

COMPUTATIONAL IDENTIFICATION OF microRNA IN FIVE WOODY OIL TREE CROPS AND THEIR miRNA TARGET SEQUENCES

CHUNTAI WU*; RUI LIU*; YU LI* and RIZHONG ZENG*

ABSTRACT

Recently discovered miRNA (microribonucleic acids) are a kind of endogenous small RNA that have important regulatory roles during eukaryote development and stress responses by controlling mRNA degradation or inhibiting translation. However, little is known about miRNA from commercially important woody oil crops. Using a comparative genomics approach and following a series of stringent filtering criteria, we identified 36 woody oil crop miRNA belonging to 22 families using expressed sequence tags and genome survey sequence databases. The numbers of miRNA found in walnut (*Juglans regia*), oil palm (*Elaeis guineensis*), tea (*Camellia sinensis*), tung oil tree (*Vernicia fordii*) and barbados nut (*Jatropha curcas*) were 17, 11, 3, 3 and 2, respectively. In addition, 275 potential target genes were predicted for 27 newly identified miRNA utilising mRNA databases of *J. curcas*, *E. guineensis*, *C. sinensis*, *J. regia*, *V. fordii* and *Olea europaea* plant species. Most of the target mRNA encoded transcription factors that regulated plant growth and development, together with a few target genes that were involved in metabolism or stress responses. Sequence alignments of 10 miRNA families from different species showed that miRNA sequences within the same family were variable. Furthermore, the authenticity of 11 novel miRNA candidates from oil palm was verified by semi-quantitative and quantitative reverse transcription-polymerase chain reaction (RT-PCR). A correlation analysis revealed a close correlation between the two PCR assay methods. The expression levels of miRNA transcripts significantly differed between the leaf and the fruit tissues of *E. guineensis*.

Keywords: microRNA, computational identification, woody oil crop, target gene, reverse transcription-polymerase chain reaction.

Date received: 6 March 2017; **Sent for revision:** 12 April 2017; **Received in final form:** 18 August 2017; **Accepted:** 18 September 2017.

INTRODUCTION

Microribonucleic acids (miRNA) are a broad class of endogenous, and highly-conserved, non-coding single-stranded small RNA (~21 nucleotides in length), which negatively regulate gene expression at the post-transcriptional level by binding to their

target mRNA, leading to translational inhibition and/or mRNA cleavage (Bartel, 2004; Filipowicz *et al.*, 2008). The miRNA arise from 60–160 nucleotide-long miRNA precursors (pre-miRNA) that can fold into stem-loop (hairpin) structures. To date, many types of miRNA have been reported in plants (Zhao *et al.*, 2007), with the majority of characterised miRNA involved in controlling developmental processes (Willmann and Poethig, 2007), signalling pathways (Yoshikawa *et al.*, 2005) or stress responses (Leung and Sharp, 2010). Although a large quantity of plant miRNA have been increasingly identified *in silico* and deposited in miRBase V17.0 (<http://>

* Key Laboratory of Biology and Genetic Resources of Rubber Tree, Scientific Observation and Experiment Station of Tropical Crops in Danzhou, Ministry of Agriculture, Rubber Research Institute, Chinese Academy of Tropical Agriculture Sciences, Baodaoxincun, Danzhou 571737, P R China.
E-mail: hnzrz@aliyun.com

mirbase.org/index.shtml), only a small number have been discovered and functionally identified *in planta*. The precise functions of many other plant miRNA remain to be elucidated.

The most popular experimental technology to seek novel miRNA is direct cloning (Lu *et al.*, 2005a,b; Zhang *et al.*, 2006a). However, it is difficult to detect miRNA having a low level of expression using this technique (Adai *et al.*, 2005). High-throughput next generation sequencing, and comparative genome-based computational strategies have been shown to be successful for the discovery of many low abundance or tissue-specific miRNA in plants (Fahlgren *et al.*, 2007; Wei *et al.*, 2009). Compared to high-throughput next-generation sequencing, comparative genomics does not require extensive technical expertise since all of the sequences used for computational prediction of miRNA are readily accessible from the National Centre for Biotechnology Information (NCBI) GenBank Databases (<https://www.ncbi.nlm.nih.gov/genbank/>). Comparative genomics is based on expressed sequence tags (EST) and genome survey sequence (GSS) databases (Liang *et al.*, 2007). Hundreds of characterised plant miRNA have been identified by screening EST and GSS databases of plant species with known miRNA. Computational approaches thus provide an ideal method to identify miRNA in plants, particularly in species which do not have fully sequenced genomes.

