

RESEARCH

Open Access



Identification of long non-coding RNAs and microRNAs involved in anther development in the tropical *Camellia oleifera*

Lingshan Kong^{1,2,3†}, Yanjing Zhuo^{4†}, Jieru Xu^{1,2}, Xiangxu Meng^{1,2}, Yue Wang^{1,2}, Wenxiu Zhao^{1,2}, Hanggui Lai⁵, Jinhui Chen^{1,2*} and Jian Wang^{1,3*}

Abstract

Background: Explored the molecular science of anther development is important for improving productivity and overall yield of crops. Although the role of regulatory RNAs, including long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), in regulating anther development has been established, their identities and functions in *Camellia oleifera*, an important industrial crop, have yet not been clearly explored. Here, we report the identification and characterization of genes, lncRNAs and miRNAs during three stages of the tropical *C. oleifera* anther development by single-molecule real-time sequencing, RNA sequencing and small RNA sequencing, respectively.

Results: These stages, viz. the pollen mother cells stage, tetrad stage and uninucleate pollen stage, were identified by analyzing paraffin sections of floral buds during rapid expansion periods. A total of 18,393 transcripts, 414 putative lncRNAs and 372 miRNAs were identified, of which 5,324 genes, 115 lncRNAs, and 44 miRNAs were differentially accumulated across three developmental stages. Of these, 44 and 92 genes were predicted to be regulated by 37 and 30 differentially accumulated lncRNAs and miRNAs, respectively. Additionally, 42 differentially accumulated lncRNAs were predicted as targets of 27 miRNAs. Gene ontology enrichment indicated that potential target genes of lncRNAs were enriched in photosystem II, regulation of autophagy and carbohydrate phosphatase activity, which are essential for anther development. Functional annotation of genes targeted by miRNAs indicated that they are relevant to transcription and metabolic processes that play important roles in microspore development. An interaction network was built with 2 lncRNAs, 6 miRNAs and 10 mRNAs. Among these, miR396 and miR156 family were up-regulated, while their targets, genes (GROWTH REGULATING FACTORS and SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes) and lncRNAs, were down-regulated. Further, the trans-regulated targets of these lncRNAs, like wall-associated kinase2 and phosphomannose isomerase1, are involved in pollen wall formation during anther development.

Conclusions: This study unravels lncRNAs, miRNAs and miRNA-lncRNA-mRNA networks involved in development of anthers of the tropical *C. oleifera* lays a theoretical foundation for further elucidation of regulatory roles of lncRNAs and miRNAs in anther development.

Keywords: Anther development, MicroRNAs, Long non-coding RNAs, Network, *Camellia oleifera*

[†]Lingshan Kong and Yanjing Zhuo contributed equally to this work.

*Correspondence: jinhui.chen@hainanu.edu.cn; 285464212@qq.com

¹ Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, 572025 Sanya, P. R. China

Full list of author information is available at the end of the article

Background

Anther development is crucial for male fertility and sexual reproduction in plants [1]. Cells in an anther undergo meiosis to produce microspores, which further develop into mature pollen grains [2]. According to morphological features, anther development can be divided into 14



stages: during stages 1 to 4, cell division events occur within the developing anther primordia to establish characteristic structure of an anther. As pollen mother cells undergo meiosis at stages 5 and 6, they produce tetrads (stage 7); dissolution of the tetrad callose wall and release of microspores mark stage 8. These microspores undergo mitosis and develop into mature pollen grains during stages 9 to stage 12; this follows anther dehiscence, pollen release and stamen senescence at stages 13 to 14 [3]. Flowering is a complex reproductive process, which requires tight control of gene expression to form a regulatory network [4]. With the development of genomics and transcriptomics tools, genes which were involved in anther development have been widely studied. For instance, the genes involved in sporopollenin synthesis, such as TETRAKETIDE α -PYRONE REDUCTASE1 (*TKPR1*), ATP-Binding Cassette Transporter G26 (*ABCG26*) and POLYKETIDE SYNTHASE A (*PKSA*), play a significant role in pollen wall formation [5–7].

Following the advent of the high-throughput sequencing, a large number of non-coding RNAs (ncRNAs) were identified [8, 9]. Based on the product length, ncRNAs could be subdivided into two groups: long ncRNAs (lncRNAs, length > 200 nt) and microRNAs (miRNAs) that are about 21 nt [10]. Recent studies have revealed that plant lncRNAs may regulate gene expression in floral transition [11], flower development [12], flowering time regulation [13] and male sterility [14]. For example, Huang et al. [15] found 14 lncRNAs that were highly co-expressed with 10 pollen-associated target genes and might participate in pollen development in *Brassica rapa*. Further, many miRNA families are involved in flowering-related processes, such as the induction of floral competence, floral patterning, and the development of floral organs [4]. For example, miR156, miR156a and miR156b regulate SQUAMOSA PROMOTER BINDING PROTEIN-LIKE6 (*SPL6*), *SPL8*, *SPL12* and *SPL3* that are involved in cell proliferation in early anther development and regulation of flowering time [16–18]. In addition, miR396, miR396a and miR396b regulate GROWTH REGULATING FACTORS (*GRFs*) that are involved in development of the pistil and determination of floral organ specification [19–21].

Camellia oleifera is an evergreen shrub or a small tree, belonging to the family Theaceae, which is an important source of oil derived from woody plants. *C. oleifera* oil can be used for making a wide range of products, such as lubricants, detergent, rustproof oil, and biopesticides, medicines, cosmetics and functional foods that have high economic value [22]. Although, *C. oleifera* produces a large number of flowers, most of them are do not produce seeds in cultivation. Currently, most of the omics research is mainly focuses on the aspect of fruit

development in *C. oleifera*, while the genes involved in flower bud development and their regulatory mechanisms remain to be elucidated.

Here, we have investigated the role of lncRNA and miRNA in the tropical *C. oleifera* anther development. We used PacBio single-molecule real-time (SMRT) sequencing, RNA sequencing (RNA-seq) and small RNA sequencing to identify and characterize mRNAs, lncRNAs and miRNAs differentially expressed in different developmental stages of anthers. Gene ontology (GO) analysis indicated that genes targeted by lncRNAs and miRNAs were involved in meiosis, pollen wall formation and microspore development. Additionally, we identified a network of interactions between ncRNAs and mRNAs that might participate in anther development. Overall, this study is helpful in not only understanding of roles of miRNAs and lncRNAs in the tropical *C. oleifera* anther development, but also provides a theoretical basis for further research on the anther development in *C. oleifera* and other industrial crops.

Results

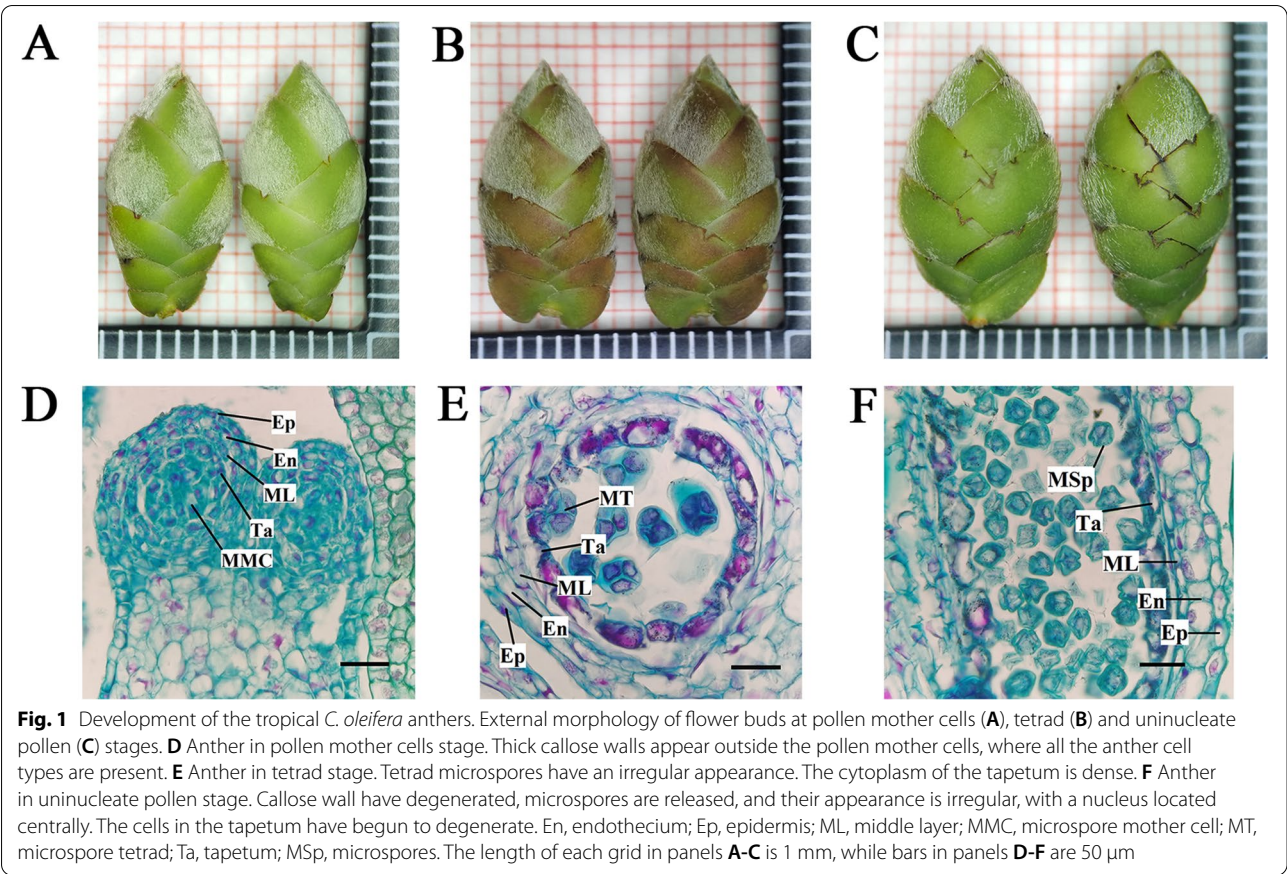
Identification of key anther development stage in the tropical *C. oleifera*

Phenotypic data was collected over 8 times of measurements. Striking contrast in floral bud appearance (length) was revealed in the 3rd round of measurement (Table 1; Fig. 1A–C). The analysis of paraffin section showed significant changes in the anthers during these measurements (Fig. 1). Specifically, microspore mother cells appeared at the 2nd measurement (Fig. 1D); at 3rd measurement, we found that meiosis had completed, tetrads were formed and pollen wall formation had begun (Fig. 1E). Subsequently, callose wall, surrounding the tetrads, had degenerated and individual microspores were released at the 4th measurement (Fig. 1F). These three

Table 1 Growth measurements of floral buds of the tropical *C. oleifera*

Number of measurements	Length (mm)	Breadth (mm)	Thickness (mm)
1st	12.29 ^{dx}	5.96 ^f	5.05 ^e
2nd	13.07 ^d	6.24 ^{ef}	5.24 ^{de}
3rd	14.33 ^c	6.87 ^{de}	5.75 ^{de}
4th	14.65 ^c	7.23 ^d	6.06 ^{cd}
5th	15.15 ^{bc}	7.61 ^{cd}	6.65 ^{bc}
6th	15.87 ^b	8.18 ^{bc}	7.23 ^b
7th	16.99 ^a	8.70 ^{ab}	8.05 ^a
8th	17.70 ^a	9.09 ^a	8.61 ^a

^x Means with the same letter within the same column are not statistically significant difference ($p < 0.05$)



stages corresponded to pollen mother cells stage (CoA1), tetrad stage (CoA2), and uninucleate pollen stage (CoA3) [3]. The floral buds from the 2nd measurement to the 4th measurement were chosen as samples for the follow up experiments.

