Comparative Profile Analysis Reveals Differentially Expressed MicroRNAs Regulate Anther and Pollen Development in Kenaf Cytoplasmic Male Sterility Line
Comparative Profile Analysis Reveals Differentially Expressed MicroRNAs Regulate Anther and Pollen Development in Kenaf Cytoplasmic Male Sterility Line

Peng Chen1*, Qiqi Shi1, Zhichen Liang1, Hai Lu1, Ru Li2

1 Key Laboratory of Plant Genetics and Breeding, College of Agriculture, Guangxi University, Nanning, China
2 College of Life Science and Technology, Guangxi University, Nanning, China

*Author for correspondence (Peng Chen: hustwell@gmail.com)
Qiqi Shi, email: 1290782669@qq.com
Zhichen Liang, email: 328636497@qq.com
Hai Lu, email: 294856020@qq.com
Ru Li, email: liruonly@163.com

Abstract

Cytoplasmic male sterility (CMS) is advantageous in extensive crop breeding and represents a perfect model for understanding anther and pollen development research. MicroRNAs (miRNAs) play key roles in regulating various biological processes. However, the miRNA-mediated regulatory network in kenaf CMS occurrence remains largely unknown. In the present study, a comparative deep sequencing approach was used to investigate the miRNAs and their roles in regulating anther and pollen development during CMS occurrence. We identified 283 known and 46 new candidate miRNAs in kenaf anther. A total of 67 differentially expressed miRNAs (DEMs) were discovered between CMS and its maintainier line. Among them, 40 and 27 miRNAs were up- and down regulated, respectively. These 67 DEMs were predicted to target 189 genes. Validation of DEMs and putative target genes were confirmed by using real-time quantitative PCR. In addition, a potential miRNA-mediated regulatory network, which mainly involves the auxin signaling pathway, signal transduction, glycolysis and energy metabolism, gene expression, transmembrane transport, protein modification and metabolism, and floral development, that mediates anther development during CMS occurrence was proposed. Taken together, our findings provide a better understanding of the molecular mechanism of miRNA regulation in pollen development and CMS occurrence in kenaf.

Key words: kenaf; cytoplasmic male sterility(CMS); microRNA; anther; pollen

Introduction

https://mc06.manuscriptcentral.com/genome-pubs
Pollen development is a crucial process in flowering plants that directly determines the male fertility (McCormick 2004). Cytoplasmic male sterility (CMS) is a maternally inherited trait in higher plants that leads to the inability to produce or release functional pollen for sexual reproduction. In agricultural production, CMS is used widely to facilitate the production of hybrid seeds because it eliminates the need for emasculation by hand. Apart from its commercial uses, CMS represents an attractive system for investigating pollen development and cytoplasm-nucleus interactions (Chase 2007).

Kenaf (Hibiscus cannabinus L.) is an important phloem fiber crop from the Malvaceae family with a 4000-year cultivation history. It has the characteristics of fast growth, salt and alkali tolerance, barren land tolerance, waterlogging resistance, and drought resistance with high fiber yield and good quality. Therefore, kenaf can be cultivated on saline and alkaline land or barren sloping land, without competing with food crops for land. It is also considered a valuable multipurpose crop due to numerous end uses, including paper, cordage, building material, animal bedding, carpet backing, forage for livestock, and biomass (Ramesh 2016). Our research group initially discovered a spontaneous male sterility mutant kenaf line (named UG93A) bearing fewer anthers that were indehiscent with abortive pollen. Genetic analysis showed that the mutant caused CMS. Using this mutant, we bred several CMS lines and pioneered kenaf “three lines” hybrid seeds utilization on a large scale. This original work greatly promoted the utilization of heterosis in kenaf (Zhou et al. 2008). It is fundamentally necessary to understand kenaf pollen development and the CMS mechanism to better utilize and produce novel CMS lines. Unfortunately, the kenaf CMS molecular mechanism remains unclear. To date, various efforts have been made to understand the pollen development and molecular mechanism of plant CMS, and these efforts have been based predominantly on the hypotheses of mitochondrial DNA (mtDNA) rearrangement, chimeric genes, RNA editing, and abnormal protein translation (Sandhu et al. 2007; Fujii et al. 2010; Heng et al. 2018). However, pollen development is actually controlled by cross-talk between mitochondria and nucleus. The inter-organellar communication between mitochondria and the nucleus is extensive, multifaceted and highly regulated. Among the many regulatory factors, microRNAs (miRNAs) are an important regulator found in recent years.

