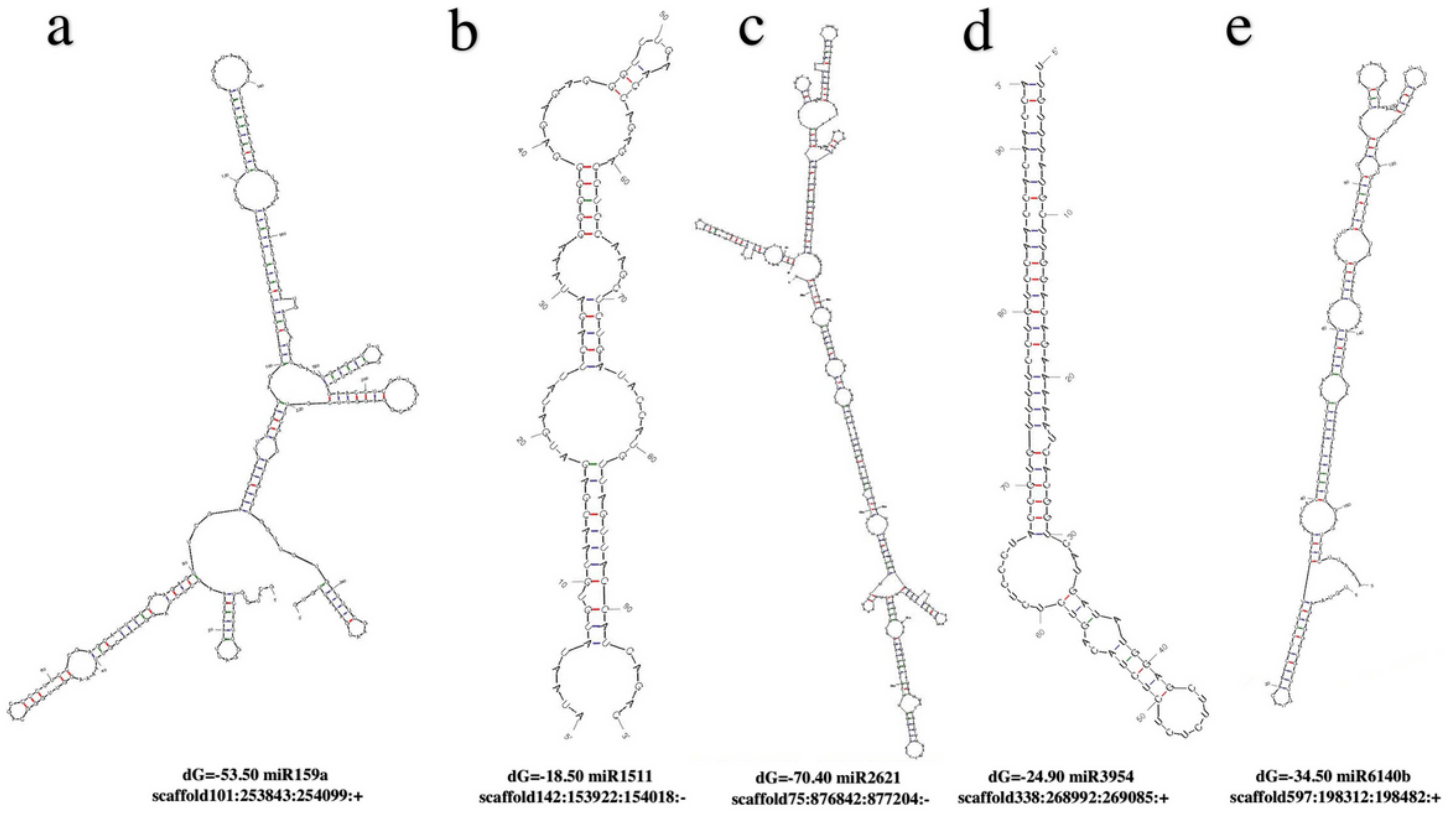

Figure 1

Analysis of stem-loop hairpin structures of pri-miRNAs in D. longan. a. pre-miR159a(scaffold101); b. pre-miR1511(scaffold142); c. pre-miR2621(scaffold75); d. pre-miR3954(scaffold338); e. pre-miR6140b(scaffold597).