Analysis of transcriptome data

To unravel the changes in genes expression during different stages of the tropical *C. oleifera* anther development, RNAs from three stages of anther, each with three biological replicates, were sequenced and analyzed by PacBio Sequel platform. A total of 391,875 polymerase reads (average read length of 72,247 bp) and 9,545,656 subreads (average read length of 2,895 bp) were produced with SMRT (Table 2). To derive a more accurate sequence information, a CCS was generated from reads that fully-passed at least twice through the insert, for a total of 346,466 CCSs (average read length of 3,141 bp) (Table 2). SMRTlink identified 289,298 full-length reads and 287,941 full-length non-chimeric (FLNC) reads (average read length of 2,980 bp; Table 2), respectively. The FLNC reads of the same transcript were clustered, and redundant reads were removed to obtain consensus reads using the ICE algorithm. Non-full-length non-chimeric reads

Table 2 Summary of reads from SMRT sequencing

Item	Number	Mean length
Polymerase reads	391,875	72,247
Subreads	9,545,656	2,895
Circular consensus sequences (CCS)	346,466	3,141
Full-length non-chimeric reads (FLNC)	287,941	2,980
Polished consensus reads	30,275	2,861
Corrected polished consensus reads	30,275	2,861
Full-length transcripts (genes)	18,393	2,812

were used to correct the consensus reads using arrow software, and 30,275 polished consensus sequences were ultimately obtained, with a mean length of 2,861 bp (Table 2). RNA-seq generated 63,625,261 (CoA1), 66,872,477 (CoA2), 67,554,470 (CoA3) raw reads for three stages, respectively. After trimming, 62,424,908 (CoA1), 65,373,048 (CoA2), 65,984,584 (CoA3) clean reads, respectively, were retained (Table 3). The raw reads of Illumina sequencing were then used to correct the SMRT data. After removing redundant sequences, a total of 18,393 transcripts, with a mean length of 2,812 bp, were retained.

The sequencing of small RNAs unveiled the dynamic regulation of miRNAs on gene expression during anther development. After sequencing of nine small RNA libraries, a total of 117,553,152 high quality reads were obtained from the CoA1 (38,868,735), CoA2 (39,821,973) and CoA3 (38,862,444), respectively (Table 3). After removing contaminant and low quality reads, 38,084,176 (CoA1), 39,286,803 (CoA2) and 38,320,966 (CoA3) clean reads, respectively, were obtained in the size range of 18–30 nt.

Identification and characterization of lncRNA and miRNA

During the tropical *C. oleifera* anther development, a total of 414 transcripts were annotated as lncRNAs (Supplementary Fig. S1 and Supplementary Table S1). These 414 putative lncRNA sequences ranged from 438 bp to 5,637 bp, with an average length of 2,367 bp (Fig. 2A and Supplementary Table S1). About half of the lncRNAs (53%) had lengths ranging from 1,500 bp to 2,500 bp, while only 10% were shorter than 1,500 bp (Fig. 2A). By combining the results of Minimap2 and FEELnc classifications, 97 lncRNAs were classified into four groups: 29 sense lncRNAs (29.90%), 61 lincRNAs (62.89%), 3 anti-sense lncRNAs (3.09%) and 4 sense intronic lncRNAs (4.12%) (Supplementary Table 1). Most of the lncRNAs were expressed at relatively low levels, with FPKM values (\log_2) ranging from -1 to 1 in all samples, as compared to expression of mRNAs (Fig. 2B). The number of lncRNAs slightly increased over time during development, a trend also found in the mRNA component of transcriptome (Fig. 2C).

Clean sRNAs reads ranged from 15 to 30 nucleotides (Fig. 2D). The majority of clean sRNA reads were 21–25 nt, and the most abundant sRNAs were 24 nt long, representing more than 60% of the total clean reads. The 21-nt sRNAs, which were mainly miRNAs, were the third most abundant size class. The abundance of 21- and 24-nt sRNAs were higher in the CoA2 library as compared with

CoA1 and CoA3 libraries. A total of 335 known and 37 novel miRNAs were identified, which accounted for 224 (CoA1), 270 (CoA2) and 247 (CoA3) known miRNAs, and 33 (CoA1), 36 (CoA2) and 34 (CoA3) novel miRNAs, respectively (Supplementary Table S2 and Fig. 2E). The first base of known miRNAs was dominantly a 'U', whereas 'A' base ratio was largest in novel miRNAs (Supplementary Fig. S2 and S3). The expression dynamics of miRNAs in three stages of anthers were analyzed using TPM values (Supplementary Table S2). About 40% miRNAs exhibited high-levels of expression (≥ 50 TPM) [23]: about 9% of these showed TPM values between 50 and 100, about 13% miRNAs had TPM values between 101 and 500, and about 17% had TPM values > 501 (Fig. 2F). Top 30 miRNAs (based on TPM) belonged to ten families of miR166, miR159, miR160, miR396, miR168, miR403, miR408, miR167, miR319, and miR395.

Analysis of differentially expressed mRNAs, lncRNAs, and miRNAs in anther development

To investigate the gene expression patterns at three stages of the tropical *C. oleifera* anther development, the FPKM values were used to normalize the reads from RNA-seq. A total of 5,324 mRNAs were differentially expressed (Fig. 3A and Supplementary Table S3) between three developmental stages. The largest number of differentially expressed mRNAs (3,586) were found by comparing CoA2 vs. CoA1, with the development of floral bud, the number of differentially expressed mRNAs decreased gradually (Fig. 3A). Of 5,324 mRNAs, 1,593, 893 and 354 mRNAs were uniquely differentially expressed in CoA2 vs. CoA1, CoA3 vs. CoA2 and CoA3 vs. CoA1, respectively, and 53 mRNAs were differentially expressed in all control groups (Fig. 3B). GO enrichment analysis of significantly differentially expressed mRNAs revealed seven significant terms of biological processes among up-regulated mRNAs in CoA2 vs. CoA1 (Supplementary Table

Table 3 Summary of Illumina sequencing

Library	Groups	Raw reads		Clean reads		Clean bases(G)	
		Total	Average	Total	Average	Total	Average
RNA-seq	CoA1	63,625,261	21,208,420.33	62,424,908	20,808,302.67	18.73	6.24
	CoA2	66,872,477	22,290,825.67	65,373,048	21,791,016	19.62	6.54
	CoA3	67,554,470	22,518,156.67	65,984,584	21,994,861.33	19.79	6.60
	Total	198,052,208	66,017,402.67	193,782,540	64,594,180	58.14	19.38
Small RNA-seq	CoA1	38,868,735	12,956,245	38,084,176	12,694,725.33	1.90	0.63
	CoA2	39,821,973	13,273,991	39,286,803	13,095,601	1.96	0.65
	CoA3	38,862,444	12,954,148	38,320,966	12,773,655.33	1.92	0.64
	Total	117,553,152	39,184,384	115,691,945	38,563,981.67	5.79	1.93

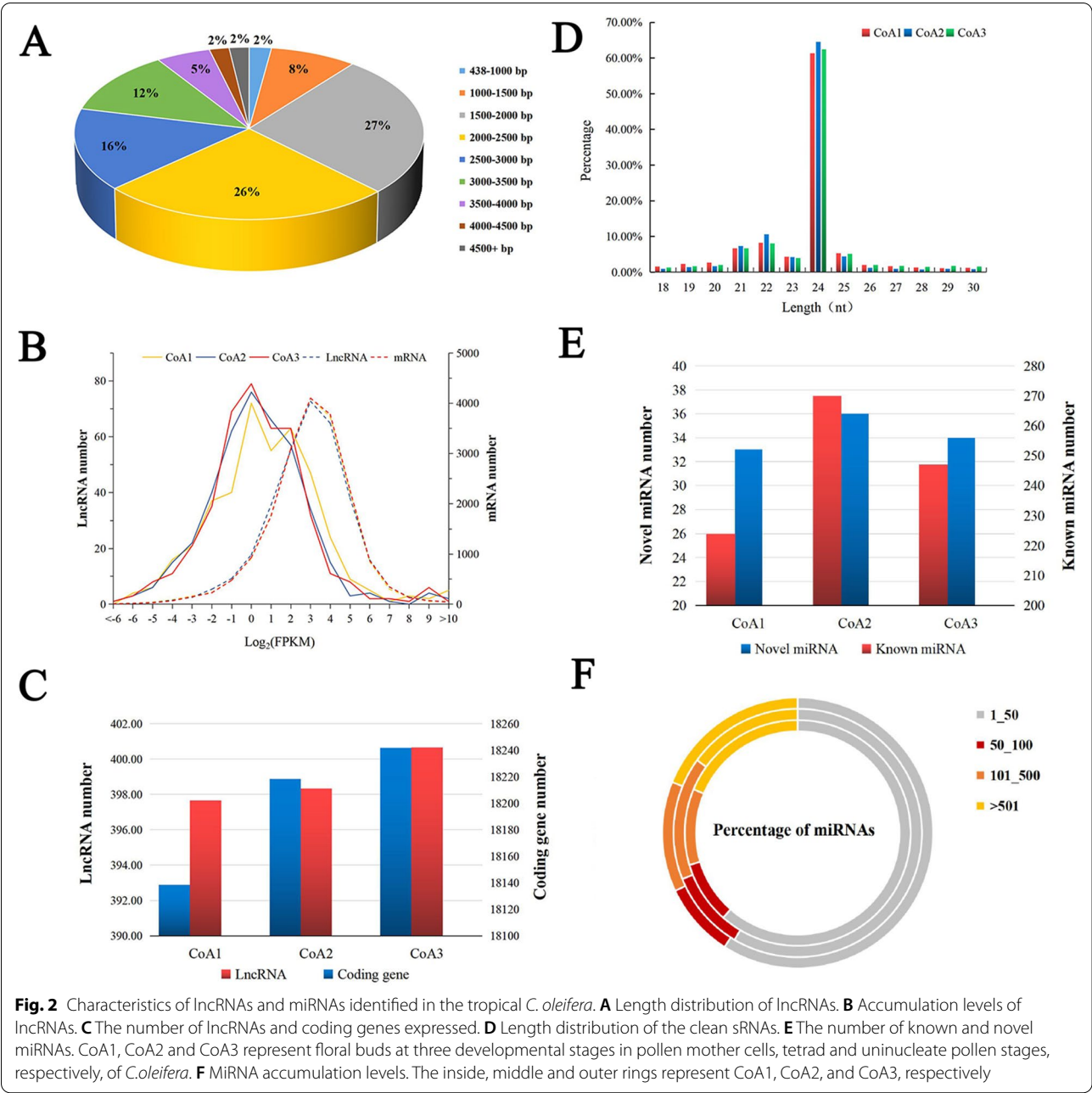
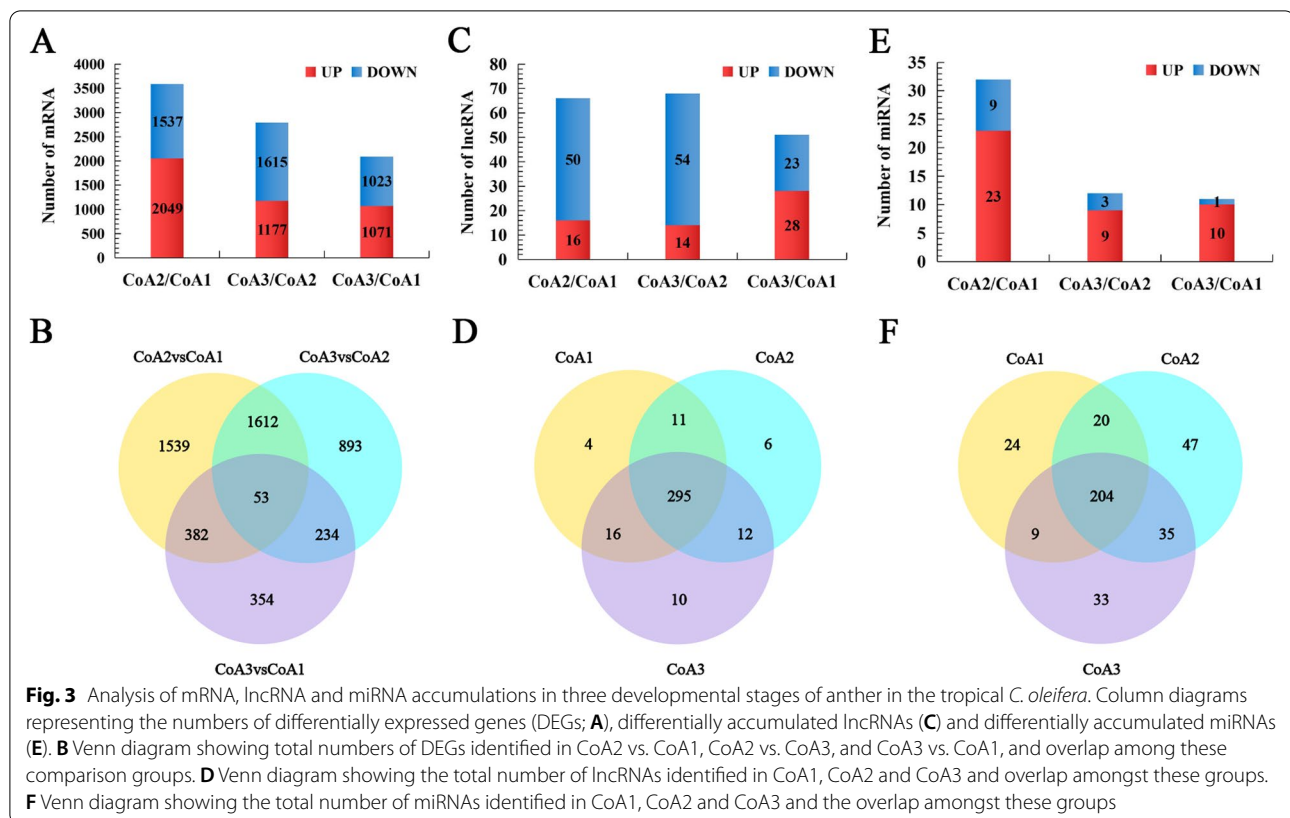