MiRNAs are a class of small (18-24 nucleotides), endogenous and non-coding RNAs that bind to target mRNAs to regulate protein expression by repressing translation, or promoting the degradation of target mRNA (Bartel 2004). The miRNA pathway is an essential mechanism that regulates different biological processes (Breakfield et al. 2012). In plants, it is involved in regulation of developmental
processes, including sexual organ development (He and Hannon 2004; Banks 2008). MiRNAs are also shown to be present and functional in mature pollen. In Arabidopsis, 33 different miRNA families were detected in mature pollen, and the majority of those families showed pollen-enriched expression when compared with leaves (Grant-Downton et al. 2009). Using next-generation sequencing technology, Wei et al. (2011) identified 292 known miRNAs and 75 novel miRNAs in rice pollen. In cotton, 16 conserved miRNA families were identified in fertile wild type (WT) and genetic male sterility (GMS) mutant anther, some conserved miRNAs were significantly differentially expressed between the GMS mutant and the WT plant during anther development (Wei et al. 2013). Similar results have also been reported in Brassica juncea (Yang et al. 2013), B. rapa (Jiang et al. 2014), Radish (Zhang et al. 2016) and Hibiscus syriacus (Kim et al. 2017).

Recently, the number of unique, mature miRNAs has grown up to 28,645 entries in the miRBase database (version 21). In contrast, there is no mature miRNAs registered for kenaf till now. Therefore, there is an urgent need to identify and further to reveal the functions of novel miRNAs regulating pollen development and other biological processes in kenaf. In the present study, to identify known and potential novel miRNAs from kenaf anthers and further to investigate DEMs in pollen development regulation, we systematically characterized the miRNA profiles of kenaf anthers from CMS mutant line (UG93A) and wild type maintainer line (UG93B) using Illumina sequencing. We obtained 283 conserved miRNAs and 46 novel miRNAs in kenaf, 67 of which were differentially expressed in the CMS and maintainer lines. The targets of these DEMs were also characterized. Furthermore, the miRNA-mediated regulatory network of CMS occurrence during anther development was constructed in kenaf. These results should lay a meaningful foundation for elucidating the roles of miRNAs in regulating pollen development in the kenaf CMS line.

Materials and methods

Plant Materials

The plant materials used in this study were UG93A (CMS mutant line) and UG93B (wild-type maintainer line, WT). Plants were grown from June to November in an experimental field of Guangxi University (108°22′E, 22°48′N) under natural conditions (the mean temperature during growth is approximately 25°C in loamy soil with a pH=6.9) and normal field management. During flowering stage (mid-October), the microspore mother cells of dual-core anthers (pollen abortion stage) were
collected for RNA isolation. Anthers were sampled at 10:00 AM, three different plants were harvested then pooled as one biological repeat, and a total of nine plants of UG93A or UG93B were sampled in triplicate. All of the materials were sampled at the same time.

**Small RNA library construction and Illumina sequencing**

For microRNA analysis, total RNA was isolated from anthers of UG93A and UG93B respectively using the mirVana miRNA Isolation Kit (Ambion, Austin, TX, USA) according to the manufacturer’s protocol. Total RNA quantity and purity was assayed with the NanoDrop ND-1000 spectrophotometer (Nano Drop) at 260/280 nm (ratio >2.0). For small RNA library construction and deep sequencing, we followed protocols of the NewGLife Company (Beijing, China). In brief, approximately 20 μg of small RNA between 18-30 nt was isolated by polyacrylamide gel electrophoresis and then ligated to 5’ and 3’ adapters. The adaptor-ligated small RNA was then converted to cDNA using Superscript II reverse transcriptase (Invitrogen). The resulting cDNA was amplified with 15 PCR cycles using Illumina’s small RNA primer sets. A total of six small RNA libraries (three for the CMS line and three for the maintainer line) were generated and subjected to Illumina sequencing according to the manufacturer’s protocol. All raw data from the small RNA libraries were deposited at the NCBI Sequence Read Archive (SRA) with the accession number SRP157875.

**Analysis of sequencing data**

The trimmed and filtered using the Perl script provided in the mirTools package (Zhu et al. 2010). During this procedure, all low quality reads were removed, including reads with 3’ and 5’ adapter contaminants, those without insert tags, and those with poly-A sequences. The resulting reads were further processed to retain sequences with length between 18 and 30 nt, inclusive. The resulting data sets were named unique reads. Unique reads were then used to query the NCBI mRNA (https://www.ncbi.nlm.nih.gov/nucest/), noncoding RNA sequences database (http://rfam.xfam.org/) and repeat Repbase (http://www.girinst.org/repbase/). The remaining unique sequences were then mapped to the kenaf reference sequences that derived from whole transcriptome of kenaf (supplemental Table S1). The matched sequences were queried against known miRNAs in miRBase 21 (http://www.mirbase.org/index.shtml) for kenaf known miRNAs identification. Then, the remaining unannotated reads were used to predict potential novel miRNAs using miRDeep* program (https://sourceforge.net/projects/mirdeepstar/files/) according to the previous criteria (Meyers et al.,
followed by the prediction of secondary structures by Mfold program (http://unafold.rna.albany.edu/?q=mfold).