Tree crops include a number of valuable woody oil crops or oil tree crops, such as barbados nut (*Jatropha curcas*), oil palm (*Elaeis guineensis*), tea (*Camellia sinensis*), walnut (*Juglans regia*), and the tung oil tree (*Vernicia fordii*). Woody oil crops are economically important crops globally and are characterised by having high oil content in their seed and/or fruit, which can be extracted and processed into edible oils, lubricants, or biodiesel, and are important resources for the cosmetics, paint and other industries. Small RNA have been studied in oil palm (Nasaruddin *et al.*, 2007; Low *et al.*, 2014) but the number of identified *Elaeis* miRNA is still limited compared with those of other crops, such as rice and wheat, and very little is known about their potential target genes. For *J. curcas*, *C. sinensis*, *J. regia* and *V. fordii*, there are no experimentally or computationally identified miRNA sequences or structural information. To identify novel miRNA in woody oil crops, all annotated plant miRNA sequences were aligned with the EST and GSS databases of *J. curcas*, *E. guineensis*, *C. sinensis*, *J. regia* and *V. fordii* using a computer-based approach. Furthermore, using the *Arabidopsis thaliana* and *Oryza sativa* DNA databases, and the protein-coding nucleotide (mRNA) databases of many woody oil crops, target mRNA for the identified miRNA were predicted via specialised algorithms such as miRU (Zhang, 2005). In addition, to validate the

computationally identified miRNA and to better understand the regulation of miRNA during the leaf and fruit development of oil palm, we also studied the spatial expression profiles of miRNA.

MATERIALS AND METHODS

Reference Set of miRNA

To identify potential miRNA in the five woody oil tree crops, 4677 mature plant miRNA sequences were downloaded from the miRNA Registry Database (<http://mirbase.org/index.shtml>; Release 18.0; November 2011), including 85 from *Chlamydomonas*, 41 from *Picea*, 38 from *Pinus*, 280 from *Physcomitrella*, 64 from *Selaginella*, 703 from *Arabidopsis*, 78 from *Brassica*, 63 from *Ricinus*, 32 from *Arachis*, 408 from *Glycine*, 674 from *Medicago*, 42 from *Gossypium*, 82 from *Theobroma*, 45 from *Aquilegia*, 79 from *Citrus*, 242 from *Populus*, 37 from *Solanum*, 186 from *Vitis*, 143 from *Brachypodium*, 23 from *Hordeum*, 661 from *Oryza sativa*, 36 from *Saccharum*, 172 from *Sorghum*, 45 *Triticum*, 321 from *Zea*, and the remainder from 14 other model plant species (Griffiths-Jones, 2006; Griffiths-Jones, *et al.*, 2008). Duplicated miRNA sequences were removed using the CD-hit program (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi) (Huang *et al.*, 2010), and the remaining 2348 miRNA were used as the reference for miRNA sequence sets. A false discovery rate analysis was performed according to the procedure developed by Kim and van de Wiel (2008).

Sources of EST and GSS Data for Woody Oil Crops

The EST and GSS databases from six woody oil crops were obtained using GenBank at NCBI (<https://www.ncbi.nlm.nih.gov/>). As of April 2011, a total of 179 896 sequences including 130 171 EST and 49 725 GSS have been deposited in NCBI GenBank.

Software Used for the Characterisation of miRNA and Related mRNA in Six Woody Oil Tree Crops

Comparative software (BLAST-2.2.23) (Camacho *et al.*, 2009) was downloaded from NCBI GenBank to identify conserved miRNA in GSS and EST sequences from five woody oil crops, excluding *Olea europaea*. A perl script was used to remove any overlapping or duplicated mature miRNA sequences. The miRNA secondary structures and stabilities were predicted using RNAstructure 5.3 software (<http://rna.urmc.rochester.edu/rnastructure.html>). BLASTX (<http://www.ncbi.nlm.nih.gov/BLAST/>) was used to search for/analyse putative mRNA targets (Mathews *et al.*, 1999; Zuker, 2003).