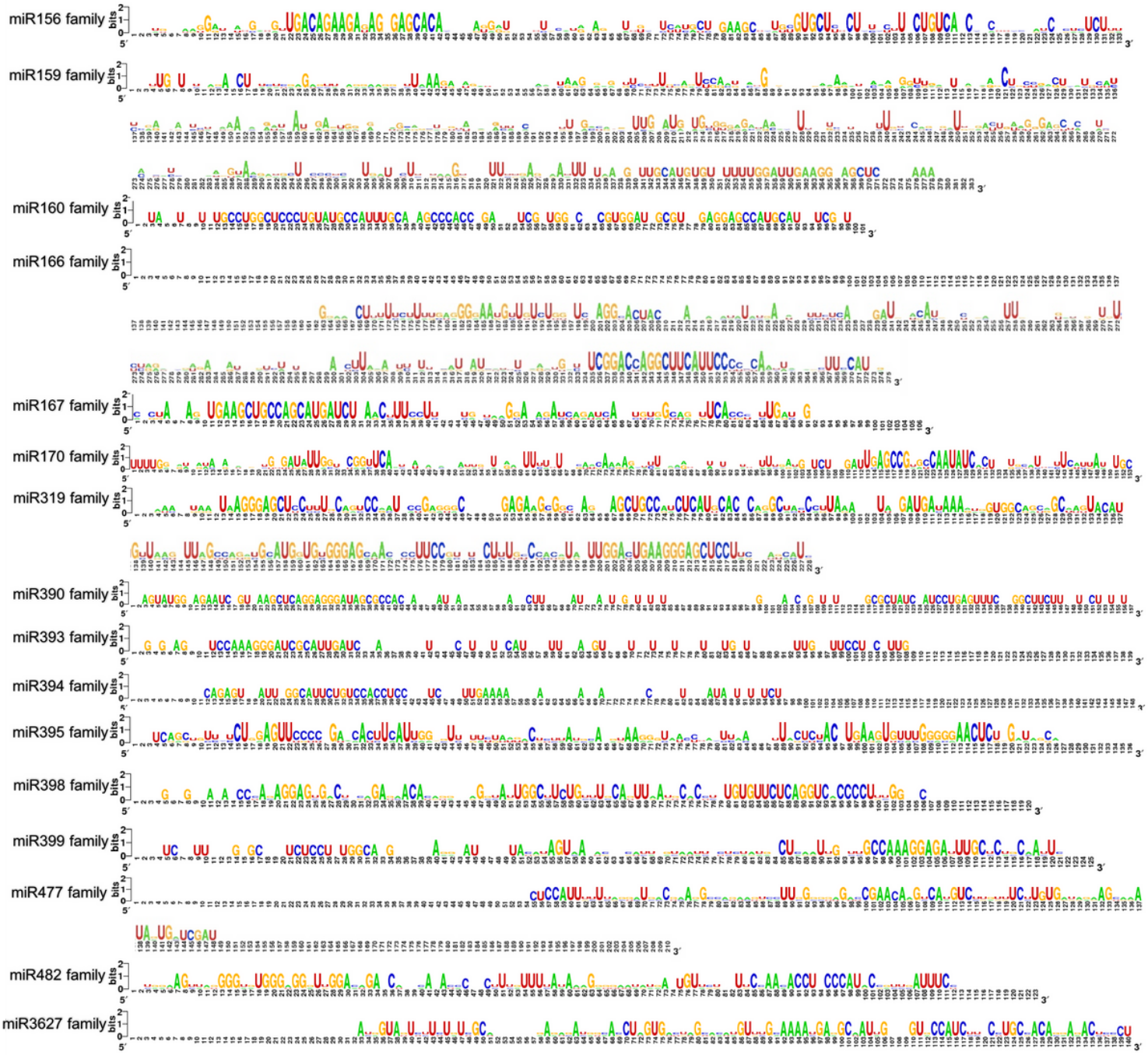

Figure 2

MiRNA precursors sequence alignments. 16 miRNA families with multiple members in longan were aligned by DNAMAN6.0.