In the DEMs analysis, the expression of miRNAs in two samples was normalized to obtain the transcripts per million (TPM) using the following formula: normalized miRNA expression level = (miRNA count/total count of clean reads) × 1000000 (Nie et al. 2015). The expression between the CMS and WT libraries was calculated as follows: Ratio = (miRNA normalized reads in the CMS library)/ (miRNA normalized reads in the WT library). The miRNAs with relative ratios (|CMS/WT|) greater than 2 and with P<0.05 were considered DEMs (Nie et al. 2015).

The potential target genes of the DEMs were predicted by the plant small RNA target analysis server (psRNATarget; http://plantgrn.noble.org/psRNATarget/) (Dai and Zhao, 2011), based on methods described by Lu et al.(2005) and with penalty scores ≤ 3 for mismatched patterns in the miRNA/mRNA duplexes. The functions of the predicted target genes were assigned according to the functions of the best hits from the BLAST search against the Nr database or Swiss-Prot annotation, with the expected value set at 0.001. To further understand the biological functions of the targets, gene ontology (GO) analysis by Blast2GO program (Conesa et al. 2005) and KEGG analysis (KOBAS2.0; http://kobas.cbi.pku.edu.cn/ home.do) were performed.

Quantitative Real-time PCR validation

DEM were subjected to quantitative RT-PCR validation using the High-Specificity miRNA QuantiMir RT Kit (RA610A-1, System Biosciences) on a LightCycler®480 system (Roche Diagnostics). At the same time, the expression patterns of DEMs’ targets were also characterized. The primers of 10 selected miRNAs and their target genes, as well as an internal control (actin), were used. Real-time PCR reactions were performed according to a previously established method (Chen et al. 2005). All the gene expression data were obtained from three individual replicates.

Results

Deep sequencing and overview of kenaf small RNAs

Six small RNA libraries from UG93A and UG93B were generated using pooled RNA isolated from the kenaf anther. Sequencing of the kenaf small RNA libraries was performed with the Illumina Hiseq2500 platform. For direct comparison, kenaf anthers used for UG93A and UG93B libraries construction were grown under the same conditions. The same amount of RNA was used to construct these two libraries.
and the samples were prepared in the same manner.

Illumina sequencing of UG93A and UG93B libraries generated a total of 17,514,332 and 19,625,176 raw reads, respectively. After data cleaning, 13,897,205 and 15,563,728 clean reads were generated from the two sRNA libraries, accounting for 79.34% and 79.30 % of the total reads ranging from 18 to 30 nucleotides in the UG93A and UG93B libraries, respectively. The remaining sequences either of low quality sequences, adapter sequences or sequences smaller than 18 nt, were excluded from further analyses. From the size distribution of reads, we found that the majority of reads, accounting for more than 60%, in the sRNA libraries were in range of 21 to 24 nt in length, which were within the typical size range for dicer-derived products; these results were in agreement with most previously reported findings (Tang et al. 2012; Wei et al. 2013; Jiang et al. 2014). Amongst these reads, small RNAs with 24 nt were the most abundant, followed by those with a 21 nt length (Fig.1 and Table S1). These small RNAs were used for further analyses. Subsequently, the small RNA sequences were compared with the Rfam database (http://rfam.xfam.org/) and GenBank (http://blast.ncbi.nlm.nih.gov). Reads that matched to ribosomal RNA (rRNA), transfer RNA (tRNA), small nucleolar RNA (snoRNA), repeat sequences, or mRNA (e-value less than 1e-3) were removed. Approximately 24.29 % of the clean reads matched from the UG93A library matched non-coding RNAs and repeat sequences, including rRNAs (21.52 %), small nuclear RNAs (snRNAs) (0.25 %), snoRNAs (0.09%), tRNAs (2.36%), and repeat sequences (0.06%). Approximately 25.23% of the clean reads from the UG93B library matched to noncoding RNAs and repeat sequences, including rRNAs (22.89%), snRNAs (0.26 %), snoRNAs (0.09 %), and tRNAs (1.92 %), and repeat sequences (0.07%). All the unannotated sequences, containing 10,521,978(75.71%) and 11,636,803(74.77%) clean reads of from the UG93A and UG93B libraries, respectively, were used for further conserved and novel miRNA prediction (Table 1).