Prediction of Putative miRNA by EST and GSS Database Analyses

We performed BLAST algorithm-based searches against the EST and GSS databases for non-redundant mature miRNA in order to identify woody oil crop orthologues of these miRNA. This used the strategy developed by Meyers *et al.* (2008), since mature miRNA are more highly conserved across plant species compared with pre-miRNA sequences. The adjusted BLASTN parameters were expected values of 1000, and 1000 descriptions and alignments. The default word size for matches between the query and database sequences was adjusted to seven. All BLAST results were saved for further analysis. Sequences with less than four nucleotide mismatches against the query miRNA were selected manually, but any identical sequences were removed. Candidate EST and GSS sequences were searched against the UniProt protein database (<http://www.uniprot.org>) to remove protein-coding sequences.

The secondary structures of the remaining sequences in the dynamically extended folding process were predicted using the RNAstructure 5.03 software (Mathews *et al.*, 1999; 2004). To distinguish miRNA from other types of RNA, the following parameters were applied (Meyers, *et al.*, 2008; Zhang *et al.*, 2005): (1) in view of the evolution of miRNA among plant species, less than three nucleotide mismatches between the query mature miRNA and the predicted mature miRNA were permitted; (2) only pre-miRNA which, when folded into a stem-loop hairpin secondary structure, contained the ~22 nucleotide mature miRNA sequence at the stem part of the hairpin were permitted; (3) a minimal-free energy index (MFEI) with their approximately >0.67 and higher negative minimum folding energy (MFE) of miRNA precursors was used to distinguish miRNA from other small RNA; (4) the A+U content of the pre-miRNA was between 30%-70%; (5) the mature miRNA sequence contained a number of nucleotide mismatches equal to or less than six with the opposing miRNA* sequence on the other hairpin arm; and, finally (6), no loops or breaks were present in the mature miRNA sequence.

Prediction of miRNA Targets

Since plant miRNA recognise their target mRNA sequences by perfect or near-perfect complementarity (Bartel, 2004), homology searching is a powerful approach by which to predict potential miRNA targets. A Plant Small RNA Target Analysis Server (psRNATarget; available at <http://plantgrn.noble.org/psRNATarget/>) has been developed to predict miRNA targets in various plant species. Predictions were performed using psRNATarget against the *Arabidopsis* and *Oryza* DNA databases

at the Institute for Genome Research (TIGR; <http://www.jcvi.org/>) or against the mRNA database for all woody oil crops, available at NCBI. The default parameters were first applied for the prediction of potential targets, but were later amended so that less than four mismatches (without gaps) were permitted between the miRNA sequence and any potential mRNA target (Zhang, 2005). We then performed BLASTX searches of the target mRNA sequences against the predicted protein signatures in the InterPro database (<http://www.ebi.ac.uk/interpro/>) to predict the functions of the potential targets.

miRNA Isolation and cDNA Synthesis

Independent miRNA samples from young oil palm leaves and fruit tissues were prepared for separate analyses using rapid miRNA extraction kits (Bioteke, Beijing, China) with miRNA lysis (MRL) and RNA washing (RW) buffers following the manufacturer's protocol. DNA removal from miRNA was performed using DNase I (Invitrogen, Carlsbad, CA, USA). The cDNA were synthesised from 0.5 µg of miRNA by multiplexed reverse transcription (RT) reaction using a pool of specific stem-loop primers. The RT reaction was carried out in a Biometra T1 Thermocycler (Biometra, Göttingen, Germany). The stem-loop primers for RT are shown in Table 1. The 50-mer RT primers for miRNA were designed according to a previous study (Chen *et al.*, 2005) and included a 44-mer stem-loop fragment, 5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC-3', and terminal six bases at the 3' end of the target miRNA. All of the DNA oligonucleotide primers were synthesised by Invitrogen Co.

