

Comparative profiling of microRNA expression in the developing seeds of high- and low-oil tea oil camellia (*Camellia oleifera*) cultivars

Bo Wu

Dalian Minzu University

Chengjiang Ruan (✉ ruan@dlnu.edu.cn)

Key Laboratory of Biotechnology and Bioresources Utilization, Ministry of Education, Institute of Plant Resources, Dalian Minzu University <https://orcid.org/0000-0001-8580-593X>

Wanchen Zhang

Dalian Minzu University

Asad Hussain Shah

University of Kotli Azad Jammu and Kashmir

Sihei Liu

Research Center of Engineering and Technology of Oiltea Camellia of Guizhou Province

Research article

Keywords: tea oil camellia, small RNA sequencing, miRNA-mRNA function modules, lipid metabolism, seed size

Posted Date: July 17th, 2019

DOI: <https://doi.org/10.21203/rs.2.11530/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Tea oil camellia (*Camellia oleifera*), an important woody oil tree, is a source of seed oil of high nutritional and medicinal values and has been widely planted in southern China. However, there are few reports on the identification of miRNAs involved in seed lipid metabolism in high- and low-oil cultivars of tea oil camellia. **Results** An miRNA sequencing database was constructed for an Illumina platform, which was used to perform high-throughput small RNA sequencing of seeds of high- and low-oil cultivars of tea oil camellia at four different developmental stages, and the important relevant miRNAs and their target genes were identified. A total of 196 miRNAs, including 156 known miRNAs from 35 families and 40 novel miRNAs, were identified, and 55 significantly differentially expressed miRNAs were found. An integrated analysis of miRNA and mRNA transcriptome sequence data and qRT-PCR-based information was performed and revealed that 10 miRNA-mRNA function modules were related to lipid metabolism and 23 miRNA-mRNA function modules were involved in the regulation of seed size. **Conclusion** Mining and studying the expression patterns and functions of miRNAs and their regulatory target genes can not only promote the development of miRNAs related to tea oil camellia in public resource databases but also provide important theoretical value and a scientific basis for the genetic improvement of new varieties of tea oil camellia in the future.

Background

Tea oil camellia (*Camellia oleifera* Abel.) is an important woody oil tree in southern China and has been widely planted with a long cultivation history of over two thousand years. Together with oil palm, olive and coconut, tea oil camellia is well known as one of the four major woody oil trees in the world [1, 2]. Tea oil camellia seeds have a high oil content of up to 50%, with rich unsaturated fatty acids (the oleic acid content accounts for over 75%) and an abundance of bioactive compounds such as vitamins, camelliaside, and tea polyphenol, and are an important source of high-grade edible vegetable oil in China [3]. The oil extracted from tea oil camellia seeds is not easily oxidized, as it has high chemical stability and could be preserved for a long time, which plays an important role in the prevention of cardiovascular diseases and cancer and in reducing cholesterol and blood lipids [4, 5]. However, the seed yield of tea oil camellia is unstable, and the seed oil content varies greatly, which seriously restricts the sustainable development of the tea oil camellia industry in China. One important strategy is the cultivation of tea oil camellia cultivars with high seed oil contents and high and stable seed yields.

MicroRNAs (miRNAs) are a class of endogenous noncoding small RNAs with lengths of approximately 20-25 nucleotides and play crucial roles in mediating the posttranscriptional regulation of gene expression by targeting mRNA degradation or inhibiting target mRNA translational in eukaryotes. Studies have confirmed the key roles of miRNAs in various biological and metabolic processes in plants, such as fatty acid biosynthesis [6], lipid metabolism [7], growth and development [8], responses to stresses [9], and cellular proliferation and differentiation [10]. To date, a total of 5,141 miRNA sequences have been deposited in the database miRBase [11, 12], and many conserved and novel miRNAs have been identified in soybean [13], *Arabidopsis* [14], *Brassica napus* [15], rice [16], peony [17] and maize [18]. miRNAs control and influence a variety of physiological processes by regulating target genes. For example, the *KAS* and *KAR* genes targeted by miR159 and miR156, respectively, are important for lipid biosynthesis [19]. The *FAD2* gene targeted by miR159b and miR5026 regulates and influences fatty acid biosynthesis [13, 20]. miR160 plays highly important roles in multiple aspects of shoot and root development by modulating ARF targeting [21]. In addition, transcription factors play important roles in regulating lipid biosynthesis [22], such as WRINKLED1 and LEAFY COTYLEDON, MYB, SPL, ARF and AP2 [23], which have been identified to be targeted by the major miRNAs [24]. Based on the high-throughput sRNA and degradome analyses of soybean seeds 15 days after flower, 55 annotated and 26 novel miRNAs, which targeted 145 genes, were identified [13]; 82% of the targeted genes were transcription factors, including the ARF, MYB, TCP, GRF and NAC families [13]. The expression level of TCPs that are targeted by miR319 controls the coordination and regulation of two sequential processes in leaf development: the negative regulation of leaf growth and positive regulation of leaf senescence in *A. thaliana* [25]. miR160 negatively regulates ARFs that significantly affect seed development in *A. thaliana* [26]. SPL10 and SPL11 are targeted by miR156 and involved in the early morphogenesis of *Arabidopsis* embryos [27]. MYBs targeted by miR159 are involved in seed size, and a double mutation of miR159 (miR159ab) increased MYB33 and MYB65 expression, which promoted the formation of small seeds [28].

Currently, miRNA-mediated regulatory networks controlling multiple biological and metabolic processes have been revealed via the functional analysis of some woody oil plants, such as oil palm [29], peony [17] and olive [30]. In developing sea buckthorn, 19 (four known and 15 novel) and 22 (six known and 16 novel) miRNAs were found to be involved in lipid biosynthesis and seed size, respectively [31], and an integrated analysis of the mRNA and miRNA transcriptomes and qRT-PCR revealed key miRNAs and their targets (miR164d-ARF2, novelmiRNA-108-*ACC*, novelmiRNA-23-*GPD1*, novelmiRNA-58-*DGAT1*, and novelmiRNA-191-*DGAT2*) that are potentially involved in the seed size and lipid biosynthesis of the sea buckthorn seed. However, there is no report on the roles of miRNAs in the different developmental stages of the seeds of high- and low-oil cultivars of tea oil camellia. To understand the potential role of miRNAs and their target genes in developing seeds of tea oil camellia, high-throughput Illumina sequencing technology analysis was used to reveal the global profile of small RNAs in the developing seeds of high- and low-oil tea oil camellia cultivars. Then, we identified some novel and known miRNAs and their target genes by the integrated analysis of mRNA and miRNA transcriptomes combined with quantitative RT-PCR. Finally, we further analyzed miRNA clusters and differentially expressed miRNAs. Our datasets and results will promote the development of tea oil camellia miRNAs in public resource databases and will provide a scientific resource and basis for further investigations of the identified miRNAs of tea oil camellia and their regulatory role in the genetic improvement of high-quality and high-yielding tea oil camellia in the future.

Results

Overview of the sequencing data from raw data to cleaned sequences

To comprehensively identify key small RNAs in the seeds of high- and low-oil cultivars of tea oil camellia at four different seed developmental stages, 16 small RNA libraries were sequenced on an Illumina HiSeq2500 platform, with two biological replicates for each seed developmental stage of each cultivar. A total of 13,572,633 and 14,183,198 raw reads were

generated from the tea oil camellia cultivars 'M3' and 'M8', respectively. Further, the raw reads were analyzed with the LC sRNA analysis pipeline ACGT101-miR, and a total of 11,653,829 (85.67%) and 12,248,627 (86.23%) valid reads were obtained for 'M3' and 'M8', respectively (Table S3 and Table S4). Based on the analysis and statistics of the raw sequencing data, the analysis of the overall length distributions of the valid data after filtration (Fig. 1) showed that most of the valid data were 20-24 nt in length, which was consistent with the typical characteristics of dicer enzymatic cleavage.

Identification and analysis of known and novel miRNAs

To detect the miRNAs involved in the development of tea oil camellia seeds, the expressed miRNAs were classified into four groups (gp1, gp2a, gp2b and gp3) according to defined criteria (Table S5). The unique sequences were selected and mapped against miRBase 21.0 by BLAST search to identify the known miRNAs and novel miRNAs. A total of 196 expressed miRNAs were identified in the developing seeds of tea oil camellia, including seven known miRNAs belonging to the gp1 group, 149 conserved and known miRNAs belonging to the groups gp2a, gp2b and gp3 and 40 novel miRNAs belonging to gp4 (Table S5). Analyses of the number and length ratios of unique miRNA sequences of different lengths revealed that the lengths of miRNAs were mainly distributed within 20~24 nt (Table S6). The number of miRNAs with a length of 21 nt was frequent and accounted for 49.55% of all miRNAs, which is consistent with the definition of miRNAs (Table S8).

The results of the statistical analysis for the expression level of the detected miRNAs and the detection rates of miRNAs by evaluating redundant miRNAs showed that 196 unique miRNAs were identified, including 156 known miRNAs and 40 novel miRNAs (Table S7). Thirty-eight of all novel miRNAs had a medium expression level; in contrast, 38 known miRNAs had a high expression level (a high expression level indicates that the number of reads in the reported miRNAs was higher than the average copy number of the data set; a medium expression level indicates that the number of reads in the reported miRNAs was higher than 10 and less than the average copy number of the data set; and a low expression level indicates that the number of reads in the reported miRNAs was less than 10), and only 4 novel miRNAs had a low expression level, namely, PC-5p-236956_11, PC-3p-145806_25,

PC-3p-253727_10 and PC-3p-150080_24. In addition, we detected the number of specifically expressed miRNAs and the number of specifically coexpressed miRNAs with a Venn diagram by comparing samples 'M3' (a) and 'M8' (b) at four different developmental stages (Fig. 2a, b). It was found that 54 and 63 coexpressed miRNAs were identified in the cultivars 'M3' and 'M8' at four different developmental stages, respectively. Mining these coexpressed miRNAs could provide a favorable foundation for the genetic improvement of tea oil camellia via breeding.

Identification of differentially expressed miRNAs

The differentially expressed miRNAs were identified by suitable differential test methods. Significantly differentially expressed miRNAs were defined with a significance threshold of P -value <0.05 . There were 55 significantly differentially expressed miRNAs identified between the samples of one cultivar at two different developmental stages and between different samples of two cultivars at the same developmental stage. A total of 34 upregulated miRNAs (red) and 21 downregulated miRNAs (green) were identified among these groups (Fig. 3). Moreover, to identify the expression levels of significantly differentially expressed miRNAs at different seed developmental stages, the significantly differentially expressed miRNAs were compared between different groups (Table 1). Five miRNAs (mtr-MIR2586a-p3_1ss15TC, ath-MIR5645b-p5_2ss19TG21TC, ath-MIR5645b-p3_2ss19TG21TC, mtr-MIR2586a-p3_2ss4TG21CT, and hpe-miR166a_1ss5AC) were upregulated in 'M3' relative to 'M8' at the early seed developmental stage. At the second seed developmental stage, three miRNAs (csi-miR1515_1ss16AG, stu-miR8016_1ss7GT, and ppe-miR172a-5p_1ss21GA) were upregulated in 'M3' relative to 'M8'. At the late seed developmental stage, four miRNAs (PC-3p-9080_426, mtr-MIR2586a-p3_1ss15TC, ghr-MIR7499-p3_2ss10GC17CG, and rco-miR167a_R+1) were downregulated in 'M3' compared with their expression in 'M8'.

In addition, the number of common and specific differentially expressed miRNAs was visually shown with a Venn diagram comparison between different groups. The results showed that there were four significantly differentially coexpressed miRNAs (csi-miR1515_1ss16AG, ccl-miR167a, aly-miR167b-3p_1ss13AG, and ppe-miR172a-5p_1ss21GA)

when the different stages in 'M3' were compared and M3T2 vs. M8T2 were compared and that there were six significantly differentially coexpressed miRNAs (csi-miR1515_1ss16AG, ccl-miR167a, aly-miR167b-3p_1ss13AG, ppe-miR172a-5p_1ss21GA, aly-miR167b-3p_1ss13AG, and ppe-miR172a-5p_1ss21GA) when 'M8' was separately compared with M3T2 vs. M8T2 (Fig. 4a). Moreover, six significantly differentially coexpressed miRNAs (aly-miR167b-3p_1ss13AG, ptc-miR172b-5p, ssl-miR828, csi-miR1515_1ss16AG, ath-miR858b and ppe-miR172a-5p_1ss21GA) were found in the comparison of the whole 'M3' and 'M8' groups, but surprisingly, only one differentially coexpressed miRNA (rco-miR167a_R+1) was found in 'M3' among all compared groups relative to M3T4 vs. M8T4 (Fig. 4b). Mining for these significantly differentially expressed and coexpressed miRNAs is helpful for exploring the expression differences and regulatory patterns between high- and low-oil cultivars of tea oil camellia at different developmental stages, which could provide a scientific basis for improved seed oil content and quality in tea oil camellia in the future. These results suggest that miRNA-mediated regulatory mechanisms may play an important role in various biological pathways during seed development in tea oil camellia.

Based on the analysis of the detected miRNAs, we further performed miRNA conservation analysis by mapping the miRNAs against selected species that have conserved evolutionary relationships and counted the frequency of occurrence of the same miRNAs in other species. The results showed that the conserved miRNA frequency was highest in the gma species, and a large amount of these conserved miRNAs also existed in the mdm, ptc and mes species (Table S9). In the family analysis, all known miRNAs were clustered into 35 miRNAs families, 24 of which (miR160, miR156, miR159, miR162, miR164, miR166, miR168, miR167, miR169, miR171, miR172, miR319, miR390, miR393, miR394, miR396, miR397, miR398, miR399, miR403, miR408, miR482, miR530, and miR535) contained many conserved miRNAs (Table S10), demonstrating that they may maintain a wide range of regulatory functions in seed development in tea oil camellia.

Prediction and identification of miRNA target genes

To further understand the functions of miRNAs in developing seeds of high- and low-oil cultivars of tea oil camellia, the target genes of the differentially expressed miRNAs were

predicted using Target Finder software based on the miRNA sequences with mRNA transcriptome sequence data. A total of 17,166 target genes were predicted, and some genes were targeted by multiple miRNAs, resulting in a total of 33,418 predicted miRNA-target gene modules (Table S11). Furthermore, the predicted target genes of known and novel miRNAs were subjected to gene ontology (GO) analysis. The target genes were classified by the hypergeometric distribution algorithm based on the molecular function, biological process and cellular component categories by GO enrichment analysis (Fig. 5). The results showed that a total of 25,988 miRNA-target function modules were related to various biological processes, 21,132 of which were found to be involved in different cellular components, and 14,675 of which were identified to be involved in molecular function. Furthermore, the major biological functions of differentially expressed target genes were identified by the significant terms in the GO enrichment analysis. A total of 107 significant GO terms were identified, which were related to nucleus, sequence-specific DNA binding, regulation of transcription, DNA-template and electron carrier activity (Fig. 6a and Table S12). To further understand the biological functions of the target genes in the developing seeds of tea oil camellia, the target Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway genes were identified by KEGG pathway enrichment analysis. A total of 16,871 miRNA-target function modules were involved in 206 individual KEGG pathways, which were involved in linoleic acid metabolism, alpha-linolenic acid metabolism and glycosylated biosynthesis. Afterwards, the significant KEGG pathway enrichment analysis showed that 31 significant KEGG pathways (Fig. 6b and Table S13) were mainly involved in pyruvate metabolism, regulation of the actin cytoskeleton, fatty acid biosynthesis and citrate cycle.