**Identification of known and novel miRNAs from kenaf**

To identify the miRNAs from kenaf, sRNA sequences identified from kenaf by deep sequencing were first compared with the currently known mature plant miRNAs (miRBase 21.0). As the results, 283 kenaf miRNAs with high sequence identity to known plant miRNAs were identified by miRDeep program according to the primary criteria described previously (Meyers et al. 2008) (Supplemental Table S2). The lack of genome information complicates the identification miRNAs in kenaf. To solve this issue, we made use of previous de novo transcriptome sequencing of kenaf (Chen et al. 2014) and
used these transcripts as references for miRNA and their targets prediction. The whole transcriptome contains 54,563 unigenes and was well annotated (Supplemental Table S3). Over 100 million unique small RNA sequences of 18–30 nt in length were detected in each material, and approximately 50% of them could be mapped to the transcriptome of kenaf. To confirm the authenticity, we further mapped the identified 283 miRNAs to kenaf transcriptome, and all these could be found in corresponding unigenes. These miRNAs were found to belong to twelve miRNA families, including miR156, miR160, miR166, miR167 etc. and were defined as conserved miRNAs. Among these families, miR166, miR156, miR169, miR172 and miR167 were the top five, each containing 60, 43, 39, 37 and 33 members respectively (Supplemental Table S2). After removing the conserved miRNA sequences, the remaining reads were further used to search against the transcriptome for candidate novel miRNAs prediction using miRDeep based on the criteria described by Meyers et al. (2008). The mfold software, was also used to confirm the novel miRNAs which identified by miRDeep program. In total, 46 novel miRNAs were identified. Their precursor and mature miRNA sequences were listed in Supplemental Table S4. Each novel pre-miRNA sequence, as well as its most abundant mature miRNA sequence, was subjected to a homology-based search against miRBase using BLASTn (e-value less than 1e-3). None of these miRNA sequences matched sequences in other species, indicating that these can be regarded as authentic candidates for kenaf-specific miRNAs. The results suggest that using the transcriptome as a reference genome may be a suitable choice for miRNA identification. These newly discovered miRNAs in kenaf anthers should have important significance in further research.

**DEMs during kenaf pollen development of UG93A and UG93B**

MiRNAs from UG93A and UG93B were used for differential expression analysis according to a previously described method (Audic and Claverie 1997). The results were considered significantly different when the $|\log_2$ fold-change (UG93A/UG93B) $| \geq 1$ and had a P value <0.01. Finally, 67 miRNAs (40 conserved and 27 novel) were differentially expressed between UG93A and UG93B. Among these, 40 and 27 miRNAs were up- and downregulated (UG93A compared with UG93B, the same below), respectively (Fig.2 and Supplemental Table S5). A total of 40 differentially expressed conserved miRNAs were found to belong to 11 miRNA families, including hac-miR156, hac-miR164, hac-miR166, hac-miR167, hac-miR169 and hac-miR397. The four most represented members families were hac-miR167, hac-miR172, hac-miR397 and hac-miR166 with 12, 9, 7 and 5 miRNAs members,
respectively. Most of the conserved miRNAs, such as hac-miR164, hac-miR167, hac-miR172, hac-miR394, hac-miR403 and hac-miR482, have been shown to play important roles in plant development and were upregulated in CMS line UG93A compared to maintainer line UG93B. For example, the conserved hca-miR167, which was found to play an important role in anther development in other plants (Liu et al. 2014; Wei et al. 2013; Wu et al. 2006), was significantly upregulated in anthers from the CMS line in the present study. Furthermore, hca-miR167 contained 12 DEMs and represented the largest DEM family. The miRNAs expression tendency within the same family was identical, with the only exception of hca-miR156. Hca-miR156b was found to be downregulated 2.61-fold, while hca-miR156e was upregulated 3.03-fold. For the newly identified 27 novel miRNAs, 14 and 13 were up- and downregulated, respectively, in CMS line UG93A.

**Prediction of miRNA Targets**

Plant miRNAs show a high degree of sequence complementarity to their targets, which allows for effective target prediction (Rhoades et al. 2002). Therefore, we performed target prediction of the DEMs between the UG93A CMS line and the UG93B maintainer line (WT). Because there was no genome sequence for kenaf, we used the kenaf whole transcriptome as a reference sequence instead. The targets expression levels were identified in our previous research, which were conducted using UG93A CMS line and the UG93B maintainer line (WT) by transcriptome sequencing (Chen et al. 2014). The miRNA targets were predicted using the psRNATarget (http://plantgrn.noble.org/psRNATarget/), a plant small RNA target analysis server, as described by Dai and Zhao (2011). As a result, 67 DEMs were predicted to target 189 genes, as shown in Supplemental Table S6. Most of the DEMs have several different target genes, for example, conserved hca-miR156, hca-miR164 and novel hac-miR55 have 19 predicted target genes; on the other hand, some of them showed the same target genes (Supplemental Table S6). The target genes encode proteins such as ARFs (auxin response factors), pollen-specific protein SF3, Floral homeotic protein APETALA, ABC transporter family members, transcription factors, and phosphofructokinase, and these proteins have documented roles in flower and pollen development according to previous studies (Wu et al. 2006; Kuromori et al. 2011; Huang et al. 2016). The conserved plant miRNAs may regulate homologous targets at identical target sites in different species (Axtell and Bowman 2008). In Arabidopsis, miRNA167 controls patterns of ARF6 and ARF8 expression, and is essential for fertility of both ovules.
and anthers (Wu et al. 2006). In the present study, the conserved hca-miR167 families were also predicted to target and regulate ARF6 (auxin response factor 6, Unigene_002669, Unigene_032770 and Unigene_002671) and ARF8 (auxin response factor 8, Unigene_004486 and Unigene_002670) expression. Along with conserved homologous targets, non-conserved targets were also identified for conserved miRNAs. Two additional genes (Unigene_045783 and Unigene_016400) were also predicted to be targeted by hca-miR167, indicating that conserved miRNAs also regulate non-conserved targets in kenaf. At the same time, the newly identified miRNAs were found to regulate many new targets, as well as some identified targets of conserved miRNAs. For example, the ARF6 was predicted to be a target of the novel hac-miR24 and the conserved hca-miR167.