Fig. 2 Characteristics of lncRNAs and miRNAs identified in the tropical *C. oleifera*. **A** Length distribution of lncRNAs. **B** Accumulation levels of lncRNAs. **C** The number of lncRNAs and coding genes expressed. **D** Length distribution of the clean sRNAs. **E** The number of known and novel miRNAs. CoA1, CoA2 and CoA3 represent floral buds at three developmental stages in pollen mother cells, tetrad and uninucleate pollen stages, respectively, of *C.oleifera*. **F** MiRNA accumulation levels. The inside, middle and outer rings represent CoA1, CoA2, and CoA3, respectively

S4). The most significantly enriched processes were ‘microtubule-based movement’ and ‘carbohydrate metabolic process’. Seventeen terms were significant among down-regulated mRNAs in CoA2 vs. CoA1, where most enriched terms were ‘regulation of transcription, DNA-templated’ and ‘regulation of cellular metabolic process’ (Supplementary Table S4). After the exploration of functional categories of differentially expressed mRNA, we found 14 genes related to pollen development, pollen wall formation, flowering time control, which may

play important roles in floral bud development of the tropical *C. oleifera* (Supplementary Table S5). A total of 115 lncRNAs were differentially accumulated in three developmental stages. With the development of anther, the number of differentially expressed lncRNAs gradually increased (Fig. 3C). The majority of lncRNAs (295) were detected in all the three developmental stages (Fig. 3D). A total of 44 miRNAs were differentially accumulated amongst the three stages at a threshold of $|\log_2FC| \geq 1$



and $q\text{-value} \leq 0.05$ (Fig. 3E), the largest number of differentially accumulated miRNAs (32) were found by comparing CoA2 vs. CoA1. However, we did not detect significant differences in accumulations of novel miRNAs across three developmental stages. Comparative analysis of 372 miRNAs (335 known and 37 novel miRNAs) showed that 204 miRNAs were common in three stages, while 24, 47 and 33 miRNAs were specific to CoA1, CoA2, and CoA3, respectively (Fig. 3F). These known miRNAs were assigned to 74 known families, of which 36.49% of the families contained a single member, 14.86% contained more than 20 members, and miR166 and miR159 families have 49 members (Fig. 4).

Target genes identification and analysis of differentially accumulated lncRNA and miRNA in anther development

A total 65 genes were predicted as potential trans-regulated targets of 48 differentially accumulated lncRNAs (Supplementary Table S6). The correlation analysis between lncRNAs and their potential target genes showed that 88.7% had a positive, while 11.3% had a negative correlation. The GO enrichment analysis of the target genes of differentially expressed lncRNAs in CoA2 vs. CoA1 and CoA3 vs. CoA2 comparisons were enrichment in the same 26 GO terms, such as 'alditol metabolic

process', 'regulation of autophagy' and 'carbohydrate phosphatase activity' (Supplementary Table S7). Additionally, the target genes of up-regulated lncRNAs were significant enriched in photosystem-related GO term such as 'photosystem II reaction center' and 'photosynthetic membrane' (Supplementary Table S7). The KEGG enrichment indicated that the target genes of up-regulated lncRNAs were significant enriched in 3 pathways, 'ascorbate metabolism', 'aldarate metabolism' and 'ribosome'; whereas target genes of down-regulated lncRNAs were significant enriched in 9 pathways that included 'RNA polymerase', 'phagosome', 'starch and sucrose metabolism' (Supplementary Table S8). Additionally, 44 target genes of 37 differentially accumulated lncRNAs were differentially expressed in the three developmental stages. Further, gene annotation indicated that 10 differentially accumulated lncRNAs that were highly co-expressed with 7 differentially accumulated target genes, such as AUXIN RESISTANT1 (*AXR1*), Phosphomannose isomerase1 (*PMI1*), wall associated kinase2 (*WAK2*) and Dicer-Like-3b (*DCL3b*). These were related to transcription, cell wall formation, and flowering time control, which indicated these lncRNAs may play important role in floral bud development in the tropical *C. oleifera* (Supplementary Table S9).

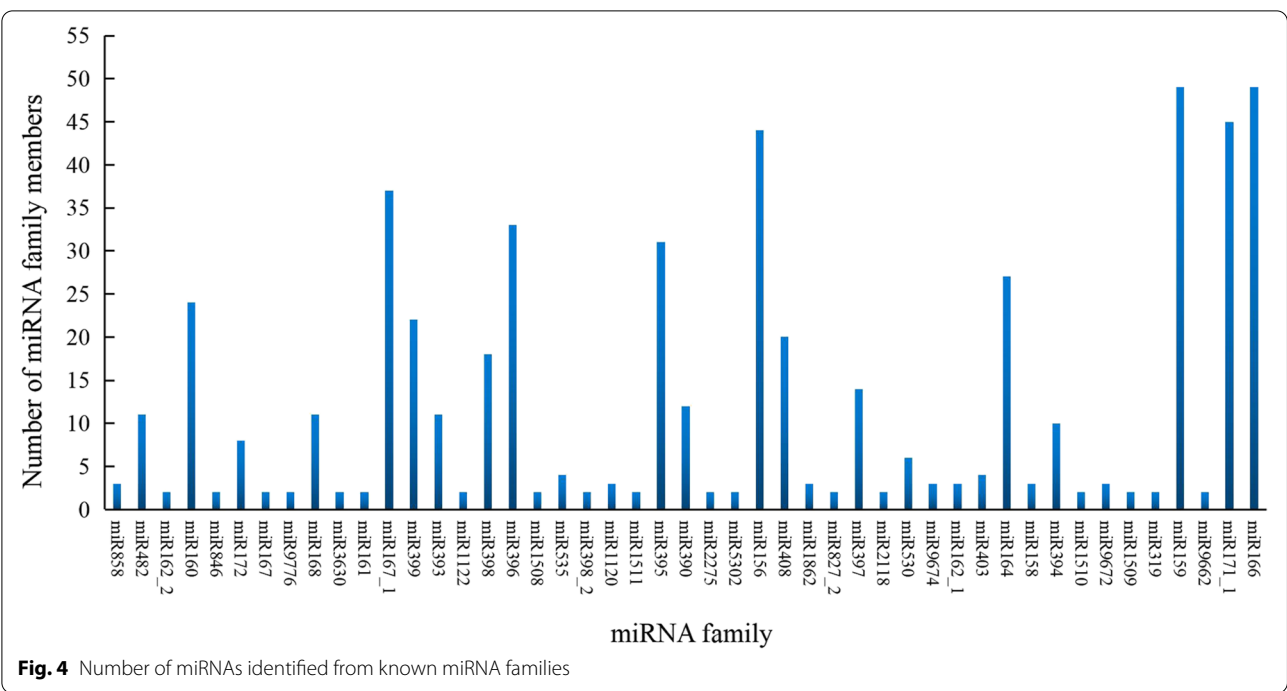


Fig. 4 Number of miRNAs identified from known miRNA families

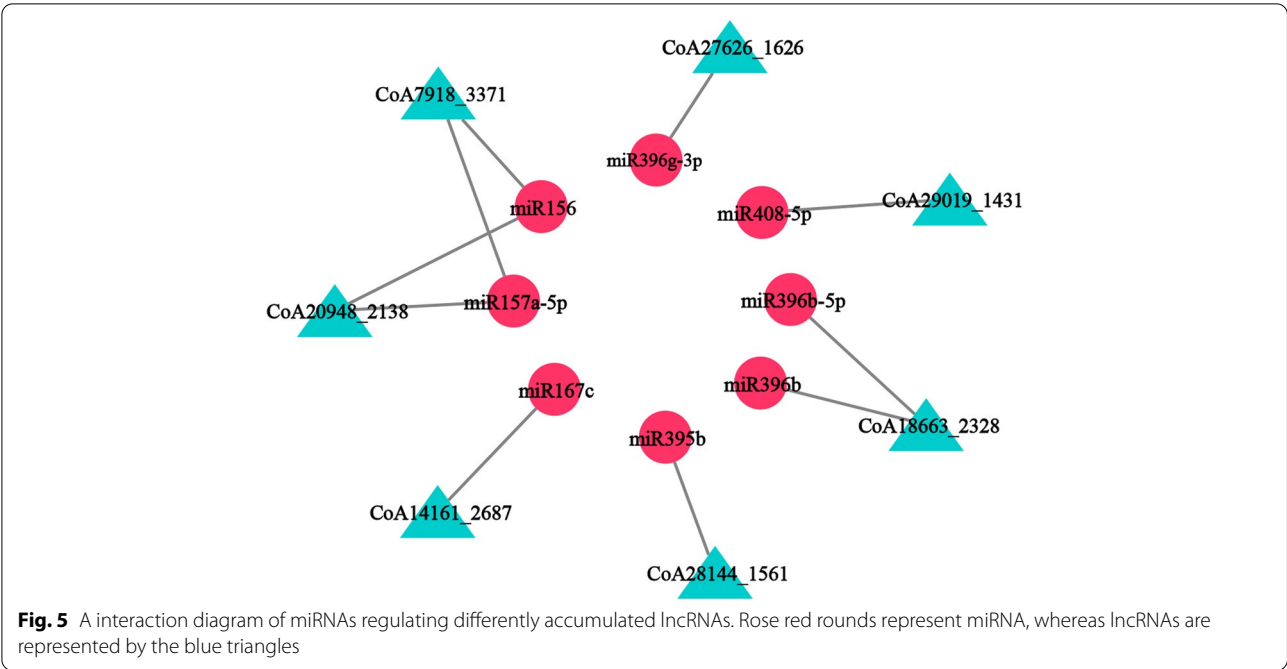
A total of 286 genes were predicted as putative targets of 44 differentially accumulated miRNAs, resulting in a total of 570 putative miRNA-target modules, as several miRNAs and targets displayed one-to-many relationships (Supplementary Table S10). Among these, 527 modules were predicted to function through miRNA-directed cleavage, while 43 modules might function through translational inhibition (Supplementary Table S10). Further, 54 (18.9%) target genes were mapped to transcription factors belonging to 15 families, amongst which, GRF, HD-ZIP, GRAS, ARF, SBP and TCP were significantly enriched. Additionally, 92 differentially accumulated genes were predicted as targets of 30 differentially accumulated miRNAs. Further, Pearson's correlation analysis indicated that 14 differently accumulated miRNAs were negatively correlated with 25 differentially accumulated target genes, and GO enrichment analysis indicated that these genes were enriched in 75 terms, such as 'regulation of transcription, DNA-templated', 'intracellular organelle' and 'organic cyclic compound binding' (Supplementary Fig. S4A). KEGG enrichment indicated that 14 target genes were enriched in 4 pathways, which included 'oxidative phosphorylation' and 'ubiquinone and other terpenoid-quinone biosynthesis' (Supplementary Fig. S5A).