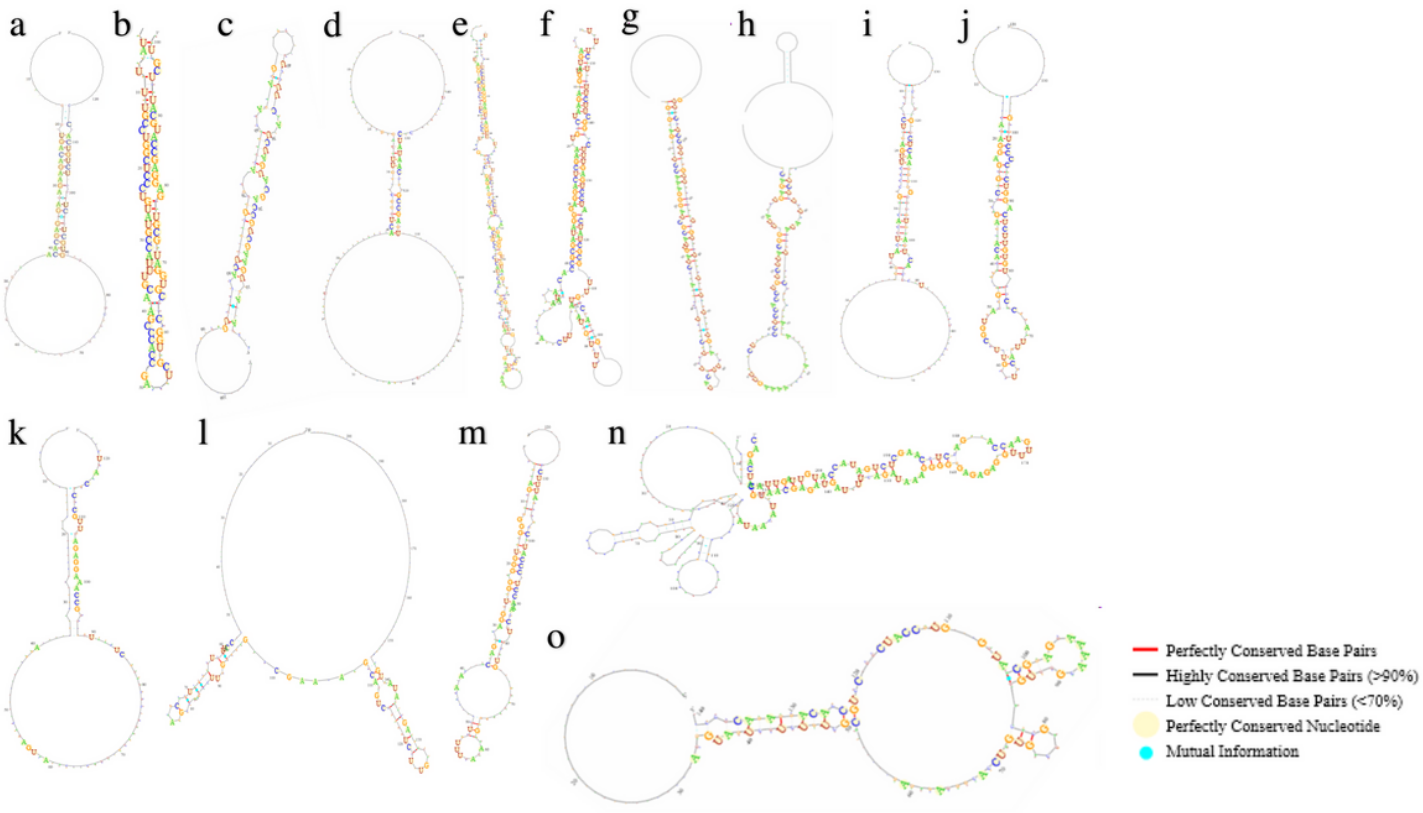

Figure 3

Consensus structures of pre-miRNA families in D.longan, the consensus structure is generated by RNAalifold. a. pre-miR156 family; b. pre-miR160 family; c. pre-miR167 family; d. pre-miR171 family; e. pre-miR319 family; f. pre-miR390 family; g. pre-miR393 family; h. pre-miR394 family; i. pre-miR395 family; j. pre-miR398 family; k. pre-miR399 family; l. pre-miR477 family; m. pre-miR482 family; n. pre-miR1511 family; o. pre-miR3627 family.