Validation of DEMs and putative target genes

To verify the expression change of the identified kenaf miRNAs, the same RNA preparations used in the miRNA sequencing were subjected to qRT-PCR assay. In this study, we randomly selected the 10 miRNAs (6 upregulated and 4 downregulated) from different families and validated them by qRT-PCR. Five conserved miRNAs (hca-miR156b, hca-miR166a, hca-miR167, hca-miR172a, and hca-miR397) and five novel miRNAs (hac-miR13, hac-miR28, hac-miR30, hac-miR43, and hac-miR55) were validated and quantified using qRT-PCR. The miRNA expression patterns were found to be similar to the sequencing results, indicating that the small RNA sequencing results were reliable. (Fig. 3).

To test if any correlation between miRNAs and their targets existed, the expression patterns of 10 identified miRNA targets based on qRT-PCR were compared. Because miRNA is responsible for the degradation its target mRNA transcripts, their expression levels could be negatively correlated. As expected, the expression levels of most miRNA were inversely correlated with these of the corresponding miRNAs. The results showed that during the anther developmental stage, hac-miR156b expressed relatively lower in the CMS line than in the WT line, while its target gene encoding an Adaptin-4 complex subunit (Unigene_000349) was expressed at a higher level in the WT line than in the CMS line, as expected (Fig. 4). In contrast, anthers from the CMS line had higher expression levels of the miRNAs, including hac-miR167, hac-miR172, hac-miR394, hac-miR13, hac-miR28 and hac-miR41; their corresponding target genes had significantly lower expression levels in the CMS line than in the WT line (Fig. 4). Positive correlations in expression levels between some miRNAs and their target genes were also observed, such as hac-miR166 and its target Unigene_007489 encoding a...
Homeobox-leucine zipper protein ATHB-15, hac-miR24 and its target Unigene_000560 encoding a Pollen-specific protein SF3, and hac-miR54 and its target Unigene_033016 MYB encoding a transcription factor MYB52. The abnormal relationships of expression levels between miRNAs and their target genes may also indicate that there are other mechanisms regulating the expression of the target genes.

GO annotation and enrichment analysis of the predicted target genes of DEMs

We conducted gene ontology (GO) annotation of the target genes of DEMs to reveal their roles in regulating pollen development in kenaf CMS line. Target genes were categorized by biological process (BP), molecular function (MF), and by their belonging to specific cellular component (CC) at GO level 2 (Fig. 5 and Supplemental Table S7). The top three molecular functions for identified target genes were catalytic activity (41.30%), binding (40.0%) and transporter activity (5.1%) (Fig. 5A). The top three cellular components were cell/cell part (33.3%), organelle (18.2%) and organelle part (5.8%) (Fig. 5B). The top three biological process categories were cellular process (26.4%), metabolic process (24.3%) and biological regulation (9.6%) (Fig. 5C).

We further conducted GO enrichment analysis based on proteins with P values less than 0.05 according to molecular functions, cellular components, and biological processes. A total of 24 terms, including cellular component biogenesis, growth, reproductive process, developmental process, biological regulation, metabolic process, organelle, transporter activity, and catalytic activity, were enriched (Fig. 5D).

Discussion

CMS is important for exploiting heterosis in crop breeding and serves as a model for investigating nuclear-cytoplasmic interactions. Plant miRNAs control multiple aspects of plant development and stress responses. They are also prominent regulators of pollen development (Štorchová 2017). In the present study, to extend our knowledge of the molecular basis for CMS occurrence in kenaf, as well as the key regulatory molecular aspects of miRNAs, six independent small RNAs libraries were constructed at anther development stage from a kenaf CMS line (UG93A) and its corresponding maintainer (UG93B) line. Kenaf CMS occurs (pollen begin abortion) at dual-core of the microspore mother cells stage, so we just focus this critical stage, and intend to compare the difference between
these near-isogenic lines, CMS line and the wild type in the present study. In the meantime, there is no doubt that comparison microRNA profiles of different stage (before pollen abortion stage and pollen abortion stage) could also provide dynamic information of pollen development. To the best of our knowledge, the present study represents the first such attempts. We think the present information is also very important to elucidate the miRNAs in regulating of pollen and anther development between CMS line and the maintainer line, and thus resulting CMS occurrence.