Next, we examined whether the lncRNAs could be targets of miRNAs. Indeed, psRNATarget program predicted 42 differentially accumulated lncRNAs as

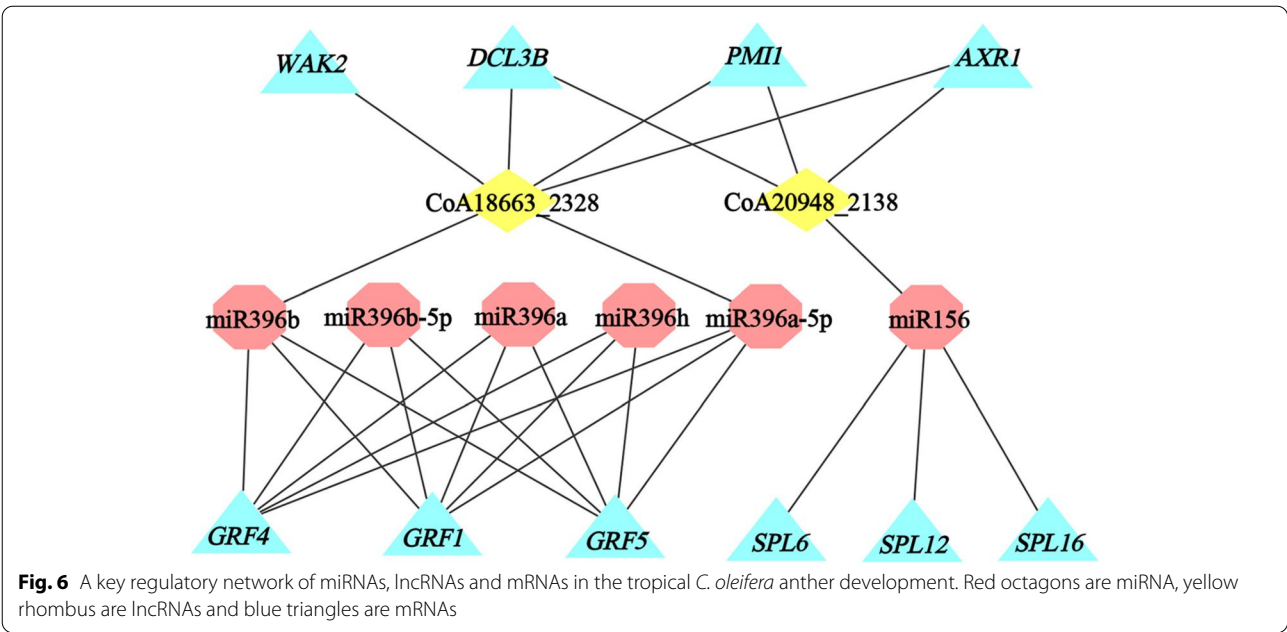
targets of 27 differentially accumulated miRNAs (Supplementary Table S11). Moreover, 27 differentially accumulated genes were targeted by these 42 differentially accumulated lncRNAs. Further, Pearson's correlation analysis indicated that 8 differently accumulated miRNAs were negatively correlated with 7 differentially accumulated lncRNAs, which target 11 differentially accumulated gene. These genes were enriched in 14 GO terms, such as 'glutathione catabolic process' and 'omega peptidase activity' (Supplementary Fig. S4B). KEGG enrichment indicated that the target genes were enriched in 2 pathways, 'RNA polymerase' and 'Fructose and mannose metabolism' (Supplementary Fig. S5B). Based on these results, we investigated miRNA-lncRNA relationships and modeled a interaction diagram of 7 miRNAs and 8 lncRNAs. It was evident that one lncRNA could potentially be targeted by 1–2 miRNAs and one miRNA could target 1–2 lncRNAs (Fig. 5).

Constructing miRNA-lncRNA-mRNA gene regulatory networks in the developing anthers

We further explored the regulatory mechanisms conferred by miRNAs during anther development. Accordingly, we selected differently accumulated genes that were targeted by lncRNAs (CoA18663_2328 and CoA20948_2138) as well as were associated with pollen wall formation (*WAK2* and *PMII*), meiosis and transcription (*AXR1* and *DCL3b*) to construct



an interaction network with miRNAs and their target genes (Fig. 6). This network defined connections between 6 miRNAs and 6 genes involved in anther development. Among them, the member of miR396 family (miR396a, miR396b, miR396a-5p, miR396b-5p, miR396h) modulated the transcript levels of genes related to formation of pollen mother cells and meiosis through their targeted regulation of *GRF* family (*GRF1*, *GRF4* and *GRF5*; Fig. 6). Additionally, miR156 targeted *SPL* family, such as *SPL6*, *SPL12* and *SPL16*, and were involved in early anther development (Fig. 6). In this network, all miRNAs were up-regulated, whereas 2 lncRNAs and 10 mRNAs were down-regulated (Supplementary Table S12).



Validation of gene expression by quantitative reverse transcription PCR (qRT-PCR)

To verify the accuracy of the RNA-Seq and small RNA-seq results, the relative expression levels analysis of 4 lncRNAs, 7 genes and 8 miRNAs were investigated with qRT-PCR. According to the results obtained from qRT-PCR, the expression pattern of most selected miRNAs, lncRNAs and genes were largely similar with the expression levels calculated using our sequencing data (Fig. 7). Although the fold-change (FC) values detected by qRT-PCR did not exactly match the FC values which calculated by sequencing, but the expression profiles were basically consistent for all selected miRNAs, lncRNAs and genes tested, which indicated that the sequencing data were reliable.

Discussion

Transcriptome mRNA in the anther development of the tropical *C. oleifera*

Anther development is a multistage and complex process that determines yield and quality of most of the industrial crops. This process is controlled by a complex network of numerous genes [24, 25]. With a plethora of plant genomes now sequenced, -omics driven technologies are playing a significant role in the study of the regulatory mechanism of anther development [25, 26]. Still, the identification and functional exploration of *C. oleifera* molecular regulatory network during anther development had been very limited, as previous investigations were mainly focused on cytological studies and single regulatory genes [27–29]. In this study, the anther development in the tropical *C. oleifera* was morphologically and cytologically monitored on temporal scales, and important stages were selected for sequencing, to systematically identify the mRNAs, lncRNAs and miRNAs, and their regulatory network.

Several studies have reported direct correlation of floral bud size and other inflorescence characteristics to anther and microspore development, as it is shown that the development of anther and microspore advances as the size of floral buds increase [30–32]. For instance, the meiotic phase of pollen mother cell can be predicted accurately by using the length-breadth ratio of floral bud in the tropical *C. oleifera* [29]. Our morphological observations and analysis of paraffin sections, combined with previous studies [27, 28], also show that significant changes in floral bud lengths reflected the changes in anthers. Therefore, it suggests that the morphological characteristics of floral bud could effectively judge the developmental stages of the tropical *C. oleifera* anthers.

In order to explore the mechanism underlying anther development, data on transcriptome of the three development stages of anther were obtained by using SMRT

sequencing and RNA-seq. By combining SMRT with Illumina sequencing, 30,275 complete transcripts were obtained, which demonstrated a more complete assembly of transcriptome than by Huddleston et al. [33]. Differences in expression of regulatory genes at various developmental stages of anthers was evident, where the largest number of DEGs (3,586) were found by comparing CoA2 vs. CoA1. These observations are supported by previous studies in *Arabidopsis* [34] and *Hordeum vulgare* [35] regarding large-scale transcriptional re-organization in pollen mother cells when they enter meiosis.

The pollen walls maintain pollen structure, provide protection from various environmental stresses and preserve pollen germination [36]. Previous studies have found that pollen walls were the most intricate cell walls in plants. They are composed of exine and intine layers that are formed in a complex series of biological events involving a large number of genes [37]. Callose synthesis has a vital function in building a properly sculpted exine, whose integrity is essential for pollen viability [38]. Recent research shows that callose synthase5 (*CalS5*) and AUXIN RESPONSE FACTOR17 (*ARF17*) are important to callose wall formation [38, 39]. In our study, *CalS5* and *ARF17* were significantly up-regulated at the tetrad stage, which indicated that they may regulated callose wall formation of the tropical *C. oleifera*. Additionally, we found that *A6*, which can regulate the degradation of callose wall and is conducive to the transformation from tetrad wall to pollen wall [40], was significantly up-regulated at uninucleate pollen stage, and might may play a similar role in the tropical *C. oleifera*. Further, sporopollenin is an important component of pollen walls and it is synthesized by a pathway of approximately eight genes in *Arabidopsis* [41]. Previous studies have shown that ABORTED MICROSPORES (*AMS*) plays a central role in coordinating sporopollen biosynthesis and the secretion of materials for pollen wall patterning [42]. *AMS* can directly regulate the expression of genes [43] involved in processes like fatty acid hydroxylation (*CYP703A2* and *CYP704B1*) [5, 44], phenolic synthesis (*PKSB* and *PKSA*, *TKPR1* and *TKPR2*) [6, 7], and lipid transport (*ABCG26*) [45]. In our study, *AMS*, *CYP703A2*, *CYP704B1*, *PKSB*, *PKSA*, *TKPR1*, *TKPR2*, and *ABCG26* were continuously up-regulated during anther development, indicating that these genes may play important roles in the establishment of the tropical *C. oleifera* pollen wall. The expression of *MYB80* [46] and *RPK2* [47], which regulates programmed cell death (PCD) of tapetum, is highest in the tetrad stage, which may indicate that more cells of tapetum started undergoing PCD in the tetrad stage; this might benefit the development of pollen wall in the tropical *C.*