Correlation analysis of differentially expressed miRNAs and their target genes involved in lipid metabolism

The seeds of the 'M3' cultivar of tea oil camellia have a high oil content and contain rich unsaturated fatty acids. The analyses of the differentially expressed miRNAs and their target genes involved in lipid metabolism revealed that multiple KEGG pathways were related to lipid metabolism during seed development by KEGG enrichment analysis ($P < 0.05$) and that 1,869 miRNA-mRNA function modules were involved in 12 pathways

related to lipid metabolism, including 1,723 predicted known miRNA-mRNA function modules and 146 novel miRNA-mRNA function modules, in which the same miRNA regulated different target genes. There were 188 target genes involved in the fatty acid biosynthesis pathway, 411 involved in the glycolysis/gluconeogenesis pathway, 118 involved in the biosynthesis of the unsaturated fatty acid pathway, and 159 related to the fatty acid metabolism pathway (Table S14). The significant KEGG enrichment analysis and function annotation were integrated, and four significant KEGG pathways related to lipid metabolism were identified, including pyruvate metabolism, fatty acid biosynthesis, glycolysis/gluconeogenesis, sphingolipid metabolism, and steroid biosynthesis. In addition, nine key miRNAs and their target genes related to the lipid metabolism pathway were identified, including one novel miRNA (PC-5p-21064_221) and eight known miRNAs that were involved in the seeds of high- and low-oil cultivars of tea oil camellia at four different stages. These genes are commonly involved in the seed oil biosynthesis pathway, such as *ACC1* (comp59939_c1) targeted by tcc-miR162 and related to the pyruvate metabolism metabolic pathway (ko00620), *KAS1* (comp67779_c0) and *S-ACP-DES6* (comp67006_c0) targeted by csi-miR166e-5p, *KAS3B* (comp61049_c0) targeted by byrgl-miR5139_L-1, *fabG* (comp63911_c0) targeted by mtr-miR156h-3p_1ss8AC, *Mcat* (comp63026_c1) targeted by cpa-miR164d, *FATB1* (comp67050_c0) targeted by PC-5p-21064_221, *MOD1* (comp52017_c0) targeted by aly-miR393a-3p_1ss12TC, *Δ9D* (comp48800_c0) targeted by ath-miR172a and involved in the fatty acid biosynthesis metabolic pathway (ko00061), and *GAPN* (comp67185_c0) targeted by gma-MIR5368-p3_1ss21CT and involved in the glycolysis/gluconeogenesis metabolic pathway (ko00010) (Table 2).

To clarify the molecular mechanism of the above identified miRNAs and their target genes related to lipid metabolism in the high- and low-oil cultivars of tea oil camellia at four different stages, the expression patterns and expression differences of miRNAs and their target genes were analyzed by cluster analysis based on relevant differentially expressed miRNAs and their target gene expression levels at different seed developmental stages (Fig. 7). The results showed that the expression level of cpa-miR164d was the highest at the early seed developmental stage and was significantly higher in 'M3' than in 'M8' at the early seed

developmental stage (June 2), and then it appeared to have a declining trend; in contrast, the expression level of the *Mcat* (comp63026_c1) gene targeted by cpa-miR164d was the lowest at the early seed developmental stage and then gradually increased in 'M3'. These results indicated that there was a negative correlation between the expression level of cpa-miR164d and its target *Mcat* gene. The expression level of mtr-miR156h-3p_1ss8AC was significantly downregulated in 'M3' relative to 'M8' at the early seed developmental stage and then started to increase gradually at the late seed developmental stage; however, the *fabG* (comp63911_c0) gene targeted by mtr-miR156h-3p_1ss8AC was significantly upregulated. The expression level of *fabG* was highest in 'M3' at the early seed developmental stage, gradually decreased and reached the lowest expression level at the fourth developmental stage. The expression level of most miRNAs in the 'M3' cultivar decreased gradually, such as aly-miR393a-3p_1ss12TC, rgl-miR5139_L-1, PC-5p-21064_221, and ath-miR172a, from the second to third seed developmental stage; in contrast, the expression levels of their corresponding target genes, *Mcat*, *KAS3B*, *FATB1* and *Δ9D*, appeared to gradually increase at these two stages.

The expression pattern of csi-miR166e-5p was similarly upregulated in the 'M3' and 'M8' cultivars; however, its target genes, *KAS1* (comp67779_c0) and *S-ACP-DES6* (comp67006_c0), appeared to have a trend of upregulation. The expression level of csi-miR166e-5p was highest at the fourth developmental stage, and there was a positive correlation between the expression levels of csi-miR166e-5p and its target genes. gma-MIR5368-p3_1ss21CT had a similar expression pattern in 'M3' and 'M8', and its expression level was low at the early seed developmental stage and then started to increase gradually. However, the *GAPN* (comp67185_c0) gene targeted by gma-MIR5368-p3_1ss21CT also appeared to have a similar expression pattern in 'M3' and 'M8', and its expression level was first downregulated at the early seed developmental stage and then upregulated, with a peak expression level at the late seed developmental stage. There was also a positively regulated relationship between the expression levels of gma-MIR5368-p3_1ss21CT and its target gene.

Based on these results, it could be speculated that these miRNAs and their target genes might be involved in lipid metabolism pathways. The analysis of their interactions and influence in different relevant lipid metabolism pathways in tea oil camellia will provide important information in the future.

Correlation analysis of differentially expressed miRNAs and their target genes involved in seed size

Seed size is one of the main factors affecting crop yield. Several miRNAs and their target genes belonging to MYB, CNR, and ARF transcription factors, which play important roles in controlling seed size, were identified (Table 3 and Fig. 8). The expression pattern of most miRNAs were slightly different between the high- and low-oil cultivars of tea oil camellia. For example, the expression levels of PC-5p-315953_8 were relatively low in 'M8' at the four seed developmental stages, but this low expression level appeared only at the first to third developmental stages in 'M3', as its expression increased sharply and showed a high expression level at the late seed developmental stage. The transcription factor *CNR2* (comp59426_c0), a target of PC-5p-315953_8, was maintained at a low expression level in 'M3' at the late seed developmental stage, but its expression level was the highest at the third seed developmental stage; however, the expression level of *CNR2* was the lowest in 'M8' at the third seed developmental stage.

The expression pattern of ath-miR858b appeared to have a gradually downregulated trend in 'M3' at different seed developmental stages, and its expression level was highest at the early seed developmental stage. The overall expression level of ath-miR858b was lower in the developing seeds of the 'M8' cultivar than in the developing seeds of the 'M3' cultivar; the lowest expression level of ath-miR858b appeared at the third developmental stage, and then it was gradually upregulated. *MYB82* (comp278917_c0), a target transcription factor of ath-miR858b, showed a declining trend in the developing seeds of the 'M3' and 'M8' cultivars of tea oil camellia; its expression level was high at the early seed developmental stage and then declined gradually. The expression level of *MYB3* (comp65157_c0), another target transcription factor of ath-miR858b, was high at the early seed developmental stage

and then gradually declined; the expression level of *MYB3* was lowest at the third seed developmental stage and was gradually upregulated from the third to the fourth developmental stage. At the third seed developmental stage, the expression level of *MYB3* was lower in 'M3' than in 'M8', and the rate of increase in 'M3' was slower than that in 'M8' from the third to the fourth seed developmental stage. In addition, the expression level of *MYB44* (comp31235_c0), another target transcription factor of ath-miR858b, showed a slight difference between the 'M3' and 'M8' cultivars. The expression level of *MYB44* in 'M3' was highest at the early seed developmental stage and then decreased rapidly to the lowest value at the second seed developmental stage; after this, it gradually increased. In contrast, *MYB44* consistently maintained a high expression level and did not change significantly in the developing seeds of 'M8'.

The expression level of mtr-miR156h-3p_1ss8AC was lowest at the early seed developmental stage in the 'M3' and 'M8' cultivars; however, the expression level of its corresponding target transcription factor *MYB44* (comp55270_c0) was highest in the 'M3' and 'M8' cultivars at the early seed developmental stage. The expression level of hpe-miR162a_L-2 in 'M3' was low during all seed developmental stages, but it was high in 'M8' at the early seed developmental stage and then gradually decreased to a low level at the late seed developmental stage. *ARF19* (comp57222_c0), a target transcription factor of hpe-miR162a_L-2, had a similar expression pattern in the 'M3' and 'M8' cultivars from the early to third seed developmental stage, with high expression levels at the early seed developmental stage; after this, its expression gradually decreased but increased again at the third seed developmental stage. From the third to fourth seed developmental stage, there were obvious differences in the expression trends for *ARF19* between the 'M3' and 'M8' cultivars. The expression level of *ARF19* was downregulated in 'M3' but upregulated in 'M8'.

mdm-miR167h_1ss22AT had a similar expression pattern in the 'M3' and 'M8' cultivars of tea oil camellia, with low expression levels at early seed developmental stages and high expression levels at late seed developmental stages. *MED27* (comp135267_c0), a target transcription factor of mdm-miR167h_1ss22AT, had the highest expression level in 'M3' at

the early seed developmental stage and the lowest expression level at the late seed developmental stage; however, the expression level of *MED27* remained high in ‘M8’ throughout all seed developmental stages. The expression level of *MED27* in ‘M8’ was lower than that in ‘M3’ at the early seed developmental stage, but it was higher than that in the ‘M3’ cultivar at other seed developmental stages.

miRNA-mediated regulation of multiple transcription factors is complex but coordinated with each transcription factor to enable their biological functions. The miRNAs that regulate the expression of target genes directly affect seed size; however, whether these predicted target genes can control seed size in tea oil camellia remains unknown.

Correlation analysis of differentially expressed miRNAs and their target genes involved in growth and development

According to the target prediction annotation information, 12 key miRNAs and their target genes related to growth and development were identified between the high- and low-oil cultivars of tea oil camellia during the different seed developmental stages. Based on the heat map analysis, the target genes targeted by relevant miRNAs were ubiquitously expressed in the developing seeds of tea oil camellia, and the different miRNA expression levels were significantly different in the ‘M3’ and ‘M8’ cultivars (Table 4 and Fig. 9). For example, the expression level of mtr-miR156h-3p_1ss8AC was significantly lower in ‘M3’ than in ‘M8’ at the early developmental stage and then increased slowly; however, the expression level of miR156h-3p_1ss8AC was consistently high in ‘M8’ throughout the whole seed developmental stage. TCP24 (comp65957_c0), a target of miR156h-3p_1ss8AC, had a high expression level in ‘M3’ at the early seed developmental stage and was significantly upregulated; however, it was consistently low in ‘M8’ at all seed developmental stages. The relationship between miR156h-3p_1ss8AC and its target, TCP24, was positively regulated.

The expression levels of PC-3p-9080_426 showed significant differences in the ‘M3’ and ‘M8’ cultivars. PC-3p-9080_426 was upregulated gradually in ‘M8’ from the early to third seed developmental stage but had the lowest expression level at the fourth seed

developmental stage; in contrast, PC-3p-9080_426 consistently had a higher expression level in 'M3' than in 'M8' at the whole seed developmental stage, although it showed a slightly decreasing trend at the late seed developmental stage. The *COL13* (comp63752_c0) gene targeted by PC-3p-9080_426 was downregulated in both cultivars of tea oil camellia at the late seed developmental stage, although its expression level was high in 'M3' at the early developmental stage and was then downregulated. The expression pattern of *ERF115* (comp58165_c0), another target gene of PC-3p-9080_426, was similar to that of the *COL13* gene in the 'M3' cultivar. However, in the 'M8' cultivar, the expression level of *ERF115* was highest at the early seed developmental stage; after this, it decreased rapidly to a low expression level at the third seed developmental stage and then had a slightly increased expression level at the late seed developmental stage.

The expression pattern of han-miR156a_L+1 was similar in the 'M3' and 'M8' cultivars of tea oil camellia. Its expression level was significantly low at the early seed developmental stage and then increased sharply, reaching the highest level at the late seed developmental stage. The two genes *SPL4* (comp60372_c0) and *SBP2* (comp62492_c0) targeted by han-miR156a_L+1 had the highest expression levels at the early seed developmental stage in the two cultivars of tea oil camellia and the lowest expression levels at the late developmental stage. The expression pattern of rco-miR156e_L+1R-1 was similar to that of han-miR156a_L+1, but the expression patterns of its three target genes, *ACS1* (comp66521_c0), *SBP2* (comp62492_c0) and *SBP1* (comp61590_c1), contrasted with that of the miRNA han-miR156a_L+1, showing a downward trend in developing seeds.

According to these results, it could be found that these miRNAs and their target genes have highly correlated relationships in expression levels during seed development, and the differences in the expression levels between the two cultivars or among the different seed developmental stages might exert a great influence on the growth and development of tea oil camellia seeds.

Correlation analysis of differentially expressed miRNAs and their target genes involved in resistance

According to the enrichment analysis of the target genes of differentially expressed miRNAs, a total of 11 functional modules of miRNAs and targeted genes related to stress resistance were identified, mainly involving two categories: abiotic stresses and biotic stresses (Table 5). In addition, two miRNA-mRNA functional modules, which may affect the yield and quality of tea oil camellia seed, were also identified. Their expression patterns and levels appeared to be obviously different between the 'M3' and 'M8' cultivars of tea oil camellia and among different seed developmental stages (Fig. 10).

The results of the heat map analysis showed that the majority of miRNAs involved in stress resistance displayed different expression profiles in two cultivars of tea oil camellia at four different seed developmental stages. The expression patterns of *ath-miR858b* and *nta-MIR6149a-p5_2ss18CT21TA* were different in the high- and low-oil cultivars. Their expression levels were consistently high in 'M3' throughout the four seed developmental stages; however, for the 'M8' cultivar, their expression levels were high at the early and second seed developmental stages, gradually decreased to the lowest level at the third seed developmental stage, and then slowly increased. In contrast, the expression level of *MYB4* (*comp31687_c0*), a target gene of *ath-miR858b*, was lowest in 'M8' at the fourth seed developmental stage; the expression level of *CPR30* (*comp65951_c0*), a target gene of *nta-MIR6149a-p5_2ss18CT21TA*, was lowest in 'M3' at the second seed developmental stage, but its highest expression level in 'M8' appeared at the third seed developmental stage relative to the other stages.

The expression level of *mtr-miR156h-3p_1ss8AC* was lowest in 'M3' at the early seed developmental stage and then increased gradually; however, the expression level of *mtr-miR156h-3p_1ss8AC* remained consistently high in 'M8' throughout the four different seed developmental stages. Interestingly, the two genes *BAG7* (*comp63682_c0*) and *NRT1.7* (*comp66713_c0*) targeted by *mtr-miR156h-3p_1ss8AC* showed low expression levels in 'M3' at the early seed developmental stage and then decreased gradually to the lowest expression level at the late seed developmental stage. Although the expression levels of these two target genes were also lowest in the 'M8' cultivar at the late seed developmental stage, they were higher than those in 'M3'.

The expression levels of han-miR156a_L 1 in the 'M3' and 'M8' cultivars were lowest at the early seed developmental stage and then increased gradually. In contrast, the expression level of its target gene, *SPL16* (comp56544_c0), was highest at the early seed developmental stage and then decreased. The expression level of ppe-miR172a-5p_1ss21GA remained high in 'M3' at all seed developmental stages, but its expression level in 'M8' was only high at the early seed developmental stage and then was significantly downregulated after the second seed developmental stage, with the lowest expression level at the fourth seed developmental stage. The expression of the *PAB2* (comp65585_c0) gene targeted by ppe-miR172a-5p_1ss21GA was lowest in 'M3' at the late seed developmental stage, and its highest expression level in 'M8' appeared at the early seed developmental stage.

Validation of differentially expressed miRNAs and their target genes by qRT-PCR

To validate the high-throughput sequencing data, six functional modules of differentially expressed miRNAs and their target genes were selected, han-miR156a_L+1-*SPL4*, rco-miR156e_L+1R-1-*SBP1*, mtr-miR156h-3p_1ss8AC-*TCP24*, ppe-miR172a-5p_1ss21GA-*PAB2*, nta-MIR6149a-p5_2ss18CT21TA-*CPR30* and PC-3p-9080_426-*ERF115*. The differences in their expression patterns and expression levels in the high- and low-oil cultivars of tea oil camellia at the four different seed developmental stages were identified by qRT-PCR analysis (Fig. 11 and Fig. 12). The results showed that six pairs of functional modules of miRNAs and their target genes showed differential expression, and the expression patterns of most miRNAs and their target genes were consistent with the results obtained by high-throughput small RNA sequencing data. All six selected miRNAs showed opposite expression trends relative to their target genes.

han-miR156a_L 1 was upregulated in high- and low-oil cultivars of tea oil camellia, and its expression level was lowest at the early seed developmental stage and highest at the late developmental stage (Fig. 11a). The expression level of han-miR156a_L 1 in the 'M3' cultivar was higher than that in the 'M8' cultivar. In contrast, the expression level of the target gene *SPL4* of this miRNA was highest at the early seed developmental stage and then decreased sharply to the lowest expression level at the late seed developmental stage (Fig.