Semi-quantitative RT-PCR

Semi-quantitative RT-PCR were performed in 25 µl volumes containing 12.5 µl 2×Taq PCR MasterMix (Tiangen Biotech, Beijing, China), 1 µl each primer (10 µmol litre⁻¹), and 1 µl cDNA (50 ng µl⁻¹). The specific miRNA primers for RT-PCR are shown in Table 1. The oil palm U6 gene was used to normalise the expression values (Table 1). Negative cDNA synthesis controls (not submitted to the RT action) for the U6 primer pair were also added to detect possible genomic DNA contamination. Amplification was performed using a Biometra T1 Thermocycler (Biometra) under the following conditions: 95°C for 3 min, followed by 35 cycles of 94°C for 15 s, 56°C for 15 s and 72°C for 20 s, ending with a final extension at 72°C for 7 min. PCR products were separated by electrophoresis on 1.0% (w/v) agarose gels with ethidium bromide staining. The separated PCR products were visualised under ultraviolet light, and the densitometry

TABLE 1. SEQUENCES OF PRIMERS USED IN REVERSE TRANSCRIPTION AND QUANTITATIVE (RT-PCR)

Primer	Primer sequence (5' → 3')
egu-156a RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGCTCT-3'
egu-156a F	5'-GGCCAGTGACAGAAGAGAG-3'
egu-156b RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGTGCTC-3'
egu-156b F	5'-CGCGCCTGACAGAAGATAG-3'
egu-160 RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGCATA-3'
egu-160 F	5'-GCTTGCCTGGTCCCTG-3'
egu-168 RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTTCCCG-3'
egu-168 F	5'-GTTTCATCGCTTGGTGCAGGT-3'
egu-169a RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGCCAA-3'
egu-169a F	5'-TGCCTCAGGGAAGTTCACC-3'
egu-169b RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCTAGCA-3'
egu-169b F	5'-GAGCTCGGCAGTCTGCGT-3'
egu-397a RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCATCAA-3'
egu-397a F	5'-GGCGGTTCATCGAGTGCAGCG-3'
egu-397b RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCATCAA-3'
egu-397b F	5'-GTAAACATCGAGTGCAGCG-3'
egu-398a RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGGGG-3'
egu-398a F	5'-GCTGCTGTGTTCTCAGGTTCG-3'
egu-398b RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGGGG-3'
egu-398b F	5'-TATATTGTGGTCCGAGGTTCG-3'
egu-414 RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGAAGA-3'
egu-414 F	5'-GCTCGCTTCATCATCATC-3'
URP	5'-GTGCAGGGTCCGAGGT-3'
U6 RT	5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAAAATTT-3'
U6 F	5'-ATTGGAACGATACAGAGAAGATTAG-3'
U6 R	5'-AATTTGGACCATTTCTCGATTGTG-3'

Note: URP - universal reverse primer. RT - reverse transcription. PCR - polymerase chain reaction.

value was determined for each PCR product using Alpha Imager™ 2200 (Alpha Innotech Corp., San Leandro, CA, USA). Then, the target fragments were recovered from agarose gels using a DNA gel extraction kit (Promega, Madison, WI, USA), and they were cloned independently into the pMD18-T vector (TaKaRa, Kyoto, Japan) by T-A ligation. The fragment sequences were determined using the Illumina HiSeq™ 2000 system (San Diego, CA, USA) with primers M13F (-47) and M13R (-48).

Real-time Quantitative RT-PCR

To explore the spatial expression profiles of the 11 newly identified miRNA in oil palm, we performed experiments on samples from four tissues collected from three mono-bodies of 13-year old *E. guineensis* using stem-loop real-time quantitative PCR. Each real-time PCR reaction was conducted in a reaction mixture containing 1.0 µl 20 ng µl⁻¹ template cDNA, 10 µl 2×SYBR green qPCR Mix (Roche), 0.5 µmol litre⁻¹ of each primer (Invitrogen, Shanghai, China) and 8.5 µl dH₂O. The final volume of reaction mixture was 20 µl. The PCR amplification was carried out on a Light Cycler 2.0 instrument (Roche Diagnostics, Mannheim, Germany) under the following cycling conditions: 3 min at 95°C, followed by 35 cycles at 94°C for 10 s, 58°C for 15 s, 72°C for 30 s, and a final extension at 72°C for 10

min. The sequences of primers specific for individual miRNA and the U6 RNA internal control are shown in Table 1. Each sample was amplified in triplicate. The data were analysed using the LightCycler 4.05 software (Roche Diagnostics).