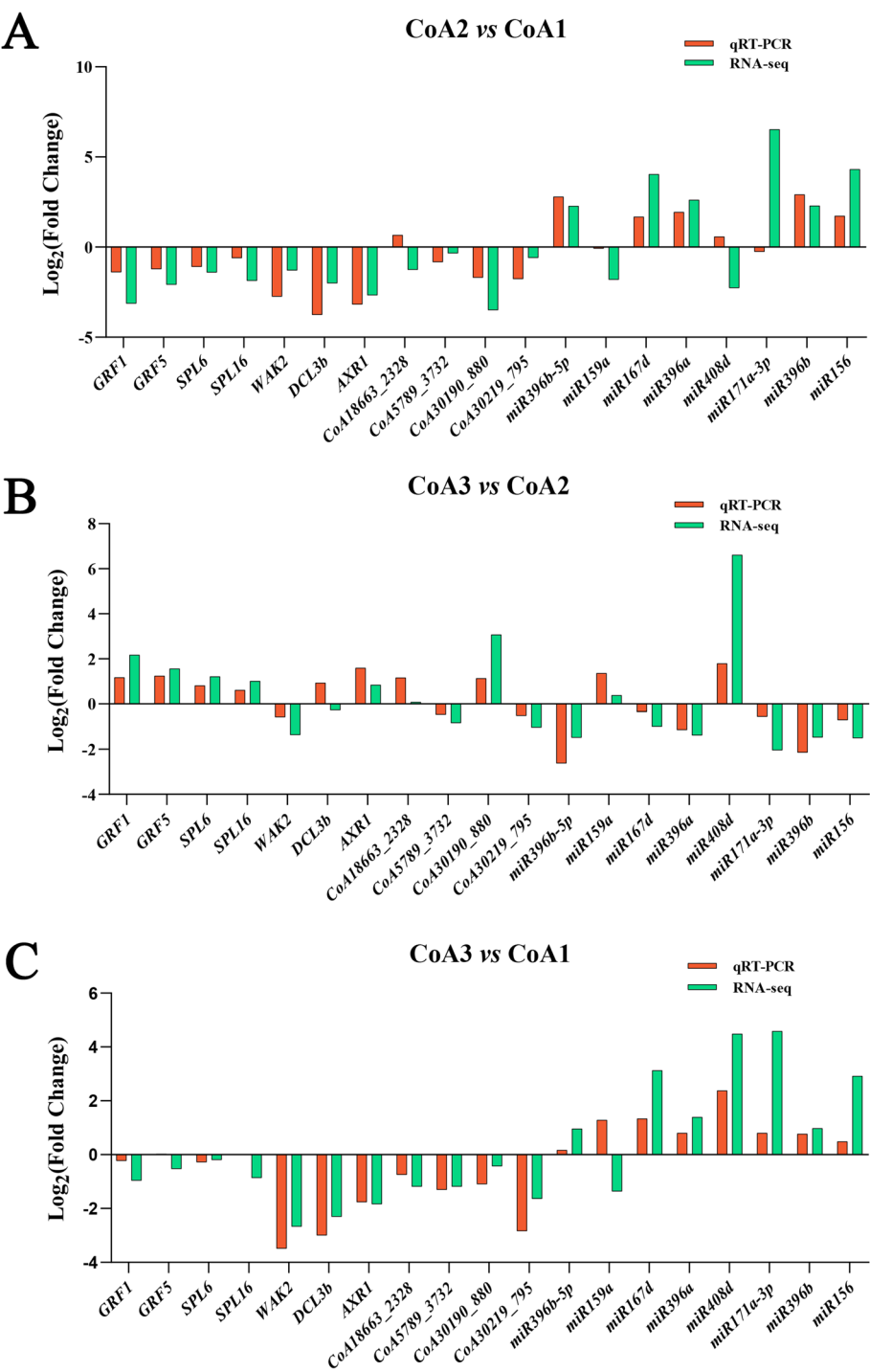


Fig. 7 qRT-PCR of the expression levels of miRNAs, lncRNAs and genes in anther from three developmental stages. The *TubA* was used as internal controls for lncRNAs and mRNAs in A, B and C; the 5S were used as internal controls for miRNAs in A, B and C

oleifera. Additionally, some genes related to the regulation of flowering time, such as *EBS* [48] and *JMJ30* [49], were also differentially expressed in the three stages of anther development, and hints towards their engagement in regulating the process of anther development in the tropical *C. oleifera*. The genes identified here would provide a reference for future studies on anther development in *C. oleifera*.

LncRNAs are involved in the regulation of anther development in the tropical *C. oleifera*

LncRNAs are unravelling as key regulators of plant development [50], and their role in plant sexual reproduction has been confirmed. For example, lncRNAs related to sexual reproduction have been systematically identified in *Gossypium hirsutum* [24], *Hordeum vulgare* [35] and *Brassica rapa* [15], but their role in the development of male buds of *C. oleifera* had still remained unknown. Here, we have identified a total of 414 putative lncRNAs in three different development stages of *C. oleifera* anthers by RNA-seq.

Previous studies in *B. rapa* [15] and *Populus* [51] indicated that lncRNAs are shorter and accumulated at significantly lower levels than protein-coding transcripts. Our study showed that the average length of lncRNAs was 2,367 bp, whereas mRNAs were of an average length of 2,811 bp. On the other hand, the accumulation level of these lncRNAs was lower than those of mRNAs, which was similar to previous studies [15]. LncRNAs could regulate gene expression in trans-acting manner [51]; such as trans-acting lncRNAs have been discovered in large numbers recently [52]. Here, trans-acting functions of lncRNAs were explored with the help of co-expression analyses, where a total of 65 genes were predicted as potential trans-regulated targets of 48 differentially accumulated lncRNAs. Previous studies have shown that autophagy is required for post meiotic anther development, including PCD-mediated degradation of the tapetum in rice [53, 54]. We found that the target genes of lncRNAs were enriched in the GO terms of 'autophagy' and in the KEGG pathway of 'phagosome', which indicated that these lncRNA may regulate the expression of genes related to autophagy and participate in the context of anther development. Additionally, several target genes have been involved in anther development and regulation of flowering time, such as *VTC2*, *PMII*, and *WAK2*. Previous studies have suggested that an increase in ascorbic acid content delays flowering [55], and that *VTC2* is necessary for ascorbate generation [56]. Additionally, *PMII* [57], and *WAK2* are involved in cell wall biogenesis [58], and these genes might also be involved in the formation of pollen walls. This result further indicates that lncRNAs are likely to play a regulatory role in anther development in the tropical *C. oleifera*.

MiRNAs regulated lncRNAs and mRNAs in anther development in the tropical *C. oleifera*

MiRNAs are reported to play essential roles in flowering by directing RNA cleavage or inhibiting translation of the target transcripts [59]. Here, we studied the regulatory roles of miRNAs during anther development in the tropical *C. oleifera* with the help of small RNA-seq analysis.

We found that 73% of differentially expressed miRNAs were concentrated in CoA2 vs. CoA1, indicating a large amount of regulatory changes between pollen mother cells stage and tetrad stage. The GO analysis of target genes of differentially expressed miRNAs showed that 9 most significantly enriched pathways belonged to 'RNA metabolism' category, which is mainly involved in transcriptional regulation. These findings support the previous reports on transcription-related genes as key miRNA targets during anther development [60, 61].

The *SPL* family of transcription factors is functionally diverse, their members in different plants species are involved in many fundamental aspects of plant growth and development, including juvenile-to-adult and floral phase transitions [62], pollen development [17], grain yield [63] and leaf initiation rate [64]. MiR156 is one of the most conserved miRNA families in plants [65]. It regulates plant growth and development, including vegetative to reproductive phase transition and floral organ development by regulating *SPLs* [66, 67]. For example, Xing et al. [17] found that miR156 targeted *SPLs* were absolutely required for male fertility in *Arabidopsis* and they worked in early anther development. Here, we identified a strong negative co-expression relationship between *SPLs* (*SPL6*, *SPL12* and *SPL16*) and miR156 during anther development of the tropical *C. oleifera*. We found that *SPLs* had high transcript levels, while miR156 had a low transcriptional level at pollen mother cell stage. These have significantly opposite expression levels at tetrad stage, which indicates that miR156-*SPL* network may play an important role in cell proliferation and meiosis in early anther development of the tropical *C. oleifera*.

MiR396 is also evolutionarily conserved among plant species [68]. Previously, it was reported that the miR396-*GRF* network is an important module affecting plant growth. For example, Yu et al. [69] found that during the grain filling stage of wheat, miR396 is involved in the development of grains by regulating the expression of *GRFs* (*GRF1*, *GRF6*, and *GRF9*). Baucher et al. [19] and Liang et al. [20] found this network plays an important role in proper development of the pistils and determination of floral organ specification. Additionally, some studies have implicated the *GRF*-*GRF*-*INTERACTING FACTOR* (*GIF*) in the specification and formation of archesporial cells and competence of carpel margin meristems [70–72]. Here, we found that *GRFs* (*GRF1*, *GRF4*, and *GRF5*) were targeted by miR396 with a strong negative correlation. Specifically, miR396s had low levels of accumulation during pollen mother cell stage, which significantly increased in the tetrad stage, while the *GRFs* have opposite expression levels. Additionally, we also found *GIF1* followed the same expression-pattern like the *GRFs*, which indicates that miR396-*GRF*-*GIF1* network

may play an important role in the formation of pollen mother cells and meiosis of the tropical *C. oleifera*.

Next, the co-expression relationship between miRNA and lncRNA were calculated by Pearson correlation coefficient. We found that two lncRNAs are negatively co-expressed with miR396- and miR156-members, respectively. Further, analysis of the target genes of lncRNAs showed that they were related to anther development. For example, *AXR1* is required for normal auxin response and is related to DNA repair [73, 74]. *AXR1* was highly expressed in pollen mother cell stage, indicating that it may participate in meiosis of pollen mother cell. Additionally, some of lncRNA-target genes were involved in cell wall biogenesis (*PMII*, *WAK2*), and in small RNA biogenesis pathway (*DCL3b*) [75]. These results strongly implicate lncRNAs with roles in the tropical *C. oleifera* anther development. We have build a comprehensive network of RNA-mediated interactions, putting together the miRNA-lncRNA, miRNA-mRNA, and lncRNA-mRNA interactions, which consists of 6 miRNAs, 2 lncRNAs, and 10 mRNAs. This network shows that the mutual regulation of miRNA, lncRNA and mRNA may play a crucial role in the tropical *C. oleifera* anther development.

Conclusions

Here, we performed a comprehensive characterization of miRNAs and lncRNA from pollen mother cell, tetrad and uninucleate pollen stages, which were resolved as critical during the tropical *C. oleifera* anther development, and predicted their potential roles. By simultaneously deploying multiple transcriptome-profiling technologies, a total of 5,324 genes, 115 lncRNAs, and 44 miRNAs were identified and their patterns of differential accumulation during three stages were analyzed. Moreover, miRNAs (like miR396a and miR156) and lncRNAs (CoA18663_2328 and CoA20948_2138), along with their target genes (such as *GRF4*, *SPL6*, *WAK2* and *PMII*), which associated with early development of anthers, pollen wall formation and flowering time regulation, were used to build an interaction network to comprehensively understand their role in anther development. Overall, this study identified potential miRNA-lncRNA-mRNA networks involved in the tropical *C. oleifera* anther development and lays down a genomics foundation for further investigations.

Methods

Plant materials

C. oleifera has a long history of cultivation in Hainan [76, 77]. Through natural domestication and breeding, it has adapted to the local climate of Hainan and has excellent economic characteristics. The plant material was identified and collected by Jinhui Chen, Jian Wang and Hanggui Lai (Hainan University) from the mountain in Tunchang

district of Hainan Province, and was planted in the tropical *C. oleifera* germplasm resource bank in Hainan University (Danzhou, Hainan, China; 109°29'25" E, 19°30'40" N) by grafting. The voucher specimens have been deposited in the herbarium of Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Hainan University, China. Floral buds were collected from 4-years old clonal *C. oleifera* trees at the tropical *C. oleifera* germplasm resource bank. The preliminary experiments in this study found that the development time of flower of the tropical *C. oleifera* was long, which takes about 9 months, the morphological differentiation and development of floral organs may take place from May to December. Therefore, measurements and acquisitions were carried out an interval of every ten days, between 24 September to 4 December 2020. The three developmental stages of anthers were selected for this study on the basis of phenotypic and histological observations as described below. These three stages corresponded to pollen mother cells stage (CoA1), tetrad stage (CoA2), and uninucleate pollen stage (CoA3).