12a). The expression level of rco-miR156e_L 1R-1 was upregulated in the 'M8' cultivar; it gradually increased from the early to third seed developmental stage, peaking at the third seed developmental stage, and then decreased slightly (Fig. 11b). The expression level of rco-miR156e_L 1R-1 in the 'M3' cultivar was upregulated continuously and reached the highest value at the late seed developmental stage but was always lower than that in the 'M8' cultivar (Fig. 11a). Interestingly, the target gene, *SBP1*, of rco-miR156e_L 1R-1 was downregulated in the 'M3' and 'M8' cultivars of tea oil camellia, and its expression level was highest at the early seed developmental stage and lowest at the late developmental stage (Fig. 12b).

The expression level of mtr-miR156h-3p_1ss8AC was upregulated in the 'M3' cultivar and was significantly lower than that in the 'M8' cultivar at the early seed developmental stage; after this, it showed a rapidly increasing trend, with the highest expression level at the late seed developmental stage (Fig. 11c). In contrast, the expression level of mtr-miR156h-3p_1ss8AC in the 'M8' cultivar was relatively high at all seed developmental stages. Surprisingly, the expression level of the target gene *TCP24* of mtr-miR156h-3p_1ss8AC in the 'M3' cultivar was also low at the early seed developmental stage (Fig. 12c) but was then rapidly upregulated; however, its expression level in the 'M8' cultivar remained low at all seed developmental stages (Fig. 12c). The expression level of miR156h-3p_1ss8AC was found to be positively correlated with the expression of its target gene in the high-oil 'M3' cultivar but negatively correlated with the expression of its target gene in the low-oil 'M8' cultivar.

The expression level of ppe-miR172a-5p_1ss21GA was highest in 'M3' at the second seed developmental stage (Fig. 11d), while the expression level of its target gene *PAB2* was lower at the second seed developmental stage than at the early and third seed developmental stages. The expression pattern of ppe-miR172a-5p_1ss21GA was similar to that of its target gene in the 'M8' cultivar, and both were first upregulated and then downregulated (Fig. 11d and Fig. 12d). The expression level of ppe-miR172a-5p_1ss21GA was found to be negatively correlated with the expression of its target gene in high-oil 'M3' but positively correlated with the expression of its target gene in low-oil 'M8'.

nta-MIR6149a-p5_2ss18CT21TA had the highest expression level in 'M3' at the third seed developmental stage, but its expression level was lowest in 'M8' at the third seed developmental stage and was significantly lower than that in 'M3' (Fig. 11e). Its corresponding target gene, *CPR30*, was low in the 'M3' cultivar at the fourth developmental stage. The expression of the target gene *CPR30* of nta-MIR6149a-p5_2ss18CT21TA in 'M3' was higher than that in 'M8' at the early seed developmental stage but was lower in 'M3' than 'M8' at the other three seed developmental stages (Fig. 12e). The expression level of PC-3p-9080_426 peaked in 'M3' at the second seed developmental stage and in 'M8' at the third seed developmental stage (Fig. 11f). It was first upregulated and then downregulated in 'M3'. The target gene *ERF115* of nta-MIR6149a-p5_2ss18CT21TA showed the lowest expression level in 'M8' at the third seed developmental stage, and its expression level first increased and then decreased in low-oil 'M3' (Fig. 12f). The expression level of *ERF115* in 'M3' was positively correlated with the expression level of PC-3p-9080_426.

The above results obtained by qRT-PCR validation are consistent with the sequencing results for the expression patterns and levels of miRNAs and their targeted genes (Fig. 11 and Fig. 12), indicating the accuracy and high efficiency of small RNA high-throughput sequencing results and reflecting the dependability of the qRT-PCR method used to identify the temporal and spatial expression characteristics and expression differences of related functional miRNAs and target genes in tea oil camellia at different seed developmental stages.

Relationship validation of the cpa-miR393_R-1-AFB2 functional module by luciferase activity assays

A target prediction analysis server psRNA target was used to assess the complementarity between cpa-miR393_R-1 and the target site of AFB2. The potential to target AFB2 3'UTR of cpa-miR393_R-1 was predicted (Fig. 13a). Luciferase activity in 293 cells cotransfected with the cpa-miR393_R-1 recombinant expression vector and the expression vector containing the 3'UTR of AFB2 fused with the reporter gene decreased by nearly 36.26% (*p*

< 0.05) compared to that in the control group (Fig. 13b). These results indicate that AFB2 is one of the target genes of cpa-miR393_R-1.

Discussion

Identification and analysis of miRNAs in seeds of high- and low-oil cultivars of tea oil camellia

Tea oil camellia is an important woody oil plant tree with high nutritional and ecological value, and it is known for its high-quality seed oil with high oleic acid. Currently, tea oil camellia production is unstable because of low seed yield and weak environmental adaptability. The exploration of the molecular mechanisms of miRNA regulation involved in various biological pathways in plants is a field of great interest to current researchers [32]. However, there are few studies and reports on miRNAs and their target genes in developing tea oil camellia seeds. In this study, a total of 11,653,829 (85.67%) and 12,248,627 (86.23%) valid reads were obtained (Table S3 and Table S4) from the developing seeds of the 'M3' and 'M8' cultivars of tea oil camellia. A total of 196 miRNAs, including 156 known and 40 novel miRNAs, were identified by standard small RNA sequencing (Table S7), and the measured miRNA length was mainly distributed in the range of 21-24 nt (Table S8). Fifty-five significantly differentially expressed miRNAs were identified, including 34 upregulated and 21 downregulated miRNAs (Fig. 3). Six common significantly differentially expressed miRNAs were identified at all seed developmental stages of the 'M3' and 'M8' cultivars (Table 1 and Fig. 4). Subsequently, these significantly differentially coexpressed miRNA expression patterns and the differences between high- and low-oil cultivars were determined and could provide a foundation for the subsequent targeted improvement of the genetic function of tea oil camellia. Studies have shown that miRNAs are highly conserved in evolutionary relationships, especially in plants [33]. The identified miRNAs in tea oil camellia were detected for cross-species conservation analysis, which showed that these identified miRNAs appeared most frequently in gma species. Most of the conserved miRNAs were from 24 miRNA families, such as miR160, miR156, miR159, miR162 and miR164. (Table S9 and Table S10).

Studies have shown that the miRNA *lin-4* can regulate two target genes, *lin-14* and *lin-28*, respectively [34], and the same target gene can also be regulated by different miRNAs, such as miR164, which can regulate NAC family genes in *Arabidopsis* [35]. In this study, for 196 identified miRNAs in the seeds of tea oil camellia, a total of 17,166 target genes were predicted, and there were multiple target genes targeted by the same miRNA. A total of 33,418 miRNA-mRNA function modules were predicted (Table S11), 25,988 of which were related to different biological processes, including 21,132 modules involved in different cell components and 4,675 modules involved in molecular function by GO enrichment analysis (Fig. 5). A total of 107 significantly enriched GO terms were identified and were related to nucleus, sequence-specific DNA binding, regulation of transcription, DNA-templated, and electron carrier activity according to the significant GO enrichment analysis (Fig. 6a and Table S12). Moreover, a total of 16,871 miRNA targets were involved in 206 individual KEGG pathways, which were related to linoleic acid metabolism, alpha-linolenic acid metabolism and glycosphingolipid biosynthesis according to the KEGG enrichment analysis. In total, 31 significantly enriched KEGG pathways (Fig. 6b and Table S13) were mainly involved in pyruvate metabolism, regulation of the actin cytoskeleton, fatty acid biosynthesis and the citrate cycle according to the significant KEGG pathway enrichment analysis. These results indicate that miRNAs regulate different biological functions and metabolic pathways in developing seeds of tea oil camellia. Excavating these important miRNA-mRNA function modules could provide a clear understanding of the biological functions and metabolic pathways of miRNA-regulated target genes during seed development in woody oil trees, providing favorable resources for improving tea oil camellia in the future.

Effects of related miRNAs and their target genes on lipid metabolism in tea oil camellia

Studies have shown that miRNAs are involved in lipid metabolism by regulating targeted genes, such as the 59 identified miRNAs in *B. napus* involved in lipid metabolism [36]. miR156 influences lipid biosynthesis by regulating the expression level of the target gene *KAR* [19]. Based on the analyses of the identified miRNAs in the seeds and other tissues of rapeseed at different developmental stages, it was found that the expression of miR156 was

highest in the seeds [37]. Our study also found that mtr-miR156h-3p_1ss8AC had high expression levels in the tea oil camellia seeds. miR159 targets and regulates the expression of the *KAS* gene, which has an important effect on plant lipid biosynthesis [19]. Previous studies found that the overexpression of the *FATB1* gene in *Arabidopsis* seeds resulted in a 4-fold increase in palmitic acid content in seed oil [38]. *KASI* is involved in the formation of C16 or C18 saturated fatty acids catalyzed by the fatty acid synthase complex [39]. *ACCase* is a rate-limited enzyme that controls the flow of carbon to fatty acid synthesis, and the high expression level of the *ACCase* gene is closely related to the high yield and oil content of oil palm [40, 41]. Currently, there are reports on the key genes functioning in lipid and fatty acid biosynthesis pathways in tea oil camellia, such as *ACCase* [42], *SAD* [43], *DGAT1* [44] and *FAD2* [45]. However, there are no reports on the regulation of related miRNAs involved in lipid biosynthesis in tea oil camellia seeds.

In this study, a total of 1,869 miRNA-mRNA function modules involved in 12 KEGG pathways related to lipid metabolism were identified (Table S14), and four significant pathways were found, including pyruvate metabolism, fatty acid biosynthesis, glycolysis/gluconeogenesis/sphingolipid metabolism and steroid biosynthesis. Ten miRNA-mRNA function modules were related to the significantly relevant lipid metabolism pathways, such as the *ACC1* gene targeted by tcc-miR162 related to the pyruvate metabolism metabolic pathway, the genes of *KAS1* and *S-ACP-DES6* targeted by csi-miR166e-5p, the *KAS3B* gene targeted by byrgl-miR5139_L-1, the *fabG* gene targeted by mtr-miR156h-3p_1ss8AC, the *Mcat* gene targeted by cpa-miR164d, the *FATB1* gene targeted by PC-5p-21064_221, the *MOD1* gene targeted by aly-miR393a-3p_1ss12TC, the *Δ9D* gene targeted by ath-miR172a and involved in the fatty acid biosynthesis metabolic pathway, and the *GAPN* gene targeted by gma-MIR5368-p3_1ss21CT and involved in the glycolysis/gluconeogenesis metabolic pathway (Table 2). The analysis of the expression patterns and expression levels for miRNA-mRNA function modules involved in lipid metabolism (Fig. 7) provides a clear perspective for understanding the role of miRNAs in lipid biosynthesis of tea oil camellia seed, indicating the potential for the utilization of miRNAs to improve the seed oil content and oil components in tea oil camellia in the future.

Effects of related miRNAs and their target genes on seed size in tea oil camellia

Studies have shown that some miRNAs and their target genes are related to seed size [13, 46], and most of the target genes belong to transcription factors, mainly ARF [19, 47], CNR [48] and MYB [49]. In this study, 23 miRNA-mRNA functional modules related to the seed size of tea oil camellia were identified (Table 3), and the expression levels of ARF, CNR, MYB, MED and other related target genes were highly correlated with the expression levels of their regulated miRNAs, in which most of them maintained negatively regulated relationships (Fig. 8). These results indicate that these miRNAs and their target genes may play important roles in controlling the seed size of tea oil camellia, providing a good foundation for further clarification of the relevant miRNA regulatory target genes that participate in the regulatory mechanism of tea oil camellia seed size, but the network mechanism for the regulation of seed size needs to be evaluated further.

Effects of related miRNAs and their target genes on growth and development in tea oil camellia seed

miR156 participates in the early morphogenesis of *Arabidopsis* embryos by regulating target SPL10 and SPL11 [27], and it may also mediate the targeted regulation of *SBP* expression levels to affect flowering time in plants [47]. miR159 mediates target MYB transcription factors to regulate seed growth and development [28]; the genes of *AFB2* and *T1R1* are auxin receptor genes regulated by miR393, which is involved in hormone regulation and signal transduction in plants [50]. Studies have also shown that miRNA171 controls the growth and development of flowers, roots and stems by targeting SCL transcription factor family members in *Arabidopsis* [51].

In this study, 12 miRNA-mRNA functional modules, which mainly targeted *SCL16*, *SPL4*, *SBP2*, *TCP24*, *T1R1* and other genes, related to the growth and development of tea oil camellia seed were identified (Table 4). The expression levels of the targeted genes were highly correlated with their miRNAs in the seeds of high- and low-oil cultivars of tea oil camellia at different developmental stages (Fig. 9). The differential expression of miRNAs

in different cultivars or at different developmental stages had a significant influence on the expression of their targeted genes. These results indicate that these miRNAs-mRNAs may play important regulatory roles in seed growth and development in tea oil camellia.

Effects of related miRNAs and their target genes on stress response in tea oil camellia

Different environmental stress factors induce high or low expression of related miRNAs to external stress [52], such as miRNA167, which plays an important regulatory role in drought stress, and its expression is upregulated when subjected to drought stress [52]. High salt stress induced the downregulation of the expression of miRNA172a in *Arabidopsis* and rice [53, 54], but low temperature promoted the expression levels of miR172 in *A. thaliana* and inhibited the expression of miRNA156 in *Populus euphratica* under low temperature conditions [55]. miR156 regulated members of the SPL family involved in different biological processes, such as growth and development, secondary metabolic synthesis and abiotic stress in plants [56]. The overexpression of soybean miR394a in *Arabidopsis* reduced water loss in *A. thaliana* and improved its drought resistance [57]. The *NRT1.7* gene, which is a member of the NRT family with a retransport function of nitrates, has an important impact on yield and nitrogen use efficiency in plants [58]. In this study, according to the results of target gene enrichment analysis, 11 miRNA-mRNA function modules involved in plant stress response-related pathways were identified, and two of them may affect the yield and quality of seeds of tea oil camellia (Table 5 and Fig. 10).

Conclusion

Tea oil camellia has been widely planted in southern China for the production of edible oil. To further understand the miRNA-regulated mechanisms for lipid metabolism in tea oil camellia seeds, an miRNA sequencing database was constructed for an Illumina platform and used to perform small RNA high-throughput sequencing with the seeds of high- and low-oil cultivars of tea oil camellia at four different developmental stages. A total of 196 miRNAs, including 156 known miRNAs from 35 families and 40 novel miRNAs, were

identified, and 55 significantly differentially expressed miRNAs were found. Ten miRNA-mRNA function modules involved in lipid metabolism and 23 miRNA-mRNA function modules involved in the regulation of seed size were discovered. Our data not only promote the development of miRNAs related to tea oil camellia in public resource databases but also provide important theoretical value and a scientific basis for the genetic improvement of new varieties of tea oil camellia in the future.

Methods

Plant materials

Two cultivars, 'M3' with a high seed oil content and 'M8' with a low seed oil content, of tea oil camellia (voucher No. 03024, identified by Chengjiang Ruan and deposited at Dalian Minzu University), which were growing in the field in Yuping Dong Autonomics County (27°28'—27°31' north latitude, 108°34'—109°09' east longitude), Guizhou Province, China under natural climate conditions, were selected as sample trees. The orchard had a mean annual rainfall of 1174.1 mm, a mean annual temperature of 16.4 °C and a mean annual sunshine 1227.8. The collection of all samples completely complies with local and national legislation permission. Cultivars of 'M3' and 'M8' were selected from natural population of tea oil camellia and bred by Chengjiang Ruan. Fruits of each cultivar were harvested on June 2, July 4, August 5, and September 3 in 2016. Seeds obtained by stripping the fruit shells were immediately wrapped in tin foil and placed into liquid nitrogen. The frozen seeds were transported to the laboratory and stored at -80°C for the following experiments.