Statistical Analyses

Data were analysed statistically with SAS 8.0 software (SAS Institute Inc., Cary, NC, USA). Comparisons were made using Duncan’s new multiple range method. Correlations were calculated using the Pearson coefficient. In all of the tests, a probability value of less than 0.05 was considered to be statistically significant.

RESULTS AND DISCUSSION

Identification of Potential miRNA in Woody Oil Crops

In the present study, a homology search procedure was applied to identify conserved miRNA in five woody oil crops. We used 2348 known mature miRNA sequences with any duplicated sequences removed (false discovery rate *q*-value = 1.40884E-04) from 33 different plant species as a reference dataset, against which BLASTN searches

were performed using 120 630 EST from *J. curcas* (46 862), *E. guineensis* (40 737), *C. sinensis* (25 411), *J. regia* (5213) and *V. fordii* (2407); and 50 906 GSS from *J. curcas* (1569), *C. sinensis* (1) and *J. regia* (49 336). After the first round of searches, a total of 1247 probable orthologues were obtained. Within these 1247 predicted miRNA, 504 miRNA were identified through an EST database search and 743 miRNA through a GSS database search.

After the subsequent elimination of duplicates (16) and protein-coding sequences (1195), the total number of potential miRNA was reduced to 36. Of these, 20 miRNA were found in the EST database, and 16 were present in the GSS database. Hence, 0.017% of EST and 0.031% of GSS contain potential miRNA. From the structural (RNAstructure 5.3) analysis, 36 potential miRNA have been identified in total (Table 2).

The screening protocol for potential miRNA was also based on the A+U content. The percentage composition of the four nucleotides (A, C, G and U) in the pre-miRNA of woody oil crops is not uniform (Table 2). Uracil is the dominant base in the pre-miRNA sequences from the five woody oil crops and comprises $29.51 \pm 5.44\%$ of total nucleotides, followed by adenine, guanine and cytosine ($20.90 \pm 4.76\%$). The sequences of the miRNA precursors had A+U content ranging from 36.5% to 67.3%, with an average of $54.4 \pm 7.54\%$ (Table 2), which shows high similarity with proportions observed in other plant species (Stellwag, 2008). Consistent with this general notion, the majority of the identified miRNA precursors in woody oil crops contain more A+U nucleotides than G+C (Zhang, 2006b, c). It is also important to note that the formation of the stem-loop hairpin secondary structure is one of the most important characteristics of pre-miRNA, although this structure is not a unique feature of miRNA, since other RNA such as mRNA, rRNA, and tRNA can also form similar hairpin structures. For this reason, uniform systems have been established for annotating new miRNA. The two criteria of negative minimal fold energy (MFE) and minimal fold energy index (MFEI) have generally been accepted. The woody oil crops' pre-miRNA have low and negative MFE (-14.2 – -79.2 kcal mol⁻¹), with an average of -40.01 ± 15.43 kcal mol⁻¹, which indicates high thermodynamic stability of the corresponding sequences. The MFEI is also an important criterion for distinguishing miRNA from the other types of RNA. In this study, the identified pre-miRNA had a high MFEI (0.70–1.28), with an average of 0.90 ± 0.17 . These results suggest that the woody oil crop miRNA identified in this research were selected with high stringency. The 36 miRNA represent 16 miRNA families in five woody oil crops (Figure 1). The miR414 has six members; miR167 and miR169 have four members; miR156 and miR398 have three members; miR159, miR171,

miR395, miR397, and miR399 have two members; and the other miRNA families such as miR160, miR168, miR170, miR319, miR390, and miR403 have only one member. We obtained a short orthologous sequence for miR167 from *J. regia* and for miR169 from *E. guineensis* EST (Table 2). All of the identified miRNA in this study were reported for the first time in the five crop species. Large differences in the number of predicted miRNA sequences were observed between the five tree species with 17 miRNA identified in *J. regia* and 11 in *E. guineensis* but only 2-3 miRNA predicted in *J. curcas* and *C. sinensis*.