Measurement and anatomical observation on floral buds plant materials

Five clonal *C. oleifera* trees with same growth status were selected, and then ten similarly sized floral buds in the same place in every tree were randomly selected; their length, breadth and thickness were measured by an electronic vernier calliper (IP54, Shangshen, Shanghai, China) at an interval of every ten days. Data was analyzed with the help of SPSS version 26 software (SPSS Statistics, IBM, New York, USA). Additionally, fifteen similarly sized floral buds were gently excised, six of these floral buds were used for paraffin section experiment and another nine floral buds were used for RNA isolation. The scales were removed with the help of tweezers, 1/3 of the floral bud tissue was longitudinally cut by a sharp blade until a part of the anther was exposed. Then, the remaining (2/3 of the floral bud) tissues were immersed in FAA stationary solution (70% ethanol:formaldehyde:acetic acid, 9:1:1) immediately, which was vacuum infiltrated (25 Pa) for 30 min to fully eliminate the air from the material. These were fixed overnight at 4°C, after which the samples were dehydrated in a graded ethanol series (70%, 85%, 90%, 95% and 100% twice), made transparent with the help of N-butanol, and embedded in paraffin overnight at 65°C [78]. The embedded sample block was mounted on rotary microtome (KD-2258, Kedee, China) and sliced to a thickness of 10 µm [79]. The sections were double dyed with 1% Safranin and 1% Fast Green, and sealed with neutral resin on slides. Clear and recognizable sections were selected and analyzed using an optical microscope (CX23, Olympus, Tokyo, Japan). Stages

of anther development were classified as described in a previous study [3].

RNA isolation and sequencing

Based on the microscopic observations, samples from three important anther development stages (CoA1: pollen mother cells stage; CoA2: tetrad stage; CoA3: uninucleate pollen stage) were collected. One biological replicate comprised of three anthers from a floral bud. Three such biological replicates were used for each stage, which formed the nine samples (CoA11, CoA12, CoA13; CoA21, CoA22, CoA23; CoA31, CoA32, CoA33) for RNA extraction and sequencing. For each stage, anthers were dissected from floral buds with the help of tweezers and collected in a centrifuge tube, which were frozen in liquid nitrogen and stored at -80°C . Total RNA was extracted from anthers using EASYspin Plus Complex Plant RNA Kit (Aidlab Biotech, Beijing, China) from nine samples. Total RNA concentration was detected by NanoDrop 2000 (Thermo Scientific, USA), and quality was tested on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Only RNA samples with absorption values of 260/280 ratio between 1.8 to 2.1 and a 260/230 ratio ≥ 1.8 , and with RNA integrity number (RIN) ≥ 8.8 were selected for follow-up experiments.

For Illumina sequencing, polyadenylated mRNA was enriched by oligo (dT) magnetic beads, and then fragmentation buffer was added to break the mRNA into short pieces. Random hexamers were used to reverse transcribe mRNA into single-stranded cDNA; dNTPs, DNA polymerase I and buffer were added to synthesize double-stranded cDNA. After using AMPure XP beads to purify double-stranded cDNA, they were subjected to end repair, the addition of the poly-A tail, ligation of the sequencing linker, and fragment size selection. Finally, 9 cDNA libraries were subjected to PCR enrichment and sequenced on an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA).

PacBio SMRT sequencing library preparation

To prepare the SMRTbell library, equal amounts of 9 total RNA samples were combined to generate an RNA pool for SMRT sequencing. Oligo-dTs were used to enrich for mRNAs containing poly-A tails in this pool, and SMARTer PCR cDNA synthesis kit (Clontech, USA) was used to synthesize cDNA with mRNA as template. PCR was used to amplify the cDNAs. cDNA was subjected to injury repair, end-repair, ligated to SMRT dumbbell-type linkers, unligated linker sequences at both ends of cDNA were removed, primers were added, and DNA polymerase was used to form a complete/full-length transcriptome library, which

was sequenced using manufacturer's instructions. The subread sequences were obtained by processing the raw sequence data with the help of SMRTlink v8.0 software. Circular consensus sequences (CCSs) were obtained following the correction between subreads. Full-length sequences containing 5' primers, 3' primers, and poly-A tails were clustered using the Iterative Isomer Clustering (ICE) algorithm. Finally, the consensus sequences were polished to obtain high-quality sequences for subsequent analysis.

Illumina and PacBio data analysis

Illumina data was mapped to the PacBio data by using LoRDEC software [80] to obtain clean reads. Redundant sequences were removed with the help of CD-HIT [81], and full-length transcripts were annotated by BLAST v2.2.26 [82], KOBAS v2.0 [83] and HMMER v3.1 [84] searches against public databases, including Swiss-Prot [85], Pfam [86], GO [87] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [88]. Full-length transcripts from this study were used as reference sequences (ref) for each of the genes, clean reads obtained by Illumina sequencing were aligned to refs, and read counts of all genes were obtained. Further, these read counts were converted into fragments per kilobase of transcript per million mapped reads (FPKM) values. Differentially expressed genes (DEGs) were determined across three developmental stages (CoA2 vs. CoA1, CoA2 vs. CoA3, and CoA3 vs. CoA1) on the basis of the criteria of $|\log_2\text{foldchange}| \geq 1$ ($\log_2\text{FC}$) and $q\text{-value} \leq 0.05$. All DEGs were mapped to individual terms in GO database, and the number of genes per term was calculated. GO enrichment analysis was then performed using Goseq software to identify significantly enriched terms in the DEGs. The enriched GO terms were checked using an $p\text{-value} < 0.05$ as the cut off for significant GO terms. KOBAS software [89] was used to perform KEGG (<http://www.genome.jp/kegg/>) pathway enrichment analysis.

LncRNA identification and classification from SMRT sequences

Five steps were adapted to identify lncRNAs from SMRT sequences of *C. oleifera*: (1) PLEK [90] was used with a default parameter of -minlength 200 to evaluate the coding potential of transcripts that lacked genome sequences and annotations. (2) Transcripts with length < 200 bp were removed. (3) Coding-non-coding index [91] was used at the default parameters to distinguish between coded and non-coded sequences; (4) coding potential calculator 2 [92] was used for the support vector machine classifier to evaluate the coding potential of a transcript

based on the biological sequence characteristics of each ORF in transcripts; the e-value was set to “1e-10”. (5) Finally, transcripts were searched against Pfam database [86] to eliminate transcripts encoding proteins and protein coding domains (cutoff E-value = 0.001). At last, transcripts without coding potentials were selected as candidate lncRNAs.

LncRNA classification was carried out according to their genomic location [93]. First, lncRNAs were classified by aligning the SMRT sequences of *C. oleifera* to the genome of *C. oleifera* [94] with the help of Minimap2 [95]. In the second analysis, FEELnc (isBest = 1) was used for this purpose [96]. Finally, the results of these two classifications were combined and lncRNAs were classified into four categories: sense lncRNAs (generic overlap with known exon), long intergenic non-coding RNAs (lincRNAs) (contained intergenic lncRNAs), antisense lncRNAs (overlapped with a known gene on the opposite strand), and sense intronic lncRNAs (falls entirely within an intron of a known gene) [97–99].

Sequencing and identification of miRNAs

Total RNAs from nine samples were used for the preparation of small RNA libraries with the help of NEBNext multiplex small RNA library prep kit for Illumina (NEB, USA) following the manufacturer's protocol. Library quality was assessed on an Agilent Bioanalyzer 2100 system after using Qubit2.0 for preliminary quantification; the library was diluted to 1 ng/μl. The libraries were sequenced (50-bp single-end reads) on an Illumina SE50 platform. The sequences, with lengths ranging between 18–30 bp, from the clean reads were mapped to the SMRT sequences by Bowtie [100]. The miRBase20.0 database was used as a reference, and srna-tools-cli (<http://srna-workbench.cmp.uea.ac.uk/>) was used to identify potential miRNAs and to draw their secondary structures. miREvo (set the parameter to “-i -r -M -m -k -p10 -g 50000”) [101] and miRDeep2 [102] were used to identify potential Dicer cleavage sites, explore secondary structures and determine minimum free energy of small RNA tags so that novel miRNAs could be predicted in the samples. miRNA candidates were obtained after removing tags originating from protein-coding genes, repeat sequences, rRNAs, tRNAs, snRNAs, and snoRNAs by mapping them against the Rfam database. For all candidate miRNAs, custom scripts were used to calculate miRNA counts and to estimate base bias at the first position of identified miRNAs. The miRNA families were identified by comparing the sequences to miRbase (<http://www.mirbase.org/ftp.shtml>) and Rfam databases (<http://rfam.sanger.ac.uk/search/>). TPM (transcript per million) value was used to estimate the expression level of miRNAs [103]. The DEGseq2 was used for

the differential expression analysis with $|\log_2FC| \geq 1$ and $q\text{-value} \leq 0.05$ as the threshold.

Predicting the potential target genes and functional annotation of lncRNAs and miRNAs

It is known that lncRNAs may act in ‘trans’ [51]. In order to predict the genes trans-regulated by differentially expressed lncRNAs, BLAST was used to search the full-length transcriptome sequences of our libraries at an e-value cutoff of ‘1e-5’ and identity = 80%. Then, Pearson correlation ($|\text{correlation}| \geq 0.8$, $P\text{-value} \leq 0.05$) was used to select potential targets on the basis of the expression correlation coefficients between lncRNAs and mRNAs.

The miRNA targets were predicted with the help of psRNATarget with expectation ≤ 3 (<https://www.zhaolab.org/psRNATarget/>) [104, 105] and the probable target mRNAs were selected on the basis of a negative Pearson's correlation coefficient between miRNA and mRNA expression levels (correlation ≤ -0.8 , $P\text{-value} \leq 0.05$). The miRNA targets in lncRNAs were predicted with the help of psRNATarget with expectation ≤ 5 , and a negative Pearson's correlation coefficient between miRNA and lncRNA expression levels ≤ -0.67 (at $P\text{-value} \leq 0.05$).

All lncRNA- and miRNA- target genes were subjected to GO and KEGG pathway enrichment analyses. Pearson's correlation coefficients were calculated in R (version 4.1.1) to construct the interaction network between miRNAs, ncRNAs and mRNAs, and the network was visualized with the help of Cytoscape (version 3.8.2).

Validation of miRNA, lncRNA and gene expression by qRT-PCR

Total RNA of all nine anther samples were reverse transcribed into cDNA using GoScript™ Reverse Transcription System (Promega, USA). Four and seven differentially expressed lncRNAs and mRNAs, which from three developmental stages of anther, were chosen for qRT-PCR verification. The specific primers for the lncRNAs and mRNAs were designed using Primer Premier v5 (Supplementary Table S13). For lncRNAs and mRNAs, qRT-PCR was performed using TB Green Premix Ex Taq II (Tli RNase H Plus; Takara, Beijing, China) in a final qRT-PCR reaction mixture containing 5.0 μL $2 \times$ TB Green Premix Ex Taq, 0.8 μL primers, 2.0 μL cDNA, and 6.0 μL ddH₂O. The alpha-tubulin (*TubA*) genes served as internal controls for normalization. In addition, eight miRNAs which related to anther development were selected for qRT-PCR verification. The miRNA RT/qPCR Detection kit (Aidlab, Beijing, China) was used to reverse transcribe total RNA into cDNA and qRT-PCR analysis for miRNAs, following the manufacturer's recommendations. The two-tailed qRT-PCR method was used to design primers for miRNAs [106] (Supplementary Table S13). The 5S ribosomal

RNA (5S) served as internal controls for normalization. The expression levels of the lncRNAs, genes and miRNAs were calculated using the $2^{-\Delta\Delta Ct}$ method against the internal controls [107]. For each biological replicate, three technical replicates were performed to ensure reproducibility and reliability.