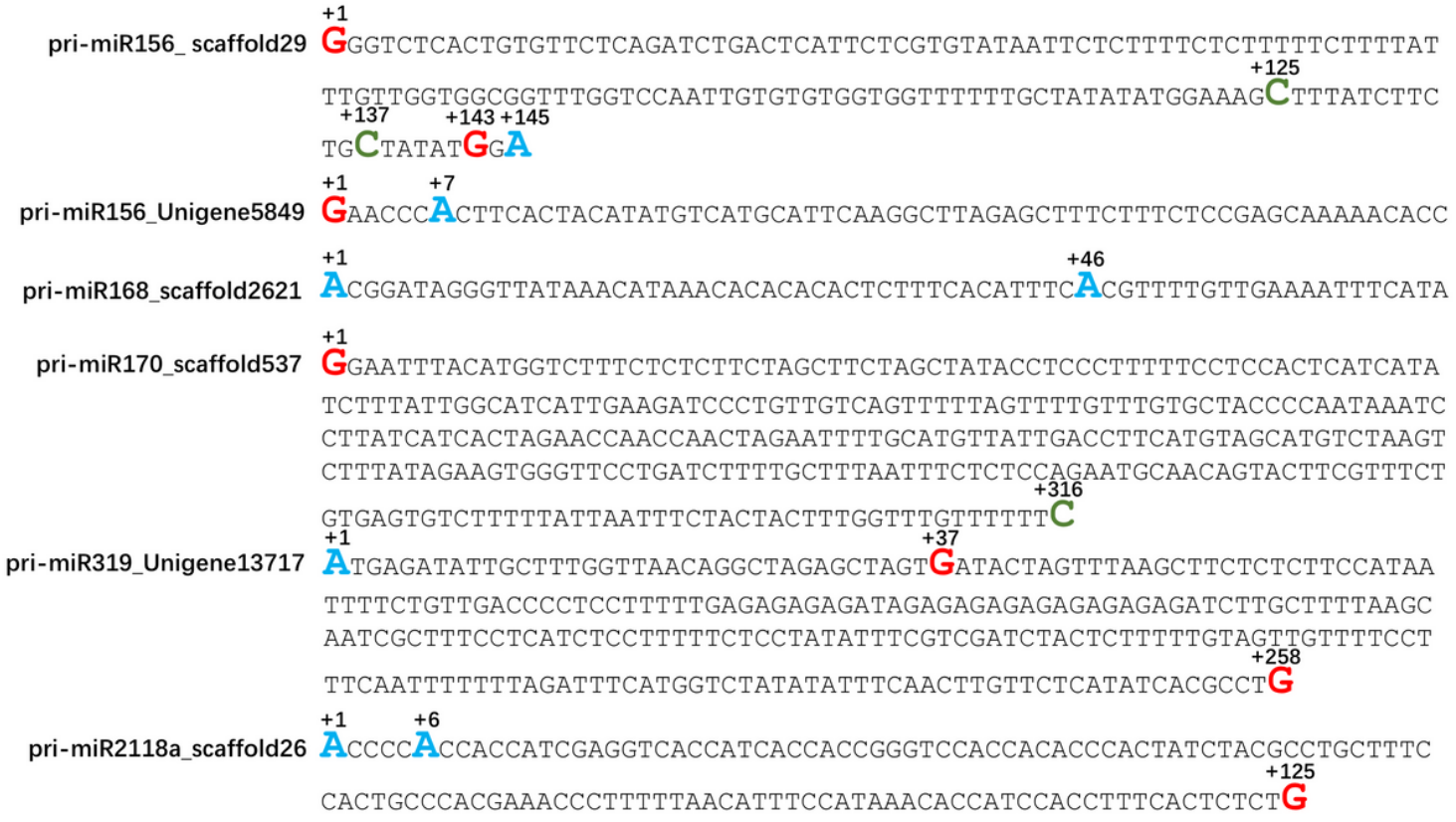

Figure 4

The location of different transcription start sites (TSSs) of miRNAs in D.longan. Red letter on behalf the TSSs is G; blue represent TSSs as A; green represent TSSs as C.