The length of the miRNA precursors varied a lot in the woody oil crops, ranging from 52 to 180 nucleotides with an average of 96.7 ± 32.0 nucleotides, similar to that observed in *Triticum aestivum*, *Glycine max* and *Nicotiana tabacum* (Wei *et al.*, 2009; Stellwag, 2008; Frazier *et al.*, 2010) (Table 2; Supplementary Figure S1, which is available via <http://jopr.mpob.gov.my/wp.content/uploads/2018/03/FigureS1.pdf>). It is suggested that the size difference of the identified miRNA within the different families may provide unique functions for the regulation of miRNA biogenesis or gene expression in plants. All of the mature miRNA sequences identified are in the stem portion of the hairpin structures with the miRNA located either at the 5' end or 3' end of the stem-loop hairpin in the pre-miRNA sequences. Of the 36 identified woody oil crop miRNA, 19 (52.8%) are at the 5' end of the miRNA precursor sequences, as shown in Supplementary Figure S1 which available via <http://jopr.mpob.gov.my/wp.content/uploads/2018/03/FigureS1.pdf>), whilst the remaining 17 miRNA (47.2%) are at the 3' end. The mature miRNA fluctuated in length from 18 to 22 nucleotides (Table 2).

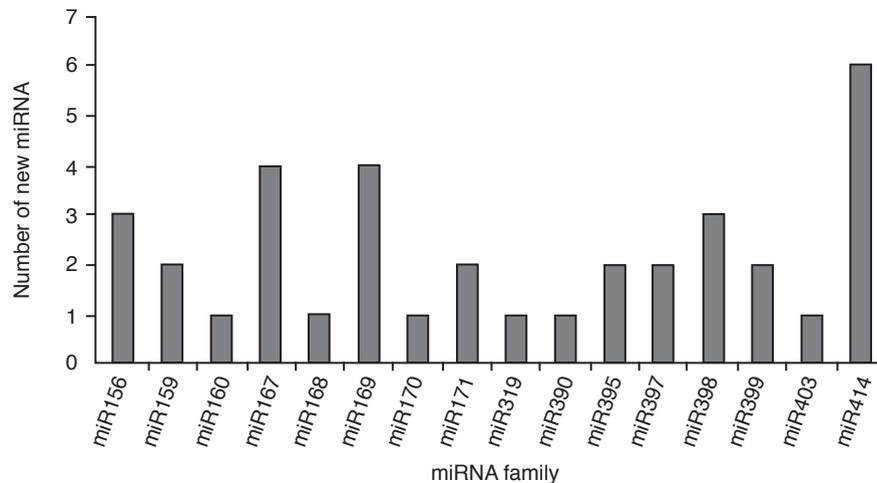
Novel Conserved Woody Oil Crop miRNA

The 36 newly identified oil tree crop miRNA contain different nucleotide identity ratios and related genetic distances (Figure 2). Some miRNA show high sequence similarity, such as miRNA159 and miRNA319, which differ by only one nucleotide, whereas others show long genetic distances and very low nucleotide identity (Figure 2). This may be an indication of their relationship and evolution. The 36 miRNA were classified into 22 miRNA families based on their sequence homology with at least 77.78% identity. We compared the sequence conservation within these 16 newly identified miRNA families (with the exception of miRNA family jre-miR403, jre-miR390, jcu-miR319, jre-miR170, egu-miR168, and egu-miR160) across the five woody oil crop species. The results of the sequence alignments of the 10 miRNA families showed that, although the nucleotide sequences of the miRNAs within the same family were quite variable, each member of

TABLE 2. NEWLY IDENTIFIED (computer-based) miRNA (microribonucleic acid) FROM EXPOSED SEQUENCE TAP (EST) AND GENOME SURVEY SEQUENCE (GSS) DATABASES OF FIVE DIFFERENT TREE OIL CROPS