Small RNA library construction and sequencing

Total RNA of each sample was extracted by using the TRIzol reagent (Invitrogen, CA, USA), according to the manufacturer's instructions. The quality and quantity of total RNA of each sample were analyzed by the Bioanalyzer 2100 (Agilent, CA, USA) and RNA 6000 Nano LabChip Kit (Agilent, CA, USA) with a RIN (RNA integrity number value) >7.0. The small RNA library of each sample was constructed with the TruSeq Small RNA Sample Prep Kit

(Illumina, San Diego, USA) in accordance with the manufacturer's protocol. Then, we performed single-end sequencing (36 bp or 50 bp) on an Illumina HiSeq2500 at LC-BIO (Hangzhou, China) following the vendor's recommended protocol.

Identification and analysis of miRNA sequencing data

To obtain accurate and reliable miRNA sequencing data, the raw read quality was evaluated by Illumina FastQC to obtain Q30 data. Then, the raw reads were processed to obtain valid reads with an in-house program, ACGT101-miR (LC Sciences, Houston, Texas, USA), to remove and filter adapter dimers, low-quality reads, low-complexity reads, common RNA families (rRNA, tRNA, snRNA, snoRNA) and repeats. To identify the known miRNAs and novel miRNAs in tea oil camellia, unique sequences with lengths of 18-25 nt were selected to map to specific species precursors in miRBase 21.0 (<http://www.mirbase.org/>) by BLAST search to identify known miRNAs and novel 3p- and 5p-derived miRNAs. Length variation at both the 3' and 5' ends and one mismatch inside of the sequence were allowed in the alignment. The unique sequences mapping to the hairpin arms of mature miRNAs of specific species were identified as known miRNAs. The unique sequences mapping to the arms of known precursor hairpins of specific species opposite to the annotated mature miRNA-containing arms were considered to be novel 5p- or 3p-derived miRNA candidates. The remaining sequences were mapped to other selected species precursors (with the exclusion of specific species) in miRBase 21.0 by BLAST search, and the mapped premiRNAs were further BLAST searched against the specific species genomes to determine their genomic locations. The two above kinds of mapped miRNAs were defined as known miRNAs. The unmapped sequences were BLAST searched against specific genomes, and the hairpin RNA structures containing sequences were predicted from the 120 nt flanking sequences by RNAfold software (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>).

Analysis of differentially expressed miRNAs

The miRNAs with differential expression based on normalized deep-sequencing counts were identified by selectively using Fisher's exact test, Chi-squared 2x2 test, Chi-squared nXn

test, Student's *t* test, or ANOVA based on the experimental design. The significance default threshold was set to $P < 0.05$, which was defined as significantly differentially expressed miRNAs.

Prediction and identification of target genes of miRNAs

To predict the target genes of differentially expressed miRNAs, computational target prediction algorithms (Target Finder) were used to identify miRNA binding sites. The GO terms and KEGG pathways of genes targeted by these differentially expressed miRNAs were annotated. The GO function and KEGG information of this species were used to perform GO function annotation and KEGG signaling pathway annotation for the targeted genes of differentially expressed miRNAs. Fisher's exact hypothesis test was performed on the GO and KEGG pathways of targeted genes, and the enrichment analysis of each GO term and KEGG pathway were separately performed. According to the prediction results of targeted genes of differentially expressed miRNAs, significance differences in the GO and KEGG information was defined with a default threshold $P < 0.05$.

Identification of the expression of miRNAs and their target genes by qRT-PCR

To validate the high-throughput sequencing data, six pairs of miRNA target genes based on the sequencing data were selected to perform qRT-PCR. Total RNA from each sample for qRT-PCR analysis was extracted as described above for the sequencing experiments. First-strand cDNA was synthesized using the Mir-X miRNAs First-Strand Synthesis Kit (TaKaRa Biotech, Dalian, China) according to the manufacturer's instructions. The U6 snRNA was selected as the reference gene, and the qRT-PCR of miRNA experiments was performed in an ABI7500 Real-time PCR instrument (Applied Biosystems, Foster, USA) using the Mir-X miRNA qRT-PCR SYBR Kit (TaKaRa Biotech, Dalian, China) and miRNA-specific primers (Table S1). The qRT-PCR experiments for each sample were analyzed with three replicates.

The predicted target genes of miRNAs were selected for qRT-PCR analyses. First-strand cDNA was synthesized using PrimeScriptTM RT Master Mix (TaKaRa Biotech, Dalian, China) according to the manufacturer's instructions. All specific primer pairs of predicted target genes (Table S2) were designed using Primer Quest online software

(<http://sg.idtdna.com/PrimerQuest/Home/Index>). The elongation factor 1-alpha (*CoEF1α*) was selected as a reference gene. The qRT-PCR experiments were performed using the SYBR Premix Ex TaqTM II Kit (Tli RNaseH Plus; TaKaRa Biotech, Dalian, China). qRT-PCR was conducted by the ABI7500 Real-time PCR instrument (Applied Biosystems, Foster, USA). The qRT-PCR analyses for each sample were performed in triplicate. The relative expression levels of each gene were calculated using the $2^{-\Delta\Delta C_t}$ method [59].

Dual-luciferase reporter assay

Fragments from the 3'UTR of AFB2 containing the predicted binding sequences for cpa-miR393_R-1 were amplified and subcloned into the pmirGLO luciferase promoter vector. The pCDNA3.1 plasmid was used as the template vector. The fragment containing the nucleotide sequences of the precursor of the cpa-miR393_R-1 was cloned into the vector to construct the recombinant vector expressing cpa-miR393_R-1 pCDNA3.1, as described earlier. The pmirGLO vector containing the 3'UTR of AFB2 was cotransfected with pCDNA3.1 or pCDNA3.1 containing pre-cpa-miR393_R-1 using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol and a previous report [60-62]. Forty-eight hours after treatment, the expressed luciferase firefly and Renilla activities were measured using a luciferase reporter assay kit (BioVision, Inc., CA, USA). Renilla was used as a transfection control.

Additional File Legends

Additional file 1: Table S1. Summary of the sequencing data from raw data to cleaned sequences from the seeds of the 'M3' cultivar of tea oil camellia. (XLSX 11 kb)

Additional file 2: Table S2. Summary of the sequencing data from raw data to cleaned sequences from the seeds of the 'M8' cultivar of tea oil camellia. (XLSX 11 kb)

Additional file 3: Table S3. Novel miRNAs in tea oil camellia seeds in this study. (XLSX 15 kb)

Additional file 4: Table S4. Known miRNAs in tea oil camellia seeds in this study. (XLSX 24 kb)

Additional file 5: Table S5. Classification of known and novel miRNAs in tea oil camellia seeds in this study. (XLSX 15 kb)

Additional file 6: Table S6. Length distribution of unique miRNAs. (XLSX 10 kb)

Additional file 7: Table S7. Conservation profiles of the identified miRNAs in tea oil camellia seeds. (XLSX 16 kb)

Additional file 8: Table S8. Family classification of the identified conserved miRNAs in tea oil camellia seeds. (XLSX 12 kb)

Additional file 9: Table S9. Predicted targets for known and novel miRNAs and their functional annotations in tea oil camellia seeds. (XLSX 1,532 kb)

Additional file 10: Table S10. Analysis of the target genes by significant GO enrichment. (XLSX 100 kb)

Additional file 11: Table S11. Analysis of the target genes by significant KEGG enrichment. (XLSX 15 kb)

Additional file 12: Table S12. KEGG enrichment analysis of differentially expressed miRNAs and their target genes related to lipid metabolism. (XLSX 73 kb)

Additional file 13: Table S13. Primers used for qRT-PCR of selected mature miRNAs. (XLSX 10 kb)

Additional file 14: Table S14. Primers used for qRT-PCR of selected target genes of miRNAs. (XLSX 11 kb)

Abbreviations

ACC: Acetyl-CoA carboxylase; ACS1: *Acy*-CoA synthetase-1; AFB2: Auxin signaling F-box 2; AP2: *Apetala2*; ARF: Auxin response factor; BAG7: BCL2-associated athanogene 7; CNR:

Colourless non-ripening; EF1 α : Elongation factor 1 α ; COL13: Collagen type XIII; CPR30: Cytochrome *P450* reductase 30; DGAT: Diacylglycerol O-acyltransferase; ERF15: Ethylene-responsive factor 15; fabG: 3-ketoacyl-acyl carrier protein reductase; FAD2: Fatty acid desaturase 2; FATB1: *Fatty* acyl-ACP thioesterase B-type 1; GAPN: Glyceraldehyde-3-phosphate dehydrogenase; GPD1: Glycerol-3-phosphate dehydrogenase; GRF: Growth-regulating factor; KAR: 3-ketoacyl-ACP reductase; KAS: 3-ketoacyl-ACP synthase; Mcat: Malonyl-CoA-acyl carrier protein transacylase; MED: Mediator Complex Subunit; NRT1.7: Nitrate nitrogen transporter 1.7; PAB2: Polyadenylate-binding protein 2; S-ACP-DES6: secretory acid phosphatase desaturase 6; SAD: Stearoyl -acyl carrier protein desaturase; SBP2: Spherical body protein 2; SCL: Stem cell leukemia; SPL: Squamosa promoter binding protein-like; SPL10: SBP-like Protein10; T1R1: *Taste* receptor T1R1; Δ 9D: Delta9-ACP desaturase

Declarations

Acknowledgements

The authors thank Wei Du, Sihei Liu and He Huang for their assistance in collecting plant materials.

Funding

This work was supported by the Key Technology Research and Development Program of Guizhou Province ([2019]2404, [[2018]5252, [2018]4016, [2017]2521, [2016]2517-1 and [2016]2519).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

Bo Wu performed the experiments of this study and wrote the manuscript. Chengjing Ruan conceived this study and designed the experiments of this study. Wanchen Zhang performed the experiments of this study. Asad Hussain Shah helped designed the experiments of this study and revised the manuscript. Sihei Liu collected the samples of this study. All authors read and approved the manuscript in its final form.

Ethics approval and consent to participate

Not applicable.

Consent for publish

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

References

1. Ohlrogge JB. Design of new plant products: engineering of fatty acid metabolism. *Plant Physiol.* 1999; 104: 821–826.
2. Li H, Zhou GY, Zhang HY, Liu JA. Research progress on the health function of tea oil. *J Med Plants Res.* 2011; 5: 485–489.
3. Lee CP, Yen GC. Antioxidant activity and bioactive compounds of tea seed (*Camellia oleifera*) oil. *J Agric Food Chem.* 2006; 54: 779–784.
4. Yu Y, Ren S, Tan K. Study on climatic regionalization and layer and belt distribution of oiltea *Camellia* quality in China. *J Nat Res.* 1999; 14: 123–127.

5. Chen J, Yang X, Huang X, Duan S, Long C, Chen J, et al. Leaf transcriptome analysis of a subtropical evergreen broadleaf plant, wild oil-tea camellia (*Camellia oleifera*), revealing candidate genes for cold acclimation. BMC Genomics. 2017; 18:211.
6. Yin DD, Li SS, Shu QY, Gu ZY, Wang LS. Identification of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony seeds. Gene. 2018; 666: 72-82.
7. Cao S, Zhu QH, Shen W, Jiao X, Zhao X, Wang MB, et al. Comparative profiling of microRNA expression in developing seeds of high linoleic and high oleic safflower (*Carthamus tinctorius*) plants. Front Plant Sci. 2013; 4: 489.
8. Niu J, Wang J, An J, Liu L, Lin Z, Wang R, et al. Integrated mRNA and miRNA transcriptome reveal a cross-talk between developing response and hormone signaling for the seed kernels of *Siberian apricot*. Sci Rep. 2016;6: 35675.
9. Mutum RD, Kumar S, Balyan S, Kansal S, Mathur S, Raghuvanshi S. Identification of novel miRNAs from drought tolerant rice variety Nagina 22. Sci Rep. 2016;6: 30786.
10. Subramanian S, Steer CJ. MicroRNAs as gatekeepers of apoptosis. J Cell Physiol. 2010; 223: 289-298
11. Cuperus JT, Fahlgren N, Carrington JC. Evolution and functional diversification of miRNA genes. Plant Cell. 2011;23: 431-442
12. Griffiths-Jones S, Saini HK, Dongen SV, Enright AJ. miRbase: tools for microRNA genomics. Nucl Acid Res. 2008;36: D154-D158.
13. Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, et al. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol. 2011;11: 5.
14. Reyes JL, Chua NH. ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. Plant J. 2007; 49: 592-606.
15. Körbes AP, Machado RD, Guzman F, Mauricio Pereira Almerão, Oliveira LFVD, Loss-MoraisG, et al. Identifying conserved and novel microRNAs in developing seeds of *Brassica napus* using deep sequencing. PLOS One. 2012;7: e50663.

16. Peng T, Sun H, Qiao M, Zhao Y, DuY. Differentially expressed microRNA cohorts in seed development may contribute to poor grain filling of inferior spikelets in rice. *BMC Plant Biol.* 2014; 14: 196.
17. Jin Q, Xue Z, Dong C, Wang Y, Chu L, Xu Y. Identification and characterization of microRNAs from tree peony (*Paeonia ostii*) and their response to copper stress. *Plos One.* 2015;10: e0117584.
18. Li D, Wang L, Liu X, Cui D, Chen T, Zhang H. Deep sequencing of maize small RNAs reveals a diverse set of microRNA in dry and imbibed seeds. *PLOS One.* 2013;8(1): e55107.
19. Ding J, Ruan C, Guan Y. Identification of microRNAs involved in lipid biosynthesis and seed size in developing sea buckthorn seeds using high-throughput sequencing. *Sci Rep.* 2018;8: 4022.
20. Belide S, Petrie JR, Shrestha P, Singh SP. Modification of seed oil composition in *Arabidopsis* by artificial microRNA-mediated gene silencing. *Front Plant Sci.* 2012; 3: 168.
21. Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton M, et al. MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. *EMBO J.* 2014; 23: 3356-3364.
22. Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, et al. Enhanced seed oil production in Canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol.* 2011; 156: 1577-1588.
23. Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, et al. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell.* 2003; 15: 5-18.
24. Manju G, Pudota BB, Shreedharan S, Wang PS. Integration of omics approaches to understand oil/protein content during seed development in oilseed crops. *Plant Cell Rep.* 2016; 36: 1-16.
25. Schommer C, Palatnik JF, Aggarwal P, Chételat A, Cubas P, Farmer EE, et al. Control of jasmonate biosynthesis and senescence by miR319 targets. *PLOS Biol.* 2008; 6: e230.

26. Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Carrington J C. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. *Plant J.* 2010; 52: 133-146.
27. Nodine MD, Bartel DP. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Gene Dev.* 2010; 24: 2678-2692.
28. Allen RS, Li J, Stahle MI, Aurelie D, Gubler F, Millar AA. Genetic analysis reveals functional redundancy and the major target genes of the *Arabidopsis* miR159 family. *Proc Natl Acad Sci USA.* 2007; 104: 16371-16376.
29. Harikrishna K, Othman RY, Hoon LS, Harikrishna JA. Computational prediction of microRNAs from oil palm (*Elaeis guineensis*) expressed sequence tags. *Asia-Pac J Mol Biol Biotechnol.* 2007; 15: 107-113.
30. Donaire LL, Pedrola LL, Rosa RDL, Liave C. High-throughput sequencing of RNA silencing-associated small RNAs in olive (*Olea europaea*). *PLOS One.* 2011; 6: e27916.
31. Ding J, Ruan CJ, Du W, Guan Y. RNA-seq data reveals a coordinated regulation mechanism of multigenes involved in the high accumulation of palmitoleic acid and oil in sea buckthorn berry pulp. *BMC Plant Biol.* 2019; 19: 207.
32. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA. Conservation and divergence of plant microRNA genes. *Plant J Cell Mol Biol.* 2010; 46: 243-259.
33. Weber MJ. New human and mouse microRNA genes found by homology search. *Febs J.* 2010; 272: 59-73.
34. Bonnet E, Wuyts J, Rouzé P, Van d PY. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *PNAS.* 2004; 101: 11511-11516.
35. Baker CC, Sieber P, Wellmer F, Meyerowitz EM. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in *Arabidopsis*. *Cur Biol.* 2005; 15: 303-315.
36. Zhao YT, Wang M, Fu SX, Yang WC, Qi CK, Wang XJ. Small RNA profiling in two *Brassica napus* cultivars identifies microRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant Physiol.* 2012; 158: 813-823.