Abbreviations

lncRNAs: Long non-coding RNAs; miRNAs: MicroRNAs; SMRT: Single-molecule real-time; RNA-seq: RNA sequencing; GO: Gene ontology; RIN: RNA integrity number; CCS: Circular consensus sequence; ICE: Iterative Isomer Clustering; ref: Reference sequences; KEGG: Kyoto Encyclopedia of Genes and Genomes; FPKM: Fragments per kilobase of transcript per million mapped reads; DEGs: Differentially expressed genes; FC: Fold change; lincRNAs: Long intergenic non-coding RNAs; TPM: Transcript per million; FLNC: Full-length nonchimeric; qRT-PCR: Quantitative Reverse-Transcription Polymerase Chain Reaction; vs.: Versus.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08836-7>.

Additional file 1: Supplementary Fig. S1. Predicted total number of lncRNAs. **Supplementary Fig. S2.** First base composition of the known miRNA (18–30 nt length) in nine samples. (A). CoA11. (B). CoA12. (C). CoA13. (D). CoA21. (E). CoA22. (F). CoA23. (G). CoA31. (H). CoA32. (I). CoA33. **Supplementary Fig. S3.** First base composition of novel miRNA in nine samples. (A). CoA11. (B). CoA12. (C). CoA13. (D). CoA21. (E). CoA22. (F). CoA23. (G). CoA31. (H). CoA32. (I). CoA33. **Supplementary Fig. S4.** GO analysis of the biological functions of target genes of lncRNAs and miRNAs. GO terms of 25 target genes of 14 differentially accumulated miRNAs (A) and 11 target genes of 7 differentially accumulated lncRNAs, which were targeted by 8 differentially accumulated miRNAs (B). **Supplementary Fig. S5.** KEGG pathway enrichment analysis of target genes of lncRNAs and miRNAs. KEGG pathway enrichment analysis of 25 target genes of 14 differentially accumulated miRNAs (A) and 11 target genes of 7 differentially accumulated lncRNAs which were targeted by 8 differentially accumulated miRNAs (B).

Additional file 2: Supplementary Table S1. lncRNAs identified in the tropical *C. oleifera*. **Supplementary Table S2.** MiRNAs identified in the tropical *C. oleifera*. **Supplementary Table S3.** The significantly differentially expressed mRNAs. **Supplementary Table S4.** Significantly enriched GO categories of differentially accumulating mRNAs. **Supplementary Table S5.** List of differentially expressed genes crucial for the tropical *C. oleifera* anther development. **Supplementary Table S6.** Potential trans-regulated target genes of differentially accumulated lncRNAs. **Supplementary Table S7.** Significant GO categories of differentially accumulating lncRNA-target genes. **Supplementary Table S8.** Significant KEGG categories of differentially accumulating lncRNA-target genes. **Supplementary Table S9.** Target genes of differentially expressed lncRNAs involved in floral bud development. **Supplementary Table S10.** Analysis of target genes for differentially accumulated miRNAs. **Supplementary Table S11.** The differentially accumulated lncRNAs predicted as targets of differentially accumulated miRNAs. **Supplementary Table S12.** Differentially expressed genes and lncRNAs in miRNA-lncRNA-mRNA network. **Supplementary Table S13.** Oligonucleotide primers used in qRT-PCR assays in this study.

Acknowledgements

Not applicable

Authors' contributions

J. C. and J. W. designed the research; L. K., J. X., X. M., Y. W. and W. Z. performed the research. J. C. and L. K. analyzed and interpreted the data; J. C., L. K. and Y. Z. wrote the paper; H. L., J. C. and J. W. supported the funding. All authors have read and approved the manuscript.

Funding

This research was funded by the Scientific Research Fund Project of Hainan University (KYQD(ZR)1830), Hainan Province Science and Technology Special Fund(ZDYF2021XDNY273) and Demonstration Funds for the Promotion of Forestry Science and Technology from the Central Government ([2020]JTG 03).

Availability of data and materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive [108] in National Genomics Data Center [109], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA006274, CRA006275 and CRA006279) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The datasets analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

With the permission by Hainan University, the materials of *C. oleifera* accessions which used in this study were identified and collected from the mountain in Tunchang district of Hainan Province, China, by Jinhui Chen, Jian Wang and Hangui Lai (Hainan University), and planted in the tropical *C. oleifera* germplasm resource bank of Hainan University, Danzhou, Hainan, China. The plant materials don't include any wild species at risk of extinction. We comply with relevant institutional, national, and international guidelines and legislation for plant study. All experiments and data analyses were conducted in Hainan University.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, 572025 Sanya, P. R. China. ²Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education/Engineering Research Center of Rare and Precious Tree Species in Hainan Province, School of Forestry, Hainan University, 570228 Haikou, P. R. China. ³School of Horticulture, Hainan University, 570228 Haikou, P. R. China. ⁴School of Public Administration, Hainan University, 570228 Haikou, P. R. China. ⁵School of Tropical Crops, Hainan University, 570228 Haikou, P. R. China.

Received: 30 May 2022 Accepted: 29 July 2022

Published online: 16 August 2022

References

- Li Z, An X, Zhu T, Yan T, Wu S, Tian Y, et al. Discovering and Constructing ceRNA-miRNA-Target Gene Regulatory Networks during Anther Development in Maize. *Int J Mol Sci*. 2019;20(14):3480.
- Scott RJ, Spielman M, Dickinson HG. Stamen structure and function. *Plant Cell*. 2004;16(Suppl):S46–60.
- Sanders P, Bul A, Weterings K, McIntire K, Hsu Y, Lee P, et al. Anther developmental defects in Arabidopsis thaliana male-sterile mutants. *Sex Plant Reprod*. 1999;11(6):297–322.
- Waheed S, Zeng L. The Critical Role of miRNAs in Regulation of Flowering Time and Flower Development. *Genes (Basel)*. 2020;11(3):319.
- Dobritsa AA, Shrestha J, Morant M, Pinot F, Matsuno M, Swanson R, et al. CYP704B1 is a long-chain fatty acid omega-hydroxylase essential for sporopollenin synthesis in pollen of Arabidopsis. *Plant Physiol*. 2009;151(2):574–89.
- Grienerberger E, Kim SS, Lallemand B, Geoffroy P, Heintz D, Souza Cde A, et al. Analysis of TETRAKETIDE α -PYRONE REDUCTASE function in Arabidopsis thaliana reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *Plant Cell*. 2010;22(12):4067–83.
- Kim SS, Grienerberger E, Lallemand B, Colpitts CC, Kim SY, Souza Cde A, et al. LAP6/POLYKETIDE SYNTHASE A and LAP5/POLYKETIDE SYNTHASE

- B encode hydroxyalkyl α -pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant Cell*. 2010;22(12):4045–66.
8. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*. 2013;193(3):651–69.
 9. Guo X, Gao L, Wang Y, Chiu DK, Wang T, Deng Y. Advances in long noncoding RNAs: identification, structure prediction and function annotation. *Brief Funct Genomics*. 2016;15(1):38–46.
 10. Bai Y, Dai X, Harrison AP, Chen M. RNA regulatory networks in animals and plants: a long noncoding RNA perspective. *Brief Funct Genomics*. 2015;14(2):91–101.
 11. Wang Z, Zhu T, Ma W, Wang N, Qu G, Zhang S, et al. Genome-wide analysis of long non-coding RNAs in *Catalpa bungei* and their potential function in floral transition using high-throughput sequencing. *BMC Genet*. 2018;19(1):86.
 12. Wu X, Shi T, Iqbal S, Zhang Y, Liu L, Gao Z. Genome-wide discovery and characterization of flower development related long non-coding RNAs in *Prunus mume*. *BMC Plant Biol*. 2019;19(1):64.
 13. Wang ZW, Wu Z, Raitskin O, Sun Q, Dean C. Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor. *Proc Natl Acad Sci U S A*. 2014;111(20):7468–73.
 14. Wang Y, Zhang H, Li Q, Jin J, Chen H, Zou Y, et al. Genome-Wide Identification of lncRNAs Involved in Fertility Transition in the Photo-Thermosensitive Genic Male Sterile Rice Line Wuxiang S. *Front Plant Sci*. 2021;11:580050.
 15. Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, et al. Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *Plant J*. 2018;96(1):203–22.
 16. Wu G, Poethig RS. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development*. 2006;133(18):3539–47.
 17. Xing S, Salinas M, Höhmann S, Berndtgen R, Huijser P. miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in *Arabidopsis*. *Plant Cell*. 2010;22(12):3935–50.
 18. Aung B, Gruber MY, Amyot L, Omari K, Bertrand A, Hannoufa A. Micro-RNA156 as a promising tool for alfalfa improvement. *Plant Biotechnol J*. 2015;13(6):779–90.
 19. Baucher M, Moussawi J, Vandeputte OM, Monteyne D, Mol A, Pérez-Morga D, et al. A role for the miR396/GRF network in specification of organ type during flower development, as supported by ectopic expression of *Populus trichocarpa* miR396c in transgenic tobacco. *Plant Biol (Stuttg)*. 2013;15(5):892–8.
 20. Liang G, He H, Li Y, Wang F, Yu D. Molecular mechanism of micro-RNA396 mediating pistil development in *Arabidopsis*. *Plant Physiol*. 2014;164(1):249–58.
 21. Yang F, Lu C, Wei Y, Wu J, Ren R, Gao J, et al. Organ-Specific Gene Expression Reveals the Role of the Cymbidium ensifolium-miR396/Growth-Regulating Factors Module in Flower Development of the Orchid Plant *Cymbidium ensifolium*. *Front Plant Sci*. 2022;12:799778.
 22. Luan F, Zeng J, Yang Y, He XR, Wang BJ, Gao YB, et al. Recent advances in *Camellia oleifera* Abel: a review of nutritional constituents, biofunctional properties, and potential industrial applications. *J Funct Foods*. 2020;75(9):104242.
 23. Wang A, Ji Z, Xuan R, Zhao X, Hou L, Li Q, et al. Differentially Expressed MiRNAs of Goat Submandibular Glands Among Three Developmental Stages Are Involved in Immune Functions. *Front Genet*. 2021;12:678194.
 24. Li Y, Qin T, Dong N, Wei C, Zhang Y, Sun R, et al. Integrative Analysis of the lncRNA and mRNA Transcriptome Revealed Genes and Pathways Potentially Involved in the Anther Abortion of Cotton (*Gossypium hirsutum* L.). *Genes (Basel)*. 2019;10(12):947.
 25. Verma N. Transcriptional regulation of anther development in *Arabidopsis*. *Gene*. 2019;689:202–9.
 26. Gómez JF, Talle B, Wilson ZA. Anther and pollen development: A conserved developmental pathway. *J Integr Plant Biol*. 2015;57(11):876–91.
 27. Zou F, Yuan DY, Duan JH, Tan XF, Zhang L. A study of microsporogenesis and male gametogenesis in *camellia grijsii* hamce. *Adv J Food Sci Technol*. 2013;5(12):1590–5.
 28. Gao C, Yuan DY, Wang BF, Yang Y, Liu DM, Han ZQ. A cytological study of anther and pollen development in *Camellia oleifera*. *Genet Mol Res*. 2015;14(3):8755–65.
 29. Zhang X, Tong H, Han Z, Huang L, Tian J, Fu Z, et al. Cytological and morphology characteristics of natural microsporogenesis within *Camellia oleifera*. *Physiol Mol Biol Plants*. 2021;27(5):959–68.
 30. Yang J, Kang X. Microsporogenesis and flower development in *Eucalyptus urophylla* \times *E. tereticornis*. *Breed Sci*. 2015;65(2):138–44.
 31. Yao PQ, Li GH, Long QY, He LG, Kang XY. Microsporogenesis and Induction of Unreduced Pollen with High Temperatures in Rubber Tree Clone RRIM 600. *Forests*. 2017;8(5):152.
 32. Arrieta M, Colas I, Macaulay M, Waugh R, Ramsay L. A Modular Tray Growth System for Barley. *Methods Mol Biol*. 2020;2061:367–79.
 33. Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, et al. Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res*. 2014;24(4):688–96.
 34. Yang H, Lu P, Wang Y, Ma H. The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J*. 2011;65(4):503–16.
 35. Barakate A, Orr J, Schreiber M, Colas I, Lewandowska D, McCallum N, et al. Barley Anther and Meiocyte Transcriptome Dynamics in Meiotic Prophase I. *Front Plant Sci*. 2021;11:619404.
 36. Jiang J, Zhang Z, Cao J. Pollen wall development: the associated enzymes and metabolic pathways. *Plant Biol (Stuttg)*. 2013;15(2):249–63.
 37. Ma X, Wu Y, Zhang G. Formation pattern and regulatory mechanisms of pollen wall in *Arabidopsis*. *J Plant Physiol*. 2021;260:153388.
 38. Dong X, Hong Z, Sivaramakrishnan M, Mahfouz M, Verma DP. Callose synthase (Cal5S) is required for exine formation during microgametogenesis and for pollen viability in *Arabidopsis*. *Plant J*. 2005;42(3):315–28.
 39. Yang J, Tian L, Sun MX, Huang XY, Zhu J, Guan YF, et al. AUXIN RESPONSE FACTOR17 is essential for pollen wall pattern formation in *Arabidopsis*. *Plant Physiol*. 2013;162(2):720–31.
 40. Hird DL, Worrall D, Hodge R, Smartt S, Paul W, Scott R. The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana* A6 gene displays similarity to beta-1,3-glucanases. *Plant J*. 1993;4(6):1023–33.
 41. Wang K, Guo ZL, Zhou WT, Zhang C, Zhang ZY, Lou Y, et al. The Regulation of Sporopollenin Biosynthesis Genes for Rapid Pollen Wall Formation. *Plant Physiol*. 2018;178(1):283–94.
 42. Xu J, Ding Z, Vizcay-Barrena G, Shi J, Liang W, Yuan Z, et al. ABORTED MICROSPORES Acts as a Master Regulator of Pollen Wall Formation in *Arabidopsis*. *Plant Cell*. 2014;26(4):1544–56.
 43. Shi J, Cui M, Yang L, Kim YJ, Zhang D. Genetic and Biochemical Mechanisms of Pollen Wall Development. *Trends Plant Sci*. 2015;20(11):741–53.
 44. Morant M, Jørgensen K, Schaller H, Pinot F, Möller BL, Werck-Reichhart D, et al. CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell*. 2007;19(5):1473–87.
 45. Quilichini TD, Friedmann MC, Samuels AL, Douglas CJ. ATP-binding cassette transporter G26 is required for male fertility and pollen exine formation in *Arabidopsis*. *Plant Physiol*. 2010;154(2):678–90.
 46. Phan HA, Iacuone S, Li SF, Parish RW. The MYB80 transcription factor is required for pollen development and the regulation of tapetal programmed cell death in *Arabidopsis thaliana*. *Plant Cell*. 2011;23(6):2209–24.
 47. Mizuno S, Osakabe Y, Maruyama K, Ito T, Osakabe K, Sato T, et al. Receptor-like protein kinase 2 (RPK2) is a novel factor controlling anther development in *Arabidopsis thaliana*. *Plant J*. 2007;50(5):751–66.
 48. Piñero M, Gómez-Mena C, Schaffer R, Martínez-Zapater JM, Coupland G. EARLY BOLTING IN SHORT DAYS is related to chromatin remodeling factors and regulates flowering in *Arabidopsis* by repressing FT. *Plant Cell*. 2003;15(7):1552–62.
 49. Gan ES, Xu Y, Wong JY, Goh JG, Sun B, Wee WY, et al. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in *Arabidopsis*. *Nat Commun*. 2014;5:5098.
 50. Shafiq S, Li J, Sun Q. Functions of plants long non-coding RNAs. *Biochim Biophys Acta*. 2016;1859(1):155–62.
 51. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7–21.
 52. Yuan JH, Yang F, Wang F, Ma JZ, Guo YJ, Tao QF, et al. A long noncoding RNA activated by TGF- β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*. 2014;25(5):666–81.