**The auxin signaling pathway is involved in pollen development**

Auxin plays critical roles during plant growth and development, and many of those roles are mediated by the auxin response transcription factor (ARF) family (Mallory et al. 2005). In *Arabidopsis*, miR167 regulates the auxin signal by cleaving ARF6 and ARF8 transcripts during anther development, and miR167 overexpression causes male sterility (Ru et al. 2006; Wu et al. 2006). In cotton, miR167 is upregulated in the genetic male sterile (GMS) line compared to WT and decreases the expression level of its target, ARF4 (Wei et al. 2013). The downregulation of auxin response factors 6 and 8 (ARF6 and ARF8) by upregulating miRNA 167 leads to floral development defects and female sterility in tomato (Liu et al. 2014). hac-miR167 was predicted to target ARF6 and ARF8. Compared with wild-type anthers, hac-miR167 was expressed at a relatively higher level in CMS mutant anthers, which led to a 2-fold downregulation of ARF6 and ARF8. In addition to the conserved hac-miR167, the newly found hac-miR24 was also predicted to regulate ARF6 and ARF8. The downregulated expression of ARF6 and ARF8 by hac-miR167 and hac-miR24 may affect the auxin response pathway in the CMS mutant (Fig. 6A).

**Signal transduction regulation**

Signal transduction pathways are the basic mechanisms controlling cell growth, proliferation, metabolism and many other processes. The miRNAs and their regulated targets that involved in signal transduction pathways were summarized in Fig. 6B. Calmodulin (CaM) is an important intermediate of calcium-mediated signal transduction. In plants, the role of CaM in regulating a variety of calcium-dependent signaling pathways within the cell has been reported (Ma et al. 1999; Liu et al. 2003). The CaM signal transduction pathway plays key roles in pollen germination and tube growth (Ma et al. 1999; Yang et al. 2014). Calmodulin-binding transcription activators (CAMTAs, also called signal response genes, SRs) belong to a conserved family of transcription factors. CAMTAs are
regulated by both biotic and abiotic stresses, such as heat, cold, high salinity, drought, UV, wounding, and phytohormones (Galon et al. 2010). Both CaM and CAMTA5 were downregulated by hac-miR156, hac-miR164, hac-miR6, and hac-miR55 in the kenaf CMS line.

**Gene expression regulation**

Anther and pollen development are complex biological processes involving a series of specific gene expression regulations. Many miRNAs were found to play key roles in regulating this process. In the present study, miRNAs such as hac-miR5, hac-miR10, hac-miR12, hac-miR17, hac-miR18, hac-miR39, hac-miR40 and hac-miR46 were found to target different transcription factors (TFs) and regulate gene expression. The characterized TF targets included Dof zinc finger protein, poly (A)-binding protein, MYB, and NACs (summarized in Fig. 6C). Taking hac-miR18 and its target MYB52 as an example, hac-miR18 was characterized as being upregulated by 2.59-fold, while its target MYB52 was found to be decreased by 1.60-fold. Plant MYB proteins comprise a large, multifunctional family and play very important roles in many biological processes, such as primary and secondary metabolism, developmental control and cell fate determination (Chen et al. 2014). In cotton, a MYB transcription factor is specifically expressed in pollen and is involved in the regulation of late anther/pollen development (Li et al. 2013). An MYB domain protein that regulates sugar partitioning is required for rice pollen development (Zhang et al. 2010). NAC domain-containing proteins (NACs) are associated with diverse processes, including various developmental programs, senescence, and biotic and abiotic stress responses (Puranik et al. 2012). Seven NAC domain-containing proteins (Unigene_005037, Unigene_005051, Unigene_005055, Unigene_029333, Unigene_032044, Unigene_032062 and Unigene_005050) were found to be the targets of hac-miR5, hac-miR36, hac-miR39 and hac-miR46 and were upregulated in the CMS line in the present study.