37. Huang D, Koh C, Feurtado JA, Tsang EW, Cutler AJ. MicroRNAs and their putative targets in *Brassica napus* seed maturation. BMC Genomics. 2013; 14: 140-140.
38. Dörmann P, Voelker TA, Ohlrogge JB. Accumulation of palmitate in *Arabidopsis* mediated by the acyl-acyl carrier protein thioesterase *FATB1*. Plant Physiol. 2000; 123: 637-643.
39. Yasuno R, von Wettstein-Knowles P, Wada H. Identification and molecular characterization of the beta-Ketoacyl-[Acyl Carrier Protein] synthase component of the *Arabidopsis* mitochondrial fatty acid synthase. J Biol Chem. 2004; 279: 8242-8251.
40. Andre C, Haslam RP, Shanklin J. Feedback regulation of plastidic acetyl-CoA carboxylase by 18:1-acyl carrier protein in *Brassica napus*. PNAS. 2012; 109: 10107-10112.
41. Nakkaew A, Chotigeat W, Eksomtramage T, Phongdara A. Cloning and expression of a plastid-encoded subunit, beta-carboxyltransferase gene (*accD*) and a nuclear-encoded subunit, biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis*). Plant Sci. 2008; 175: 497-504.
42. Wang B, Tan X, Jiang J, Zhang L. Molecular cloning and expression of two genes encoding accase subunits of *Camellia oleifera* (Theaceae). Pakistan Journal of Botany. 2018; 50: 103-110.
43. Zhang DQ, Tan XF, Hu FM. The cDNA cloning and characteristic of stearyl-ACP desaturase gene of *Camellia oleifera*. Acta Hort. 2008; 769: 55-61.
44. Xia EH, Jiang JJ, Huang H, Li X, Huang Y. Transcriptome analysis of the oil-rich tea plant, *Camellia oleifera*, reveals candidate genes related to lipid metabolism. PLOS One. 2014; 9: e104150.
45. Zeng Y, Tan X, Zhang L, Jiang N, Cao H. Identification and expression of fructose-1,6-bisphosphate aldolase genes and their relations to oil content in developing seeds of tea oil tree (*Camellia oleifera*). PLOS One. 2014; 9: e107422.
46. Xu R, Li Y. Control of final organ size by mediator complex subunit 25 in *Arabidopsis thaliana*. Development. 2011; 22: 4545-4554.

47. Gupta M, Bhaskar PB, Sriram S, Sriram S, Wang PH. Integration of omics approaches to understand oil/protein content during seed development in oilseed crops. *Plant Cell Rep.* 2016; 36: 1-16.
48. Guo M, Simmons CR. Cell number counts--the fw2.2 and CNR genes and implications for controlling plant fruit and organ size. *Plant Sci.* 2011; 181: 1-7.
49. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell.* 2009; 138: 750-759.
50. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. Specific effects of microRNAs on the plant transcriptome. *Dev Cell.* 2005; 8: 517-527.
51. Sunkar R, Zhu JK. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell.* 2004; 16: 2001-2019.
52. Llave C, Carrington JC. Cleavage of scarecrow-like mRNA targets directed by a class of *Arabidopsis* Science. 2002; 297: 2053-2056.
53. Liu HH, Tian X, Li YJ, Wu CA, Zheng CC. Microarray-based analysis of stress-regulated microRNA in *Arabidopsis thaliana*. *RNA.* 2008; 14: 836-843.
54. Jian X, Zhang L, Li G, Zhang L, Wang XJ, Cao XF. Identification of novel stress-regulated microRNAs from *Oryza sativa* Genomics. 2010; 95: 47-55.
55. Shen JQ, Xie KB, Xiong LZ. Global expression profiling of rice microRNAs by one-tube stem-loop reverse transcription quantitative PCR revealed important roles of microRNAs in abiotic stress responses. *Mol Genet Genomics.* 2010; 284: 477.
56. Wang JW, Czech B, Weigel D. miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell.* 2009; 138: 738-749.
57. Ni Z, Hu Z, Jiang Q, Zhang H. Overexpression of gma-MIR394a confers tolerance to drought in transgenic *Arabidopsis thaliana*. *Biochem Biophys Res Commun.* 2012; 427: 330-335.
58. Fan SC, Lin CS, Hsu PK, Lin SH, Tsay YF. The *Arabidopsis* nitrate transporter NRT1.7, expressed in phloem, is responsible for source-to-sink remobilization of nitrate. *Plant*

Cell. 2009; 21: 2750-2761.

59. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protocols. 2008; 3: 1101-1108.
60. Agatheeswaran S, Pattnayak NC, Chakraborty S. Identification and functional characterization of the miRNA-gene regulatory network in chronic myeloid leukemia lineage negative cells. Sci Rep. 2016; 6: 32493.
61. Kumari B, Jain P, Das S, Ghosal S, Hazra B, Trivedi AC, et al. Dynamic changes in global microRNAome and transcriptome reveal complex miRNA-mRNA regulated host response to *Japanese encephalitis* Virus in microglial cells. Sci Rep. 2016; 6: 20263.
62. Qian P, Jiang T, Wang X, Song F, Shen X. Bmo-miR-275 down-regulates expression of *Bombyx mori* sericin gene 2 in vitro. PLOS ONE. 2018; 13: e0190464.

Tables

Table 1 Significantly differential expressed miRNAs between seeds of ‘M3’ and ‘M8’ cultivars of tea oil camellia at four development stages

Comparison groups	miRNA name	P-value (t-test)	Up/Down
M3T1 vs. M8T1	mtr-MIR2586a-p3_1ss15TC	2.85E-03	Up
	ath-MIR5645b-p5_2ss19TG21TC	1.15E-02	Up
	ath-MIR5645b-p3_2ss19TG21TC	1.15E-02	Up
	mtr-MIR2586a-p3_2ss4TG21CT	1.94E-02	Up
	hpe-miR166a_1ss5AC	3.35E-02	Up
M3T2 vs. M8T2	rco-miR156e_L+1R-1	8.38E-03	Down
	csi-miR1515_1ss16AG	1.41E-02	Up
	stu-miR8016_1ss7GT	1.63E-02	Up
	dpr-miR167a_R+1	1.86E-02	Down
	ccl-miR167a	2.53E-02	Down
	tcc-miR390a	3.52E-02	Down
	ath-miR390a-5p	3.52E-02	Down
	aly-miR167b-3p_1ss13AG	3.86E-02	Down
	ppe-miR172a-5p_1ss21GA	4.49E-02	Up
	han-miR156a_L+1	4.92E-02	Down
M3T3 vs. M8T3	PC-3p-9080_426	1.36E-02	Down
	mtr-MIR2586a-p3_1ss15TC	1.49E-02	Down
M3T4 vs. M8T4	ghr-MIR7499-p3_2ss10GC17CG	1.24E-02	Down
	rco-miR167a_R+1	4.23E-02	Down

Table 2 Predicted miRNAs and their target genes involved in lipid metabolism

miRNA ID	Gene ID	Gene name	Gene Annotation	KEGG Pathway
tcc-miR162	comp59939_c1	<i>ACC1</i>	acetyl-CoA carboxylase / biotin carboxylase	Pyruvate metabolism
csi-miR166e-5p	comp67779_c0	<i>KAS1</i>	3-oxoacyl-[acyl-carrier-protein] synthase I	
csi-miR166e-5p	comp67006_c0	<i>S-ACP-DES6</i>	acyl-[acyl-carrier-protein]- desaturase	
rgl-miR5139_L-1	comp61049_c0	<i>KAS3B</i>	3-oxoacyl-[acyl-carrier-protein] synthase III	
mtr-miR156h-3p_1ss8AC	comp63911_c0	<i>fabG</i>	3-oxoacyl-[acyl-carrier protein] reductase	Fatty acid biosynthesis metabolic
cpa-miR164d	comp63026_c1	<i>Mcat</i>	[acyl-carrier-protein] S-malonyltransferase	
PC-5p-21064_221	comp67050_c0	<i>FATB1</i>	fatty acyl-ACP thioesterase B	
aly-miR393a-3p_1ss12TC	comp52017_c0	<i>MOD1</i>	enoyl-[acyl-carrier protein] reductase I	
ath-miR172a	comp48800_c0	<i>Δ9D</i>	acyl-[acyl-carrier-protein] desaturase	Glycolysis / Gluconeogenesis
gma-MIR5368-p3_1ss21CT	comp67185_c0	<i>GAPN</i>	glyceraldehyde-3-phosphate dehydrogenase (NADP)	

Table 3 Predicted miRNAs and their target genes involved in seed size

miRNA ID	Gene ID	Gene name	Gene Annotation
vvi-miR160c_L+1R-1	comp58420_c0	<i>ARF22</i>	auxin response factor 22
htu-miR160a_R+1	comp58420_c0	<i>ARF22</i>	auxin response factor 22
hpe-miR162a_L-2	comp57222_c0	<i>ARF19</i>	auxin response factor 19
tcc-miR162	comp57222_c0	<i>ARF19</i>	auxin response factor 19
mdm-miR167h_1ss22AT	comp60627_c0	<i>ARF6</i>	auxin response factor 6
dpr-miR167a_R+1	comp60627_c0	<i>ARF6</i>	auxin response factor 6
PC-3p-16403_272	comp57254_c0	<i>ARF12</i>	auxin response factor 12
mdm-miR167h_1ss22AT	comp57254_c0	<i>ARF12</i>	auxin response factor 12
aly-miR170-5p_1ss8CA	comp66127_c0	<i>ARF11</i>	auxin response factor 11
PC-3p-145806_25	comp67524_c1	<i>ARF1</i>	auxin response factor 1
mtr-miR156h-3p_1ss8AC	comp55270_c0	<i>MYB44</i>	Transcription factor MYB44
ptc-miR172b-5p	comp61335_c0	<i>MYB44</i>	Transcription factor MYB44
ath-miR858b	comp31235_c0	<i>MYB44</i>	Transcription factor MYB44
PC-5p-21064_221	comp63079_c0	<i>MYB44</i>	Transcription factor MYB44
crt-miR168_L-2	comp59139_c0	<i>MYB44</i>	Transcription factor MYB44
csi-miR166e-5p_L-1R+1_1ss21GA	comp278197_c0	<i>MYB82</i>	Transcription factor MYB82
ath-miR858b	comp278197_c0	<i>MYB82</i>	Transcription factor MYB82
ath-miR858b	comp65157_c0	<i>MYB3</i>	Transcription factor MYB3
mdm-miR167h_1ss22AT	comp135267_c0	<i>MED27</i>	Mediator of RNA polymerase II transcription subunit 27
gra-miR167a_L-2R+2	comp54724_c0	<i>MED28</i>	Mediator of RNA polymerase II transcription subunit 28
mdm-MIR167c-p3_1ss6AG	comp63676_c0	<i>MED14</i>	Mediator of RNA polymerase II transcription subunit 14
mes-MIR482-p3_2ss9TG21AG	comp50462_c0	<i>MED18</i>	Mediator of RNA polymerase II transcription subunit 18
PC-5p-315953_8	comp59426_c0	<i>CNR2</i>	Cell number regulator 2

Table 4 Predicted miRNAs and their target genes involved in growth and development

miRNA ID	Gene ID	Gene name	Gene Annotation
ccl-miR171	comp61081_c0	<i>SCL16</i>	Scarecrow-like protein 6
ssl-miR171b_2ss11CT19AT	comp61081_c0	<i>SCL16</i>	Scarecrow-like protein 6
han-miR156a_L+1	comp60372_c0	<i>SPL4</i>	Squamosa promoter-binding-like protein 4
han-miR156a_L+1	comp62492_c0	<i>SBP2</i>	Squamosa promoter-binding protein 2
rco-miR156e_L+1R-1	comp66521_c0	<i>ACS1</i>	1-aminocyclopropane-1-carboxylate synthase
rco-miR156e_L+1R-1	comp62492_c0	<i>SBP2</i>	Squamosa promoter-binding protein 2
rco-miR156e_L+1R-1	comp61590_c1	<i>SBP1</i>	Squamosa promoter-binding protein 1
mtr-miR156h-3p_1ss8AC	comp65957_c0	<i>TCP24</i>	Transcription factor TCP24
cpa-miR393_R-1	comp61948_c0	<i>T1R1</i>	transport inhibitor response 1
cpa-miR393_R-1	comp64823_c0	<i>AFB2</i>	Protein AUXIN SIGNALING F-BOX 2
PC-3p-9080_426	comp63752_c0	<i>COL13</i>	Zinc finger protein CONSTANS-LIKE 13
PC-3p-9080_426	comp58165_c0	<i>ERF115</i>	EREBP-like factor

Table 5 Predicted miRNAs and their target genes involved in stress-response

miRNA ID	Gene ID	Gene name	Gene annotation
PC-3p-16403_272	comp14497_c0	BAM1	leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1
PC-3p-16403_272	comp47735_c0	Os03g0214200	Ninja-family protein Os03g0214200
PC-5p-28629_169	comp66741_c0	ERF106	EREBP-like factor
ath-miR858b	comp31687_c0	MYB4	Myb-related protein Myb4
ath-MIR5645b-p5_2ss19TG21TC	comp59571_c0	PYL1	Absciscic acid receptor PYL1
mtr-miR156h-3p_1ss8AC	comp63682_c0	BAG7	BAG family molecular chaperone regulator 7
mtr-miR156h-3p_1ss8AC	comp66713_c0	NRT1.7	proton-dependent oligopeptide transporter, POT family
mdm-MIR167c-p3_1ss6AG	comp56349_c0	XDH1	xanthine dehydrogenase/oxidas
gra-miR167a_L-2R+2	comp62927_c0	APUM5	Pumilio homolog 5
gma-miR394a-5p_R+1_1ss18TA	comp64503_c0	SGS3	Protein SUPPRESSOR OF GENE SILENCING 3
nta-MIR6149a-p5_2ss18CT21TA	comp65951_c0	CPR30	F-box protein CPR30
han-miR156a_L+1	comp56544_c0	SPL16	Squamosa promoter-binding-like protein 16
ppe-miR172a-5p_1ss21GA	comp65585_c0	PAB2	20S proteasome subunit alpha 2