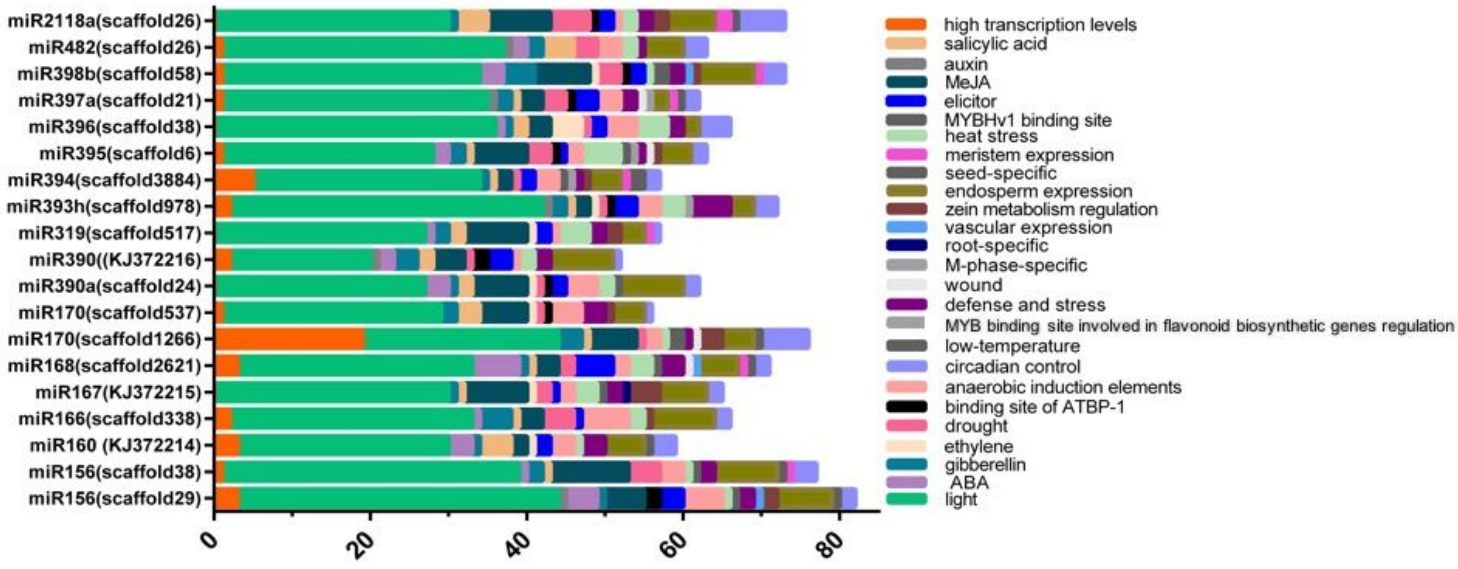

Figure 5

The predicted cis-elements in the promoter of longan miRNA promoters. The 2.5 kb upstream region from the TSSs of the 19 miRNA genes using the PlantCARE database.

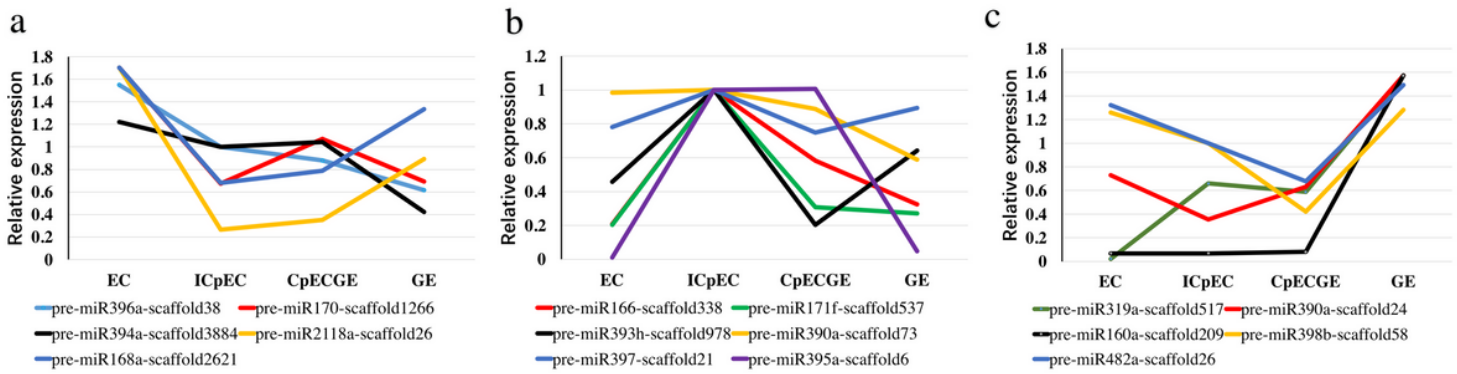

Figure 6

Expression levels of pre-miRNAs during longan early SE. FSD, EF-1α and EIF-4 α were used as a reference gene to normalize pre-miRNAs expression data. a. Five pre-miRNAs exhibited high levels in the EC stage; b. Six pre-miRNAs had the highest expression in the ICpEC stage or CpECGE stage; c. Five pre-miRNAs were highest in the GE stage.