Species	New miRNA	Mature sequence (5' → 3')	Accession	Gene source	Homologous miRNA	NM* (nt)	LM* (nt)	LP* (bp)	Location	A (%)	C (%)	G (%)	U (%)	G+C (%)	A+U (%)	MFEI* (Kcal mol ⁻¹)	MFEI*
<i>Elais guineensis</i>	egu-miR156a	UGACAGAAAGAGAGAGAGCA	EL683138.1	EST	zma-miR156k	2	19	59	5	40.7	10.2	25.4	23.7	64.4	35.6	15.3	0.729
	egu-miR156b	UGACAGAAAGAUAGAGAGCAC	EY412069.1	EST	zma-miR156k	2	20	80	5	27.5	25.0	20.0	27.5	55.0	45.0	46.2	1.283
	egu-miR160	UGCCUGGCCUCCUGAUGCC	DW247953.1	EST	ath-miR160a	1	20	135	5	18.5	23.0	39.3	19.3	37.8	62.2	59.8	0.712
	egu-miR168	UCGCUUGGUCAGGUCGGGAA	EY413183.1	EST	ath-miR168a	0	21	104	5	15.4	29.8	33.7	21.2	36.5	63.5	49.8	0.755
	egu-miR169a	UCAGGGAAGUUCACCUUGGCU	EL686486.1	EST	ath-miR169*	3	21	167	3	21.6	20.4	28.1	29.9	51.5	48.5	57.7	0.712
	egu-miR169b	GGCAGUCUGGUCUUCUAG	EL681134.1	EST	zma-miR169*	3	18	101	5	18.8	19.8	32.7	28.7	47.5	52.5	37.6	0.709
	egu-miR397a	UCAUCGAGUCGACCGUUGAUG	CN599915.1	EST	ath-miR397b	2	21	111	5	19.8	25.2	25.2	29.7	49.5	50.5	45.3	0.809
	egu-miR397b	AUCGAGUCGACCGUUGAUG	CN599915.1	EST	osa-miR397b	3	19	107	5	18.7	25.2	26.2	29.9	48.6	51.4	44.5	0.809
	egu-miR398a	UGUGUUCUCAGGUCGCCCCUG	CN600323.1	EST	osa-miR398b	0	21	101	3	16.8	33.7	27.7	21.8	38.6	61.4	52.4	0.845
	egu-miR398b	UGUGGUCGCAGGUCGCCCCUG	CN600323.1	EST	osa-miR398b	2	21	101	3	21.8	27.7	33.7	16.8	38.6	61.4	52.4	0.845
<i>Camellia sinensis</i>	csi-miR1414	UCAUCAUCAUCAUCAUCUCC	EY410728.1	EST	osa-miR414	2	21	146	5	24.7	23.3	23.3	28.8	53.4	46.6	63.2	0.929
	csi-miR156	UGACAGAAAGAUAGAGAGCA	H5396956.1	EST	ptc-miR156k	3	19	96	5	26.0	20.8	26.0	27.1	53.1	46.9	32.7	0.727
	csi-miR414a	UCUUCUUCUCAUCAUCAUC	GH709869.1	EST	ath-miR414	3	20	66	5	25.8	25.8	18.2	30.3	56.1	43.9	23.1	0.700
	csi-miR414b	UCAUCAUCAUCAUCAUCAUC	GH710834.1	EST	ath-miR414	2	21	123	5	22.0	10.6	23.6	43.9	65.9	34.1	36.1	0.859
	jre-miR159	UUGGAGAGAAAGGAGCUGGA	HR193094.1	GSS	ptc-miR159f	3	20	83	3	27.7	21.7	21.7	28.9	56.6	43.4	28.1	0.781
<i>Juglans regia</i>	jre-miR167a	AAGCUGCCAGCAUGAUCU	CV197291.1	EST	ath-miR167c	3	18	76	5	32.9	19.7	17.1	30.3	63.2	36.8	33.0	1.179
	jre-miR167b	UGAAGCUGCCAGCAUGAUCU	CV197291.1	EST	ath-miR167d	2	20	80	5	32.5	20.0	17.5	30.0	62.5	37.5	37.6	1.253
	jre-miR167c	GAUCAUGUCCAGGUGCAC	HR219203.1	GSS	zma-miR167h*	2	19	61	3	29.5	19.7	21.3	29.5	59.0	41.0	20.8	0.832
	jre-miR167d	GAUCAUGGGGAGUUUCAC	HR222747.1	GSS	zma-miR167h*	3	19	70	3	25.7	21.4	22.9	30.0	55.7	44.3	32.0	1.032
	jre-miR169a	CAGCCAAAGAUAGACUUGCCG	HR223202.1	GSS	ath-miR169a	1	20	98	3	24.5	15.3	26.5	33.7	58.2	41.8	49.8	1.215