Figures

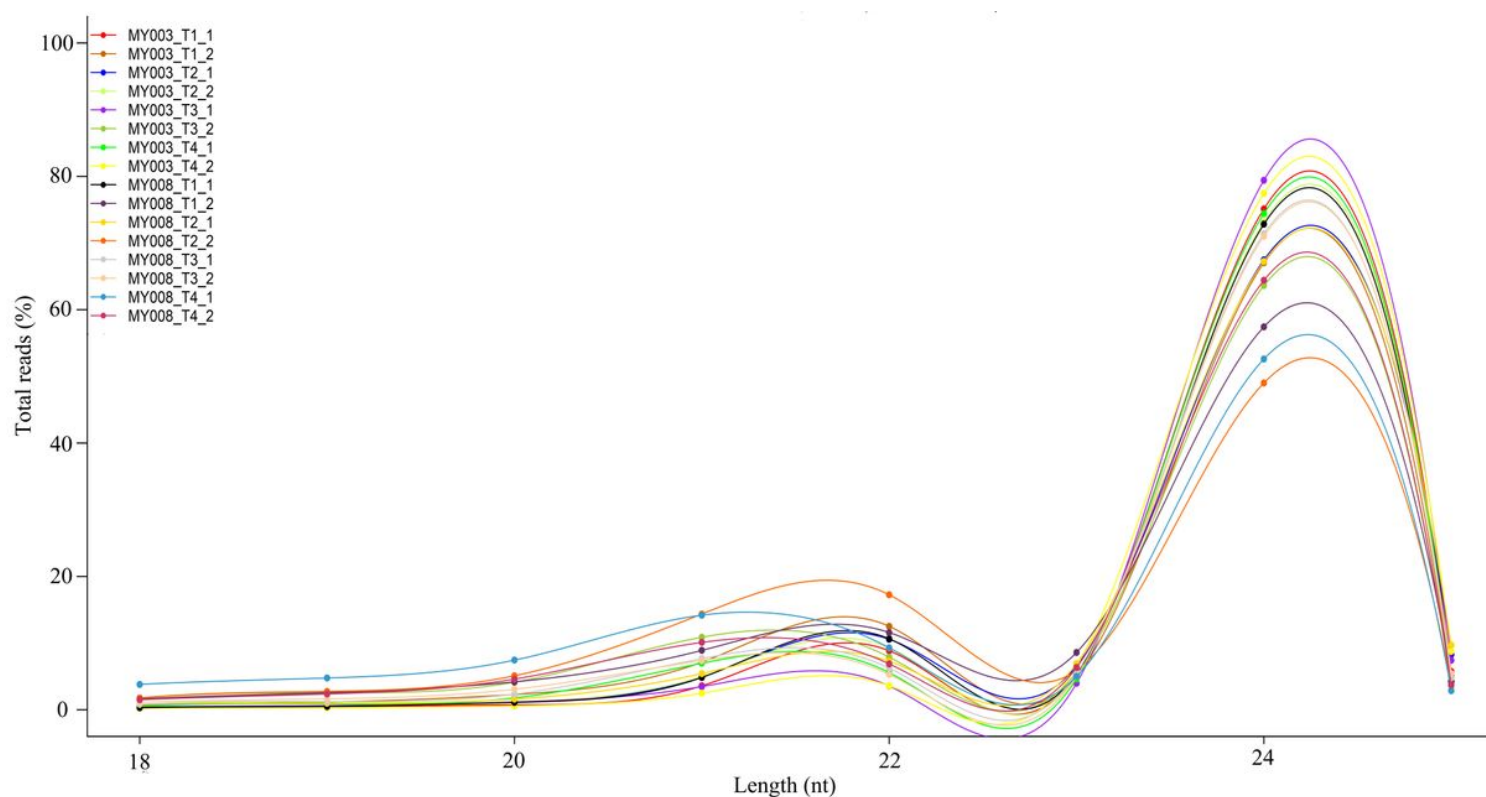


Figure 1

Length distribution of the valid data statistics from the seeds of the 'M3' and 'M8' cultivars of tea oil camellia at four different developmental stages.

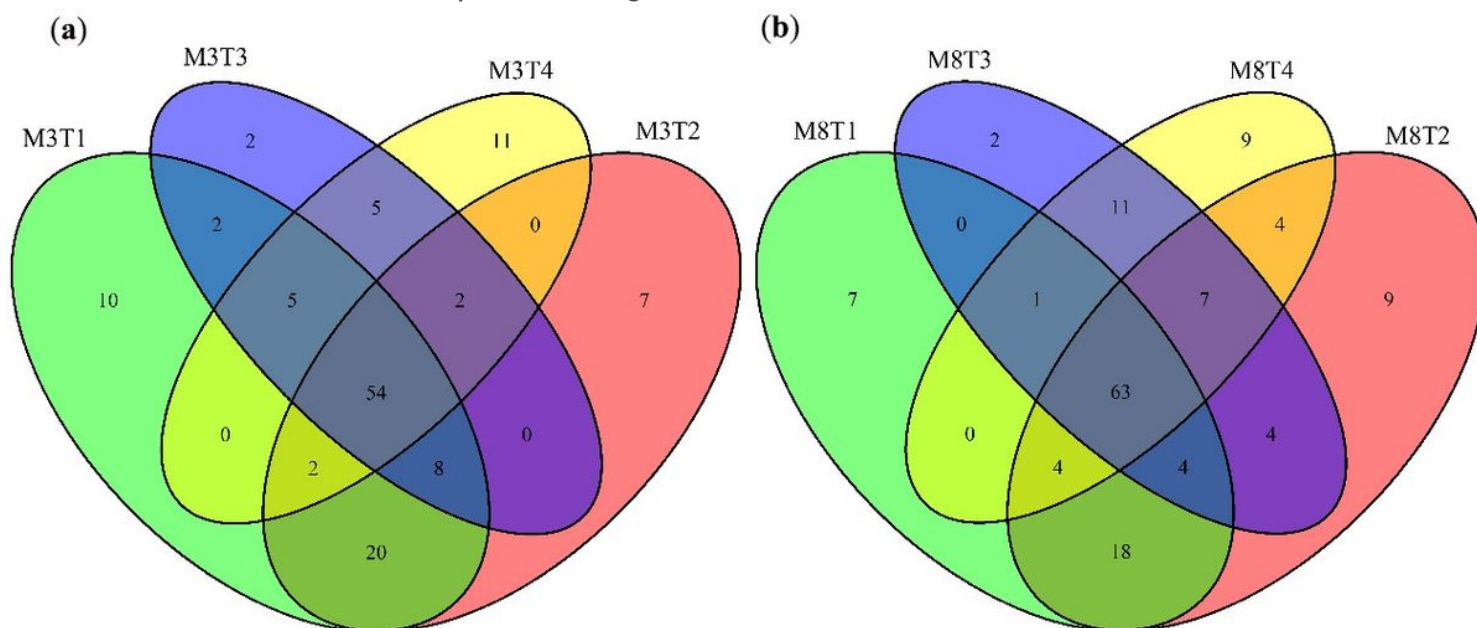


Figure 2

Venn diagrams of the detected miRNAs among the seeds of 'M3' (a) and 'M8' (b) at four different developmental stages.

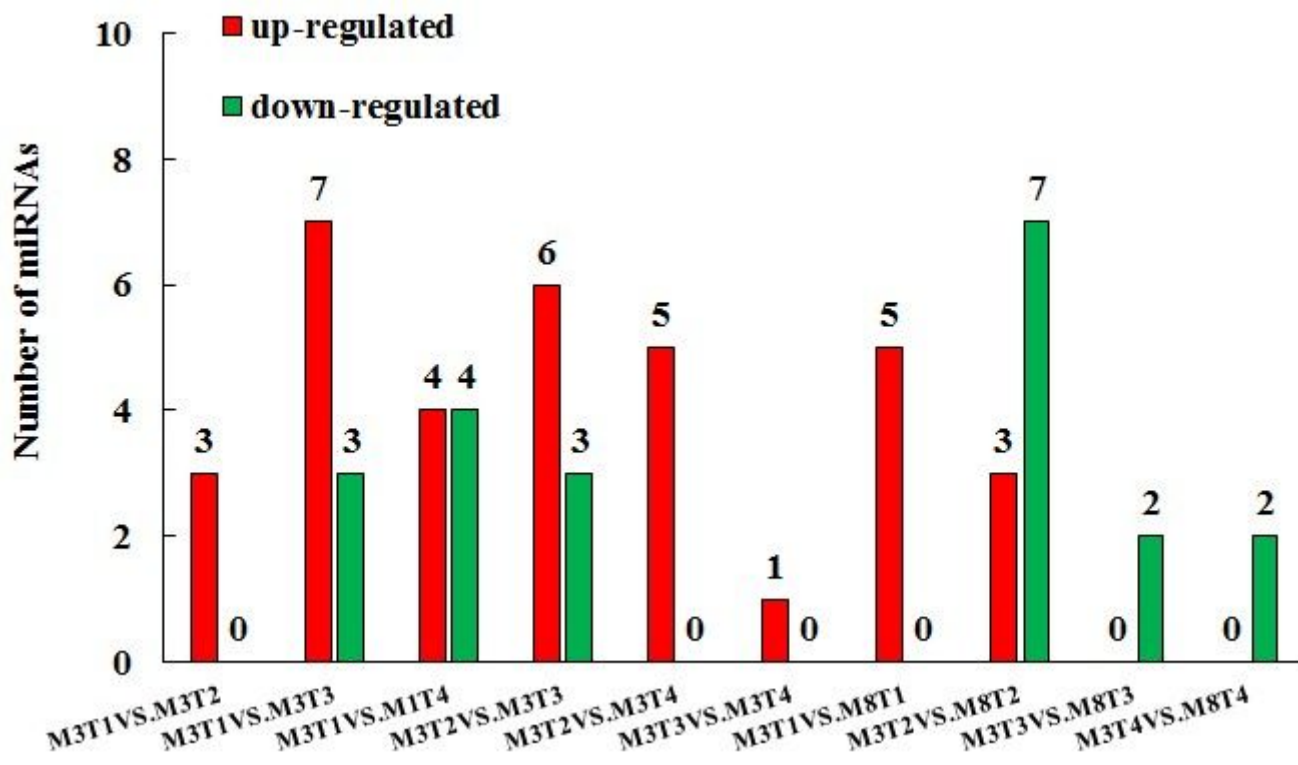


Figure 3

Bar plot of differentially expressed miRNAs among different groups. Note: the number of upregulated (red) and downregulated (green) miRNAs.

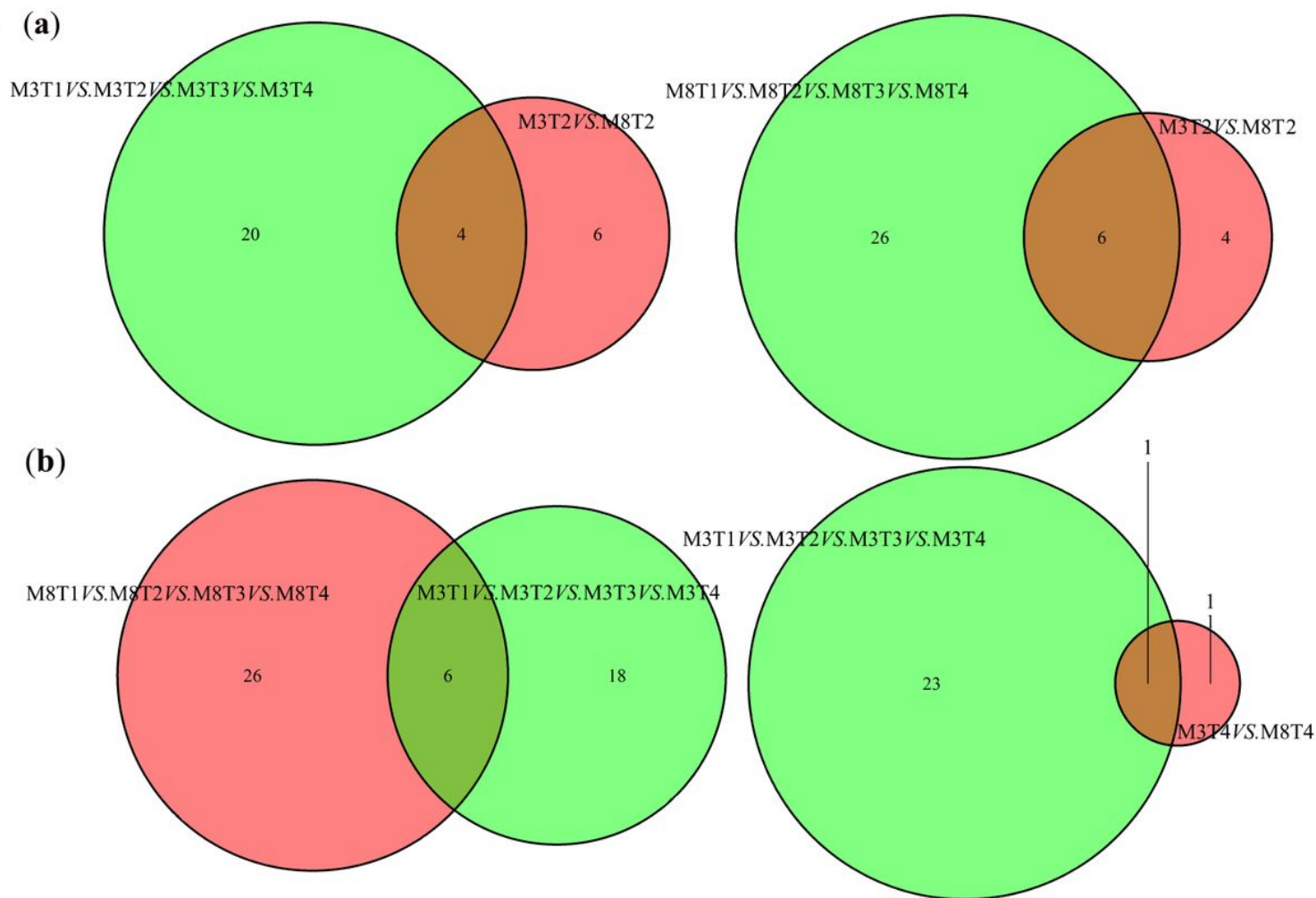


Figure 4

Venn diagrams of differentially expressed miRNAs between different comparison groups. a Distribution of a Venn diagram of differentially expressed miRNAs in M3T1 vs. M3T2 vs. M3T3 vs. M4T4 compared to M3T2 vs. M8T2 and M8T1 vs. M8T2 vs. M8T3 vs. M8T4 compared to M3T2 vs. M8T4. b Distribution of a Venn diagram of differentially expressed miRNAs in M3T1 vs. M3T2 vs. M3T3 vs. M4T4 compared to M8T1 vs. M8T2 vs. M8T3 vs. M8T4 and M3T1 vs. M3T2 vs. M3T3 vs. M4T4 compared to M3T4 vs. M8T4.

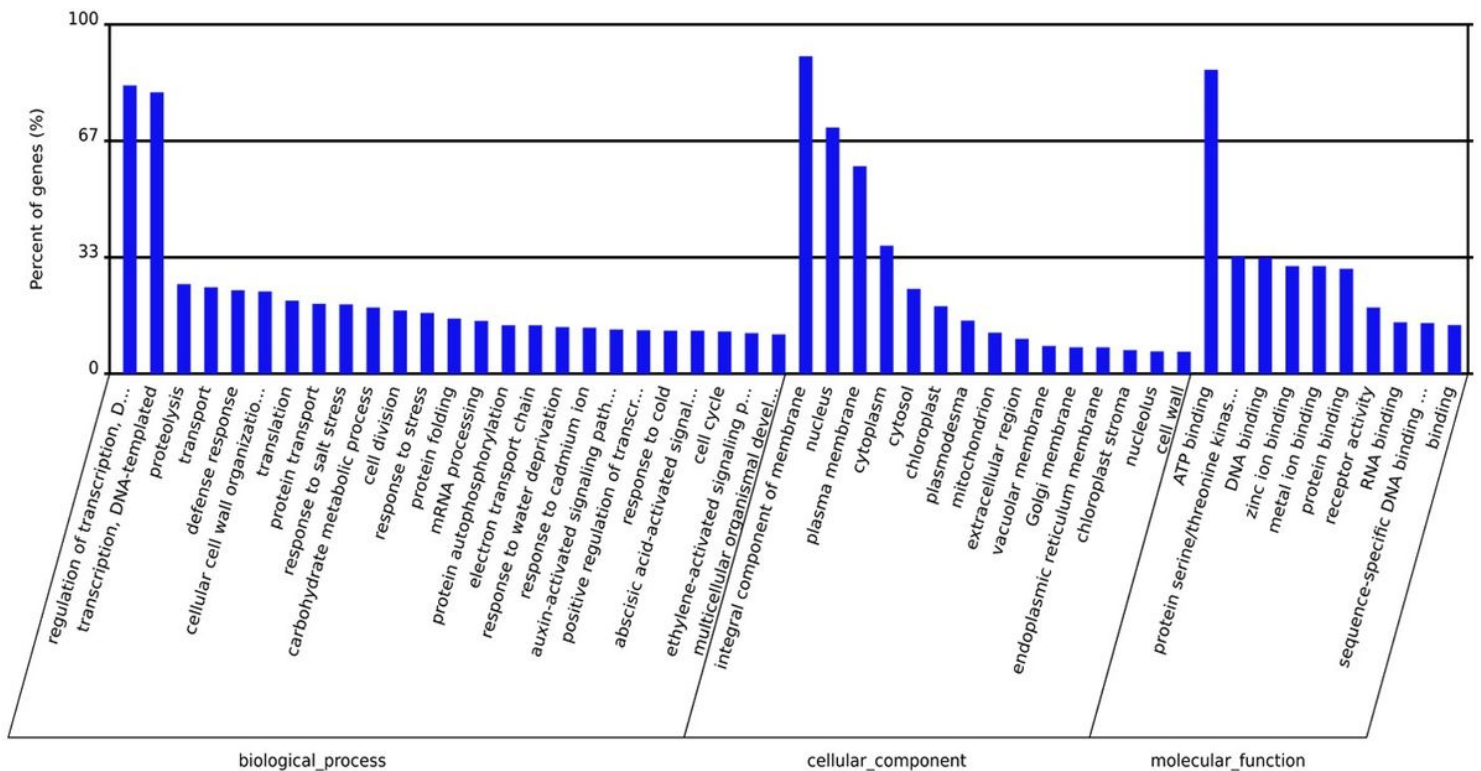


Figure 5

GO categories and distribution of genes targeted by identified miRNAs in the seeds of tea oil camellia.