53. Hanamata S, Kurusu T, Kuchitsu K. Roles of autophagy in male reproductive development in plants. *Front Plant Sci.* 2014;5:457.
54. Hanamata S, Sawada J, Toh B, Ono S, Ogawa K, Fukunaga T, et al. Monitoring autophagy in rice tapetal cells during pollen maturation. *Plant Biotechnol (Tokyo).* 2019;36(2):99–105.
55. Kotchoni SO, Larrimore KE, Mukherjee M, Kempinski CF, Barth C. Alterations in the endogenous ascorbic acid content affect flowering time in *Arabidopsis*. *Plant Physiol.* 2009;149(2):803–15.
56. Dowdle J, Ishikawa T, Gatzek S, Rolinski S, Smirnov N. Two genes in *Arabidopsis thaliana* encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling viability. *Plant J.* 2007;52(4):673–89.
57. Fang W, Yu X, Wang B, Zhou H, Ouyang H, Ming J, et al. Characterization of the *Aspergillus fumigatus* phosphomannose isomerase Pmi1 and its impact on cell wall synthesis and morphogenesis. *Microbiology (Reading).* 2009;155(Pt10):3281–93.
58. Kohorn BD, Johansen S, Shishido A, Todorova T, Martinez R, Defeo E, et al. Pectin activation of MAP kinase and gene expression is WAK2 dependent. *Plant J.* 2009;60(6):974–82.
59. Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol.* 2006;57(1):19–53.
60. Wei LQ, Yan LF, Wang T. Deep sequencing on genome-wide scale reveals the unique composition and expression patterns of microRNAs in developing pollen of *Oryza sativa*. *Genome Biol.* 2011;12(6):R53.
61. Dhaka N, Sharma S, Vashisht I, Kandpal M, Sharma MK, Sharma R. Small RNA profiling from meiotic and post-meiotic anthers reveals prospective miRNA-target modules for engineering male fertility in sorghum. *Genomics.* 2020;112(2):1598–610.
62. Wang JW, Czech B, Weigel D. miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell.* 2009;138(4):738–49.
63. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, et al. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet.* 2010;42(6):541–4.
64. Preston JC, Hileman LC. Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Front Plant Sci.* 2013;4(80):80.
65. Wang Z, Wang Y, Kohalmi SE, Amyot L, Hannoufa A. SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 2 controls floral organ development and plant fertility by activating ASYMMETRIC LEAVES 2 in *Arabidopsis thaliana*. *Plant Mol Biol.* 2016;92(6):661–74.
66. Spanudakis E, Jackson S. The role of microRNAs in the control of flowering time. *J Exp Bot.* 2014;65(2):365–80.
67. Ding XL, Ruan H, Yu LF, Li Q, Song QJ, Yang SP, et al. miR156b from Soybean CMS Line Modulates Floral Organ Development. *J Plant Biol.* 2020;63:141–53.
68. Cuperus JT, Fahlgren N, Carrington JC. Evolution and functional diversification of MIRNA genes. *Plant Cell.* 2011;23(2):431–42.
69. Yu Y, Sun F, Chen N, Sun G, Wang CY, Wu DX. MiR396 regulatory network and its expression during grain development in wheat. *Protoplasma.* 2021;258(1):103–13.
70. Lee SJ, Lee BH, Jung JH, Park SK, Song JT, Kim JH. GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR Specify Meristematic Cells of Gynoecia and Anthers. *Plant Physiol.* 2018;176(1):717–29.
71. Shimano S, Hibara KI, Furuya T, Arimura SI, Tsukaya H, Itoh JI. Conserved functional control, but distinct regulation, of cell proliferation in rice and *Arabidopsis* leaves revealed by comparative analysis of GRF-INTERACTING FACTOR 1 orthologs. *Development.* 2018;145(7):dev159624.
72. Zhang D, Sun W, Singh R, Zheng Y, Cao Z, Li M, et al. GRF-interacting factor1 Regulates Shoot Architecture and Meristem Determinacy in Maize. *Plant Cell.* 2018;30(2):360–74.
73. Leyser HM, Lincoln CA, Timpte C, Lammer D, Turner J, Estelle M. *Arabidopsis* auxin-resistance gene AXR1 encodes a protein related to ubiquitin-activating enzyme E1. *Nature.* 1993;364(6433):161–4.
74. Martinez-Garcia M, Fernández-Jiménez N, Santos JL, Pradillo M. Duplication and divergence: New insights into AXR1 and AXL functions in DNA repair and meiosis. *Sci Rep.* 2020;10(1):8860.
75. Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, et al. Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. *Plant J.* 2012;69(3):462–74.
76. Yang ZL, Zeng XQ, Chen FF, Li GY. Status of *Camellia oleifera* resources in Hainan Island. *Nonwood Forest Research.* 2015;33(3):138–44.
77. Zheng DJ, Pan XZ, Xie LS, Zeng JH, Wu YJ, Zang ZL, et al. Investigation and analysis of *Camellia oleifera* industry development status in Hainan Province. *Nonwood Forest Research.* 2015;33(1):131–5.
78. Li SF, Iacuone S, Parish RW. Suppression and restoration of male fertility using a transcription factor. *Plant Biotechnol J.* 2007;5(2):297–312.
79. Ranjan R, Khurana R, Malik N, Badoni S, Parida SK, Kapoor S, et al. bHLH142 regulates various metabolic pathway-related genes to affect pollen development and anther dehiscence in rice. *Sci Rep.* 2017;7:43397.
80. Salmela L, Rivals E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics.* 2014;30(24):3506–14.
81. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150–2.
82. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
83. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39(suppl):W316–22.
84. Eddy SR. Profile hidden Markov models. *Bioinformatics.* 1998;14(9):755–63.
85. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 2000;28(1):45–8.
86. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44(D1):D279–285.
87. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25–9.
88. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004;32(Database issue):D277–80.
89. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics.* 2005;21(19):3787–93.
90. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics.* 2014;15(1):311.
91. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 2013;41(17):e166.
92. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 2017;45(W1):W12–6.
93. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics.* 2011;27(17):2325–9.
94. Lin P, Wang K, Wang Y, Hu Z, Yan C, Huang H, et al. The genome of oil-*Camellia* and population genomics analysis provide insights into seed oil domestication. *Genome Biol.* 2022;23(1):14.
95. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100.
96. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 2017;45(8):e57.
97. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell.* 2009;136(4):629–41.
98. Shuai P, Liang D, Tang S, Zhang Z, Ye CY, Su Y, et al. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J Exp Bot.* 2014;65(17):4975–83.
99. Chen J, Quan M, Zhang D. Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and normal wood xylem by RNA-seq. *Planta.* 2015;241(1):125–43.
100. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
101. Wen M, Shen Y, Shi S, Tang T. miREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. *BMC Bioinformatics.* 2012;13:140.

102. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 2012;40(1):37–52.
103. Zhou L, Chen J, Li Z, Li X, Hu X, Huang Y, et al. Integrated profiling of microRNAs and mRNAs: microRNAs located on Xq27.3 associate with clear cell renal cell carcinoma. *PLoS One.* 2010;5(12):e15224.
104. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* 2018;46(W1):W49–54.
105. Chen J, Chen B, Zhang D. Transcript profiling of *Populus tomentosa* genes in normal, tension, and opposite wood by RNA-seq. *BMC Genomics.* 2015;16(1):164.
106. Androvic P, Valihrach L, Elling J, Sjoback R, Kubista M. Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Res.* 2017;45(15):e144.
107. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc.* 2008;3(6):1101–8.
108. Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, et al. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Genomics Proteomics Bioinformatics.* 2021;19(4):578–83.
109. CNGB-NGDC Members and Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformatics in 2022. *Nucleic Acids Res.* 2022;50(D1):D27–38.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