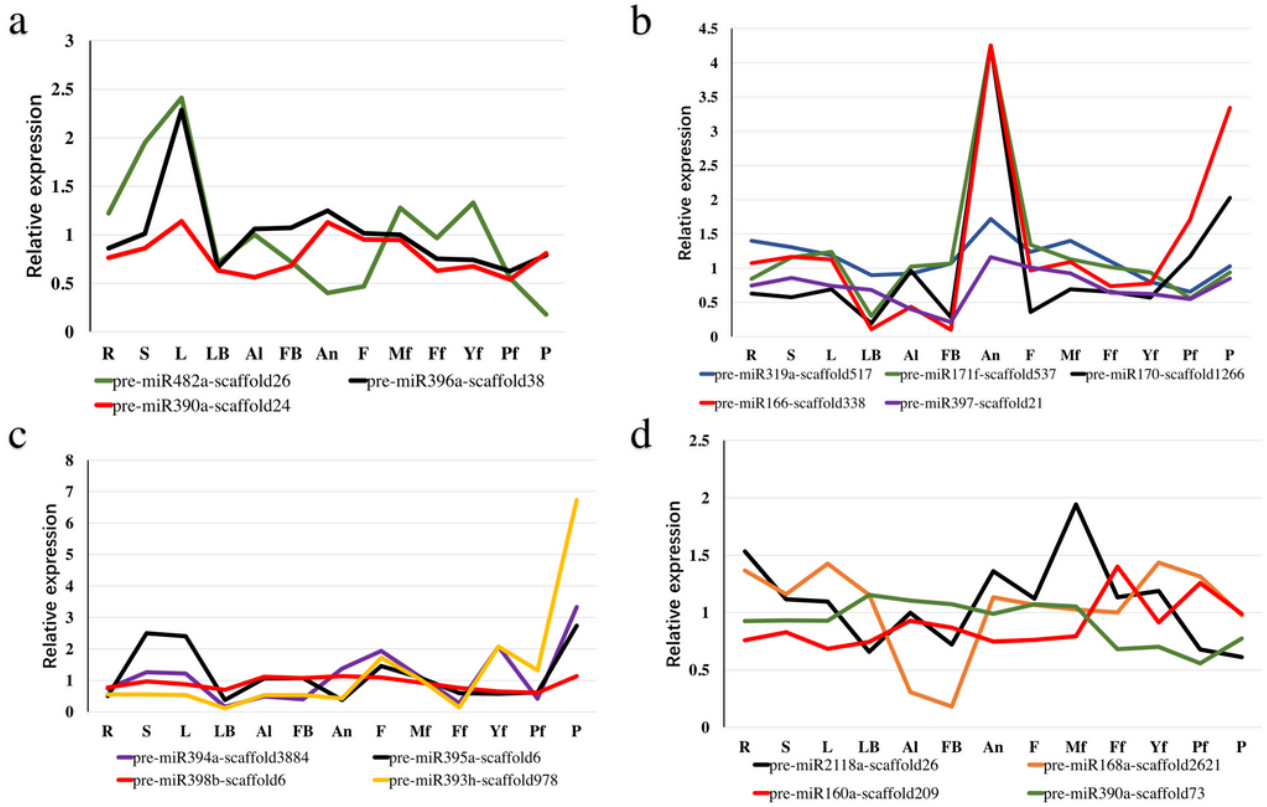

Figure 7

Expression levels of pre-miRNAs across tissue types. UBQ, EF-1α and ACTB were used as a reference gene to normalize pre-miRNAs expression data. a. Three pre-miRNAs were highly expressed in the leaf stage; b. Five pre-miRNAs were highly expressed in the anther stage; c. Four pre-miRNAs were highly expressed in longan pulp; d. Four pre-miRNAs were highly expressed in the female flowers, leaf buds, male flower, and young fruit, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1FigureS1.png
- Additionalfile3TableS1.xlsx
- Table2.xlsx
- Table1.xlsx
- Additionalfile4TableS2.xlsx
- Additionalfile2FigureS2.pdf