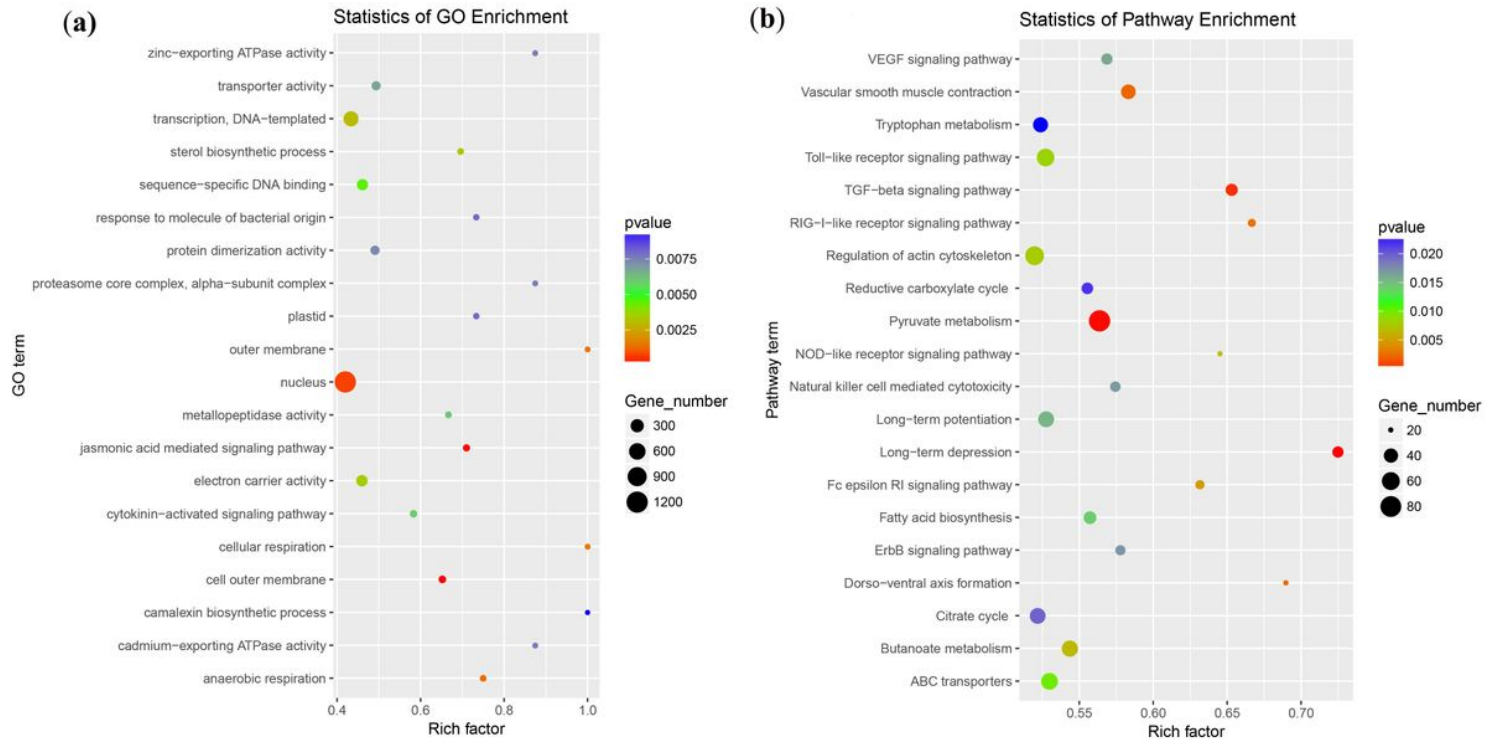


Figure 6

Analyses of significant GO enrichment (a) and KEGG enrichment (b) scatterplot distribution map based on miRNA targets.

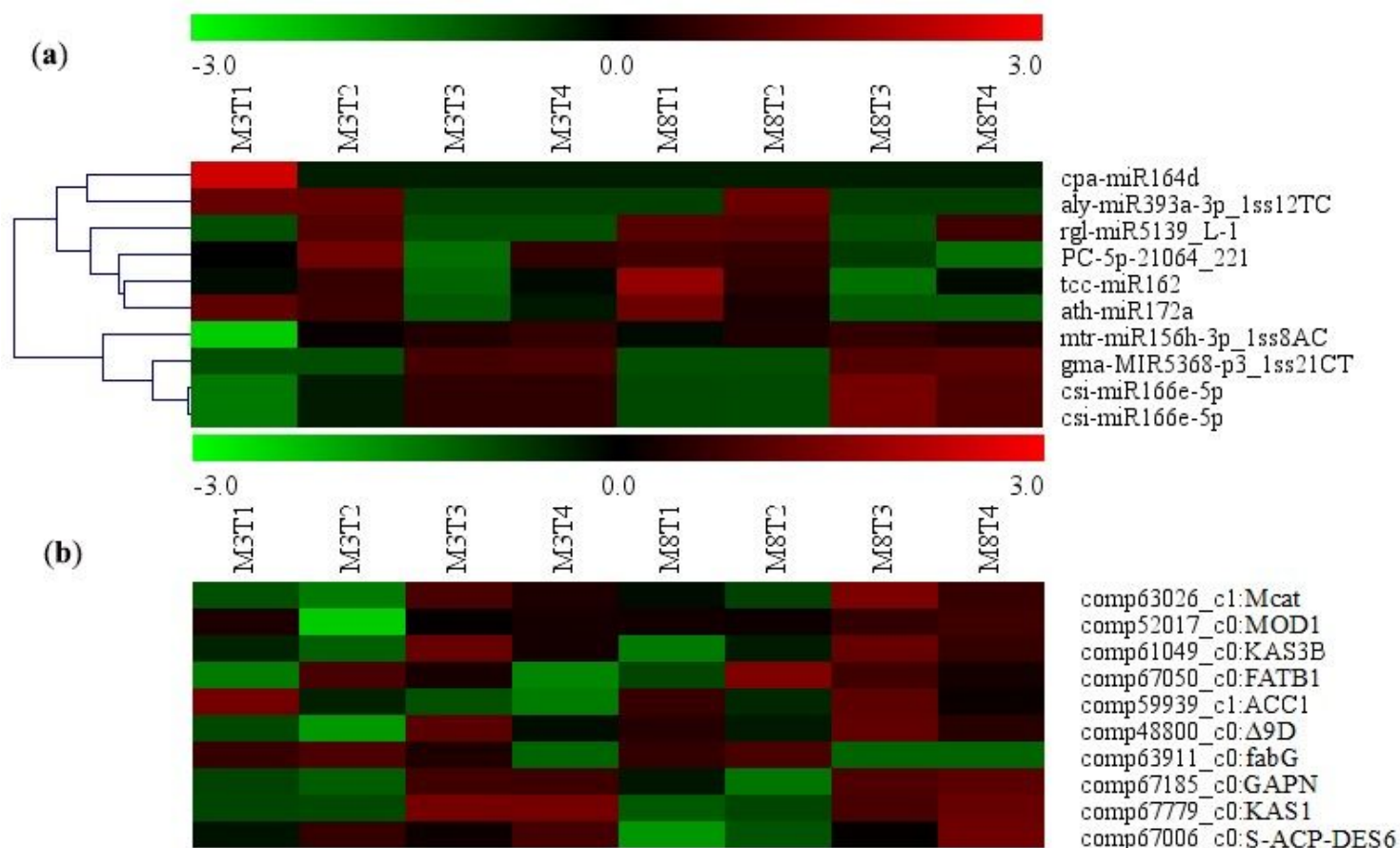


Figure 7

Heat map analysis of predicted differentially expressed miRNAs (a) and their target genes (b) involved in lipid metabolism from the seeds of the 'M3' and 'M8' cultivars at four developmental stages. Red and green indicate high levels and low levels, respectively, of miRNAs and their targets.

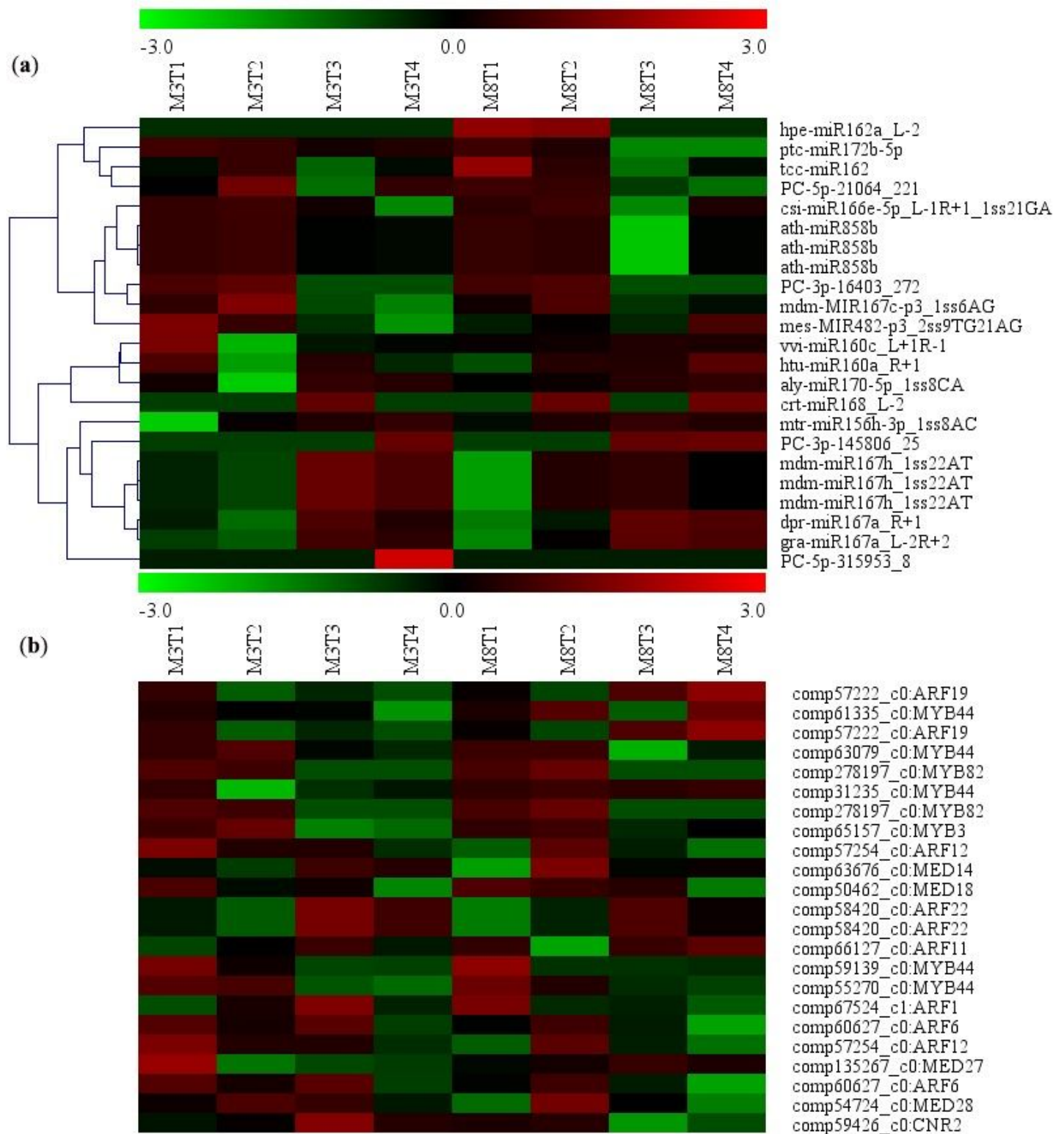


Figure 8

Heat map analysis of predicted differentially expressed miRNAs (a) and their target genes (b) involved in seed size from seeds of the 'M3' and 'M8' cultivars at four developmental stages. Red and green indicate high levels and low levels, respectively, of miRNAs and their targets.

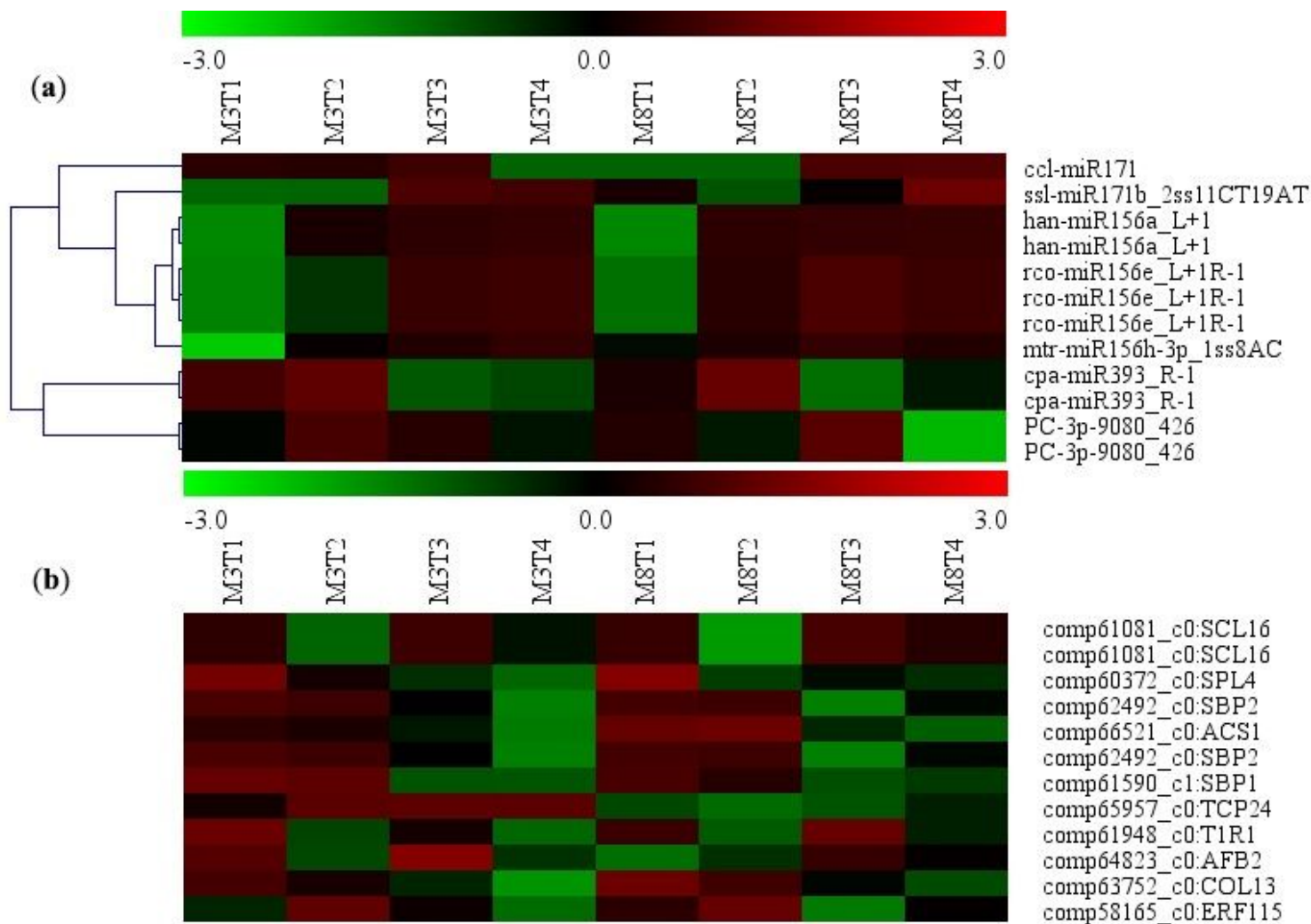


Figure 9

Heat map analysis of predicted differentially expressed miRNAs (a) and their target genes (b) involved in the growth and development of seeds from the 'M3' and 'M8' cultivars at four developmental stages. Red and green indicate high levels and low levels, respectively, of miRNAs and their targets.