	jre-miR169b	GAGCCAAAGAAUGAAUUGCCG	HR199456.1	GSS	ptc-miR169t	2	20	54	5	27.8	18.5	24.1	29.6	57.4	42.6	16.9	0.735
	jre-miR170	UGAUUGAGCCGUGCCAAUAUC	HR216622.1	GSS	ath-miR170	1	21	78	3	23.1	17.9	25.6	33.3	56.4	43.6	42.7	1.256
	jre-miR171a	UUGAGCCGUGCCAAUAUAUCAG	HR216622.1	GSS	ath-miR171b	0	21	84	3	22.6	19.0	27.4	31.0	53.6	46.4	44.6	1.144
	jre-miR171b	UAUUGGCCGUGCCAAUAUAAGA	HR216622.1	GSS	zma-miR171c*	2	21	74	3	23.0	17.6	25.7	33.8	56.8	43.2	25.5	0.750
	jre-miR390	AAGCUCAGGAGGUAAGCGCC	HR209951.1	GSS	ath-miR390a	0	21	81	5	21.0	21.0	25.9	32.1	53.1	46.9	37.0	0.974
<i>Jatropha curcas</i>	jre-miR395a	CUUAUUGGUUUGGUGGAACUC	HR207015.1	GSS	ptc-miR395a	3	21	52	5	23.1	13.5	19.2	44.2	67.3	32.7	14.2	0.835
	jre-miR395b	UGAAGUUGUUGGGGAAACUC	HR210159.1	GSS	ath-miR395b	2	20	89	3	23.6	22.5	22.5	31.5	55.1	44.9	43.7	1.093
	jre-miR398	UGUGUUCUCAGGUCACCCCUU	HR211540.1	GSS	ath-miR398a	0	21	87	3	27.6	16.1	26.4	29.9	57.5	42.5	34.3	0.927
	jre-miR399	UGCCAAAGGAGAGUUGCCUUG	HR214457.1	GSS	ath-miR399b	0	21	100	3	27.0	20.0	25.0	28.0	55.0	45.0	40.5	0.900
	jre-miR403	UUAGAUUCACGCACAACUCCG	HR202791.1	GSS	ath-miR403	0	21	74	3	25.7	24.3	21.6	28.4	54.1	45.9	31.5	0.927
<i>Vernicia fordii</i>	jcu-miR414	UCAUCAUCAUCAUCAUCAUC	CB303876.1	EST	ath-miR414	2	21	79	3	24.1	24.1	24.1	27.8	51.9	48.1	33.0	0.868
	jcu-miR159	UUGGACUGAAGGAGGCUCCU	JM428844.1	GSS	ath-miR159c	2	20	180	3	28.3	21.7	24.4	25.6	53.9	46.1	79.2	0.954
	jcu-miR319	UUGGACUGAAGGAGGCUCC	JM428844.1	GSS	ath-miR319a	2	19	178	3	28.1	21.9	24.7	25.3	53.4	46.6	79.2	0.954
<i>Vernicia fordii</i>	vfo-miR399	GGGCAACUUCGAUCCUUUGCCA	GR218775.1	EST	zma-miR399*	1	22	123	5	22.8	21.1	25.2	30.9	53.7	46.3	41.7	0.730
	vfo-miR414a	CAUCUUCAUUCUUCUUCUUC	GR218086.1	EST	ath-miR414	3	19	96	5	27.1	17.7	18.8	36.5	63.5	36.5	29.8	0.850
	vfo-miR414b	UCAUCUUCAUUCUUCUUCUUC	GR218086.1	EST	ath-miR414	3	21	86	5	29.1	17.4	19.8	33.7	62.8	37.2	29.3	0.920

Note: NM - number of mismatches; LM - length of mature miRNA; LP - length of pre-miRNA; MFE - minimal folding free-energy; MFEI - minimal folding-free energy index.



Note: In this study, only 16.67% of the family members were categorised into six different miRNA families having single members, while 83.33% of the family members were categorised into 10 different independent miRNA families. The number of members per miRNA family was greater than or equal to two. Among them, family members of three miRNA families were detected in no less than two species.

Figure 1. Size distribution of newly identified miRNA (microribonucleic acids) families in woody oil crops.