**Glycolysis and energy metabolism regulation**

Glycolysis and energy metabolism are the most basic metabolic pathways in biological metabolism. Their main physiological function is to provide the required energy and carbon sources. Phosphofructokinase (PFK) is one of the most important regulatory enzymes of glycolysis and is the foundation for respiration. PFK catalyzes ATP-dependent phosphorylation to convert fructose-6-phosphate into fructose 1,6-bisphosphate and ADP, which is one of the key regulatory and rate-limiting steps of glycolysis. PFK is able to regulate glycolysis through allosteric inhibition, and in
this way, the cell can increase or decrease the rate of glycolysis in response to the cell's energy requirements (Lenzen 2014). Pyruvate kinase is the enzyme that catalyzes the final step of glycolysis. It catalyzes the transfer of a phosphate group from phosphoenolpyruvate (PEP) to adenosine diphosphate (ADP), yielding one molecule of pyruvate and one molecule of ATP (Valentini et al. 2000). In higher plants, the demand for ATP is greatly increased during pollen development, and decreased ATP synthesis may be a causal factor in the disruption of pollen or microspore development (Chen et al. 2014). In this study, hac-miR13 and hac-miR41 were identified to target PFK2 (6-phosphofructokinase 2; Unigene_001113) and pyruvate kinase (Unigene_027486), respectively, and regulate the glycolysis pathway (Fig. 6D). Compared with the maintainer line, hac-miR13 and hac-miR41 were characterized by 4.24-fold and 2.54-fold upregulation, respectively, in the CMS line; accordingly, the targets PFK2 and pyruvate kinase were 6.97-fold and 1.49-fold downregulated, respectively. The decreased expression of PFK2 and pyruvate kinase should decrease ATP production and cause pollen to develop abnormally. In plants, there are few reports of miRNA in regulating glycolysis pathways. The regulatory role of newly discovered miRNAs in glycolysis deserves further analysis.

Material transport regulation

As shown in Fig. 6E, many miRNAs were characterized as regulating substrate and ion transporter expression in the present study. The ABC (ATP-binding cassette) transporter is a highly conserved family of ATP-binding proteins that use ATP-derived energy to transport molecules across cell membranes. These proteins transport various substrates, such as organic anions, heavy metals, hormones, fatty acids, ions and secondary metabolites, and thus contribute to almost all aspects of plant development and stress responses (Kuromori et al. 2011). For example, in Arabidopsis, ABCG26 is required for pollen development, and the ABCG26 mutant results in a male sterile phenotype (Kuromori et al. 2011). Hac-miR10 and hac-miR13 were predicted to target ABC transporters 9 (Unigene_054515) and 26 (Unigene_000787), respectively. Interestingly, ABC transporter 9 was upregulated 6.54-fold, while ABC transporter 26 was downregulated 4.55-fold. In addition, other substrate transporters, such as potassium transporter (Unigene_010796), calcium-transporting ATPase (Unigene_006560, Unigene_028785), boron transporter (Unigene_034388), and glucose transporter (Unigene_012392), were differentially regulated by hac-miR41. Intracellular ion homeostasis plays vital roles in the growth and development of higher plants, and a disruption in the balance may lead to
abnormal pollen development.

**Protein modification and metabolism regulation**

Most cellular proteins are continuously synthesized and degraded within the lifespan of a cell. Efficient protein turnover is essential for many aspects of cell physiology and development (Hellmann and Estelle 2002). Many novel miRNAs and their targets were identified in the present study and involved in protein modification and metabolism regulation (summarized in Fig. 6F). The ubiquitin/proteasome pathway represents one of the most important proteolytic systems in eukaryotes and was proposed to be involved in pollen germination and pollen tube growth (Sheng et al. 2006). Three ubiquitin genes were shown to be downregulated by nine miRNAs in kenaf in the present study. In addition, two protein phosphatases, protein phosphatase 2A and 2C, were upregulated by hac-miR24 and hac-miR17, respectively. As a phosphatase enzyme, protein phosphatase removes a phosphate group from the phosphorylated amino acid residue of its substrate protein and plays key roles in protein posttranslational modification (Fuchs et al., 2012; Chen et al. 2018). The protein phosphatase 2C was found to play an important role in pollen germination in rice (Fujii and Toriyama 2008).

**Floral development regulation**

In the present study, four genes (Unigene_007489, Unigene_00749, Unigene_007492 and Unigene_007493) encoding ATHB15 and two genes (Unigene_010682 and Unigene_010684) encoding homeobox-leucine zipper protein REVOLUTA (HD-ZIP protein REV) were identified as conserved targets of hac-miR166 (summarized in Fig. 6G). Previous studies in *Arabidopsis* have shown that miR166 and its targets play an important role in shoot apical and lateral meristem formation, organ polarity, and vascular development (McConnell et al. 2001; Kim et al. 2005; Jung and Park, 2007). The overexpression of miR166 drastically reduced ATHB15 levels in CMS plants compared with those in wild-type plants and resulted in reduced floral organs (Williams et al. 2005). However, in a strange way, both hac-miR166 and its target ATHB15 were found to be 2- to 4-fold downregulated in CMS pollen when compared with WT pollen. This may indicate that a specific regulatory pattern exists in kenaf CMS pollen. However, the relationship between cytoplasmic male sterility and the lower level of hac-miR166 in CMS mutant anthers relative to WT anthers is currently unknown and needs further study. To date, miR172 and miR164 are the two miRNAs with the clearest functions that are known to be involved in floral organ formation. MiR172 controls inner whorl organ formation by restricting the
expression of the APETALA2 (AP2) gene, whereas the miR164 gene targets NAM-related genes to establish boundaries between the whorls of organs and between organs within whorls, which define the sizes of the primordia and the resulting floral organs (Luo et al. 2013; Huang et al. 2016). Nine miR172 family hac-miRNAs and one hac-miR164 were identified to be upregulated in CMS kenaf pollen, and their targets, such as AP2 and WD repeat-containing protein 11, were found to be downregulated. hac-miR172, hac-miR18 and hac-miR18 target a pentatricopeptide repeat-containing (PPR) protein, which has been widely reported to be involved in many stages of mRNA processing, including splicing and RNA editing (O'Toole et al. 2008). PPR proteins from various higher plants also act to suppress the expression of mitochondrial genes associated with cytoplasmic male sterility (Hu et al. 2012; Huang et al. 2015).

**Conclusion**

In this study, we employed comparative high-throughput sequencing approaches to characterize conserved and novel candidate miRNAs and their targets between the CMS mutant line and the wild-type maintainer line in kenaf. A total of 283 conserved and 46 novel miRNAs of kenaf were identified, 67 of which were differentially expressed in CMS and maintainer lines. The targets of these DEMs were also characterized. As summarized in Fig. 6, the DEMs and their targets participated in various biological processes, such as the auxin signaling pathway, signal transduction, glycolysis and energy metabolism, gene expression, transmembrane transport, protein modification and metabolism, and floral development. Our results provide insight into the CMS mechanism and pollen development in kenaf from the perspective of miRNA regulation.

**Acknowledgments**

This work was supported by the National Natural Science Foundation of China (31560341), and the earmarked fund for Modern Agro-industry Technology Research System (CARS-16-E14). We thank Professor Ruiyang Zhou for providing the kenaf UG93A and UG93B seeds in present study.

**Compliance with ethical standards**

The authors have declared that no competing interests exist.

**References**
Fuchs, S., Grill, E., Meskiene, I., and Schweighofer, A. 2012. Type 2C protein phosphatases in plants.


Legends to Figures and Tables

Fig.1 Size distribution of small RNA sequences derived from the UG93A and UG93B libraries.

Figure 2 Differentially expressed miRNAs between CMS line and maintainer line (MT) in kenaf. A. differentially expressed conserved miRNAs, B. differentially expressed novel miRNAs.

Figure 3 Validation of miRNAs expression by qRT-PCR

Figure 4. Expression correlation between miRNAs and their targets

Figure 5 Target genes GO analysis of the differentially expressed miRNAs. A. GO based on biological process. B. GO based on molecular function. C. GO based on cellular component.

Figure 6 The proposal schematic model of miRNA-mediated regulatory network of anther and pollen development during CMS occurrence in kenaf

Table 1 Statistics of small RNA sequences from UG93A and UG93B
Supplementary figure legend

1. gen-2018-0207.R3Suppl a
   statistical of raw reads and length distribution of mapped reads

2. gen-2018-0207.R3Suppl b
   identified conserved miRNAs

3. gen-2018-0207.R3Suppl c
   Kenaf Unigene sequence and annotation

4. gen-2018-0207.R3Suppl d
   Novel microRNAs in kenaf

5. gen-2018-0207.R3Suppl e
   Differentially Expressed miRNAs

6. gen-2018-0207.R3Suppl f
   target annotation and expression

7. gen-2018-0207.R3Suppl g
   Targets GO and GO enrichment
Table 1 Statistics of small RNA sequences from UG93A and UG93B

<table>
<thead>
<tr>
<th>Types</th>
<th>UG93A</th>
<th>UG93B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clean Reads</td>
<td>Percentage</td>
</tr>
<tr>
<td>Total</td>
<td>13897205</td>
<td>100.00%</td>
</tr>
<tr>
<td>rRNA</td>
<td>2991008</td>
<td>21.52%</td>
</tr>
<tr>
<td>snRNA</td>
<td>34976</td>
<td>0.25%</td>
</tr>
<tr>
<td>snoRNA</td>
<td>12363</td>
<td>0.09%</td>
</tr>
<tr>
<td>tRNA</td>
<td>328660</td>
<td>2.36%</td>
</tr>
<tr>
<td>Repeat sequence</td>
<td>8220</td>
<td>0.06%</td>
</tr>
<tr>
<td>Unannotated</td>
<td>10521978</td>
<td>75.71%</td>
</tr>
</tbody>
</table>
Fig. 1

140x99mm (300 x 300 DPI)
Fig. 2

209x170mm (300 x 300 DPI)
Fig. 3

209x140mm (300 x 300 DPI)
Fig. 4

209x140mm (300 x 300 DPI)
Fig. 5

209x199mm (300 x 300 DPI)
Regulation of CMS occurrence in kenaf

Fig. 6

209x209mm (300 x 300 DPI)