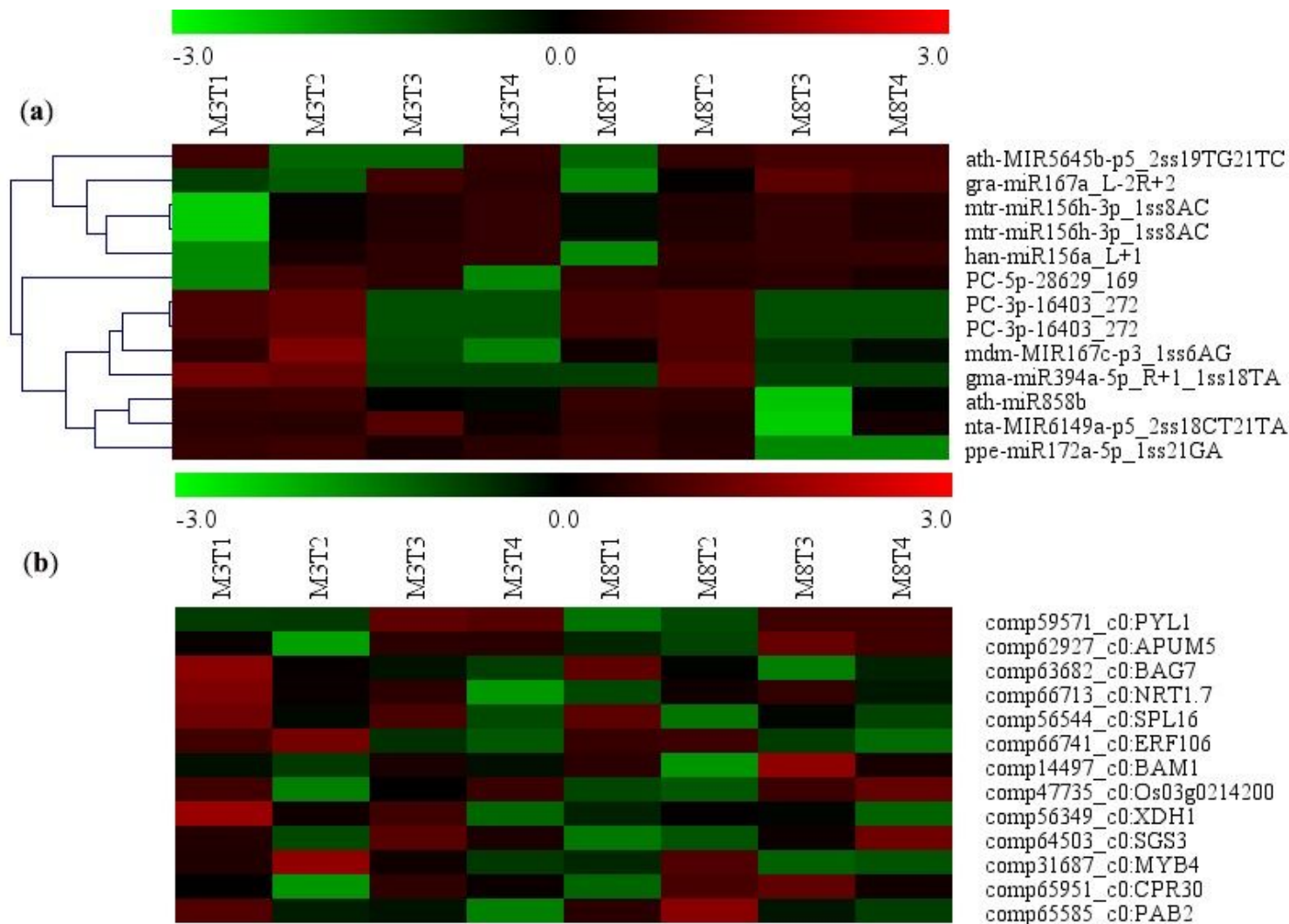


Figure 10

Heat map analysis of predicted differentially expressed miRNAs (a) and their target genes (b) related to resistance from the seeds of the 'M3' and 'M8' cultivars at four developmental stages. Red and green indicate high levels and low levels, respectively, of miRNAs and their targets.

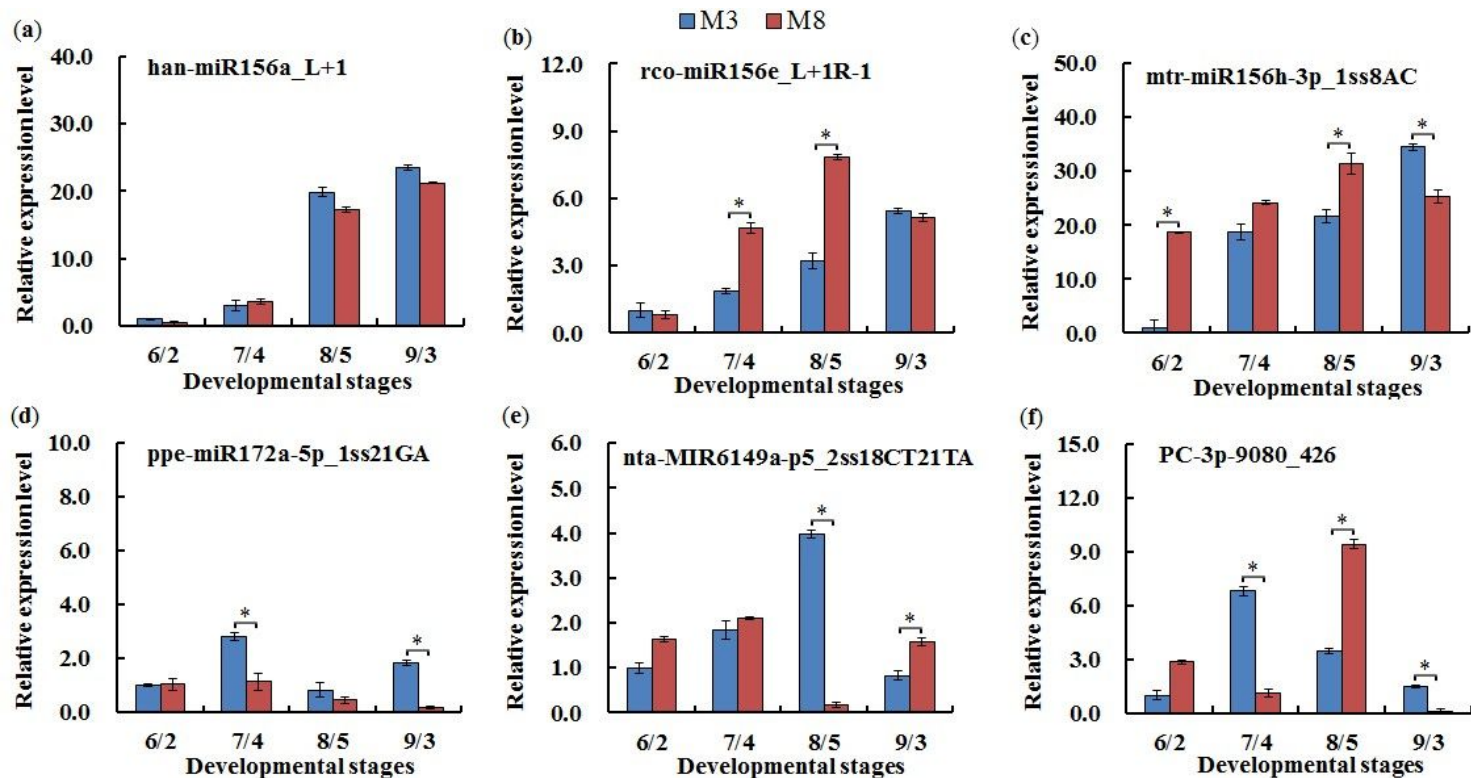


Figure 11

Expression differences of selected miRNAs in the seeds of the 'M3' and 'M8' cultivars of tea oil camellia at four different developmental stages. a han-miR156a_L+1. b rco-miR156e_L+1R-1. c mtr-miR156h-3p_1ss8AC. d ppe-miR172a-5p_1ss21GA. e nta-MIR6149a-p5_2ss18CT21TA. f PC-3p-9080_426. * indicates a significant difference in relative gene expression between the two cultivars at the same harvest time at the level of 0.05.

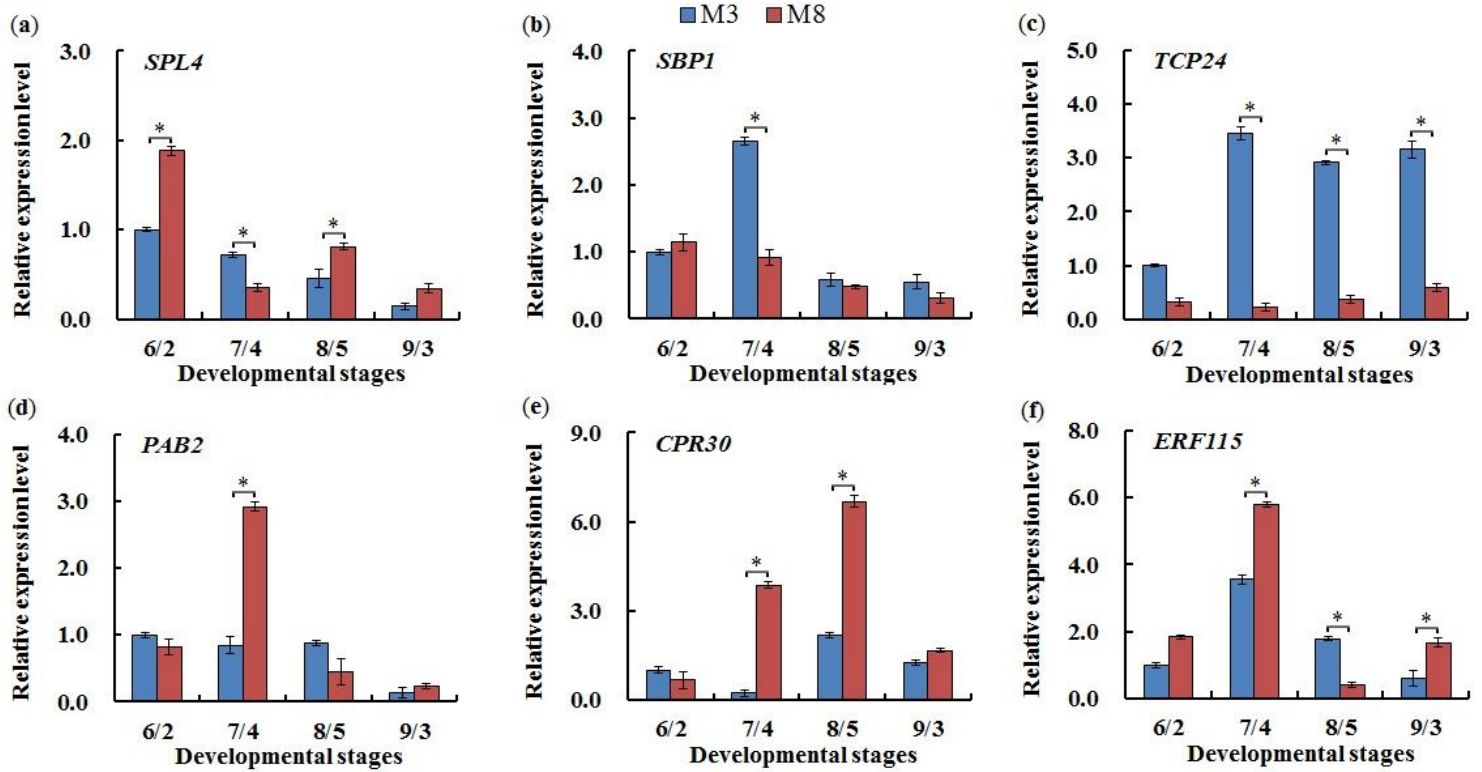


Figure 12

Expression differences of the 6 selected predicted targets in the seeds of the 'M3' and 'M8' cultivars at four different developmental stages. a SPL4. b SBP1. c TCP24. d PAB2. e CPR30. f ERF115. * indicates a significant difference in the relative gene expression between the two cultivars at the same harvest time at the level of 0.05.

a cpa-miR393_R-1 1 TCCAAAGGGATCGCATTG-ATC 21
 |||||
 AFB2 1579 AGGTTTCCCTAGCGTAACAAAG 1558

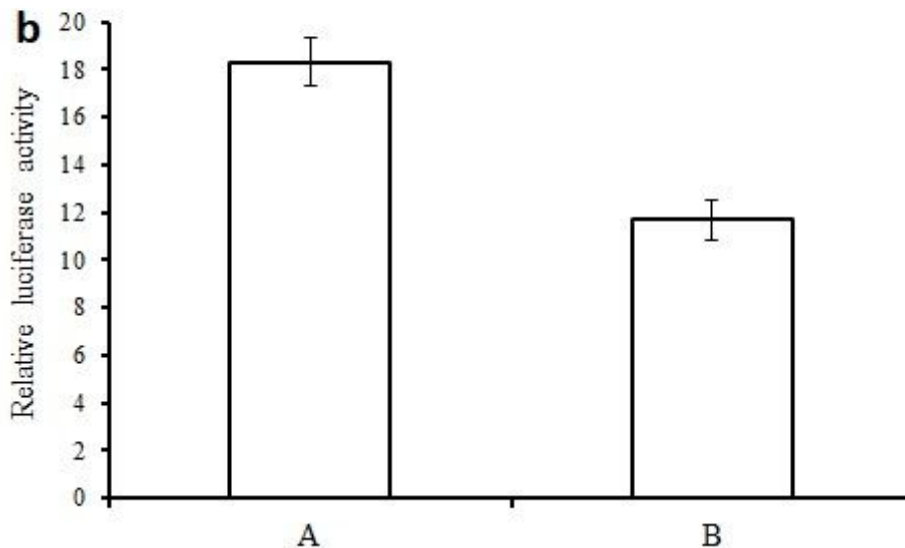


Figure 13

The validation of cpa-miR393_R-1 and target AFB2. (a) Prediction of the binding sites of cpa-miR393_R-1 in AFB2 mRNA using psRNATarget. (b) Effect of cpa-miR393_R-1 expression on luciferase activity in transfected cells: (A) pCDNA3.1 + pmirGLO-AFB2 and (B) pCDNA3.1- cpa-miR393_R-1 + pmirGLO- AFB2. Data are represented as the mean \pm standard deviation (SD) from three independent experiments.

Supplementary Files

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

- [TableS11.xlsx](#)
- [TableS3.xlsx](#)
- [TableS1.xlsx](#)
- [TableS5.xlsx](#)
- [TableS4.xlsx](#)
- [TableS2.xlsx](#)
- [TableS8.xlsx](#)
- [TableS7.xlsx](#)
- [TableS6.xlsx](#)
- [TableS14.xlsx](#)
- [TableS13.xlsx](#)
- [TableS10.xlsx](#)
- [TableS9.xlsx](#)
- [TableS12.xlsx](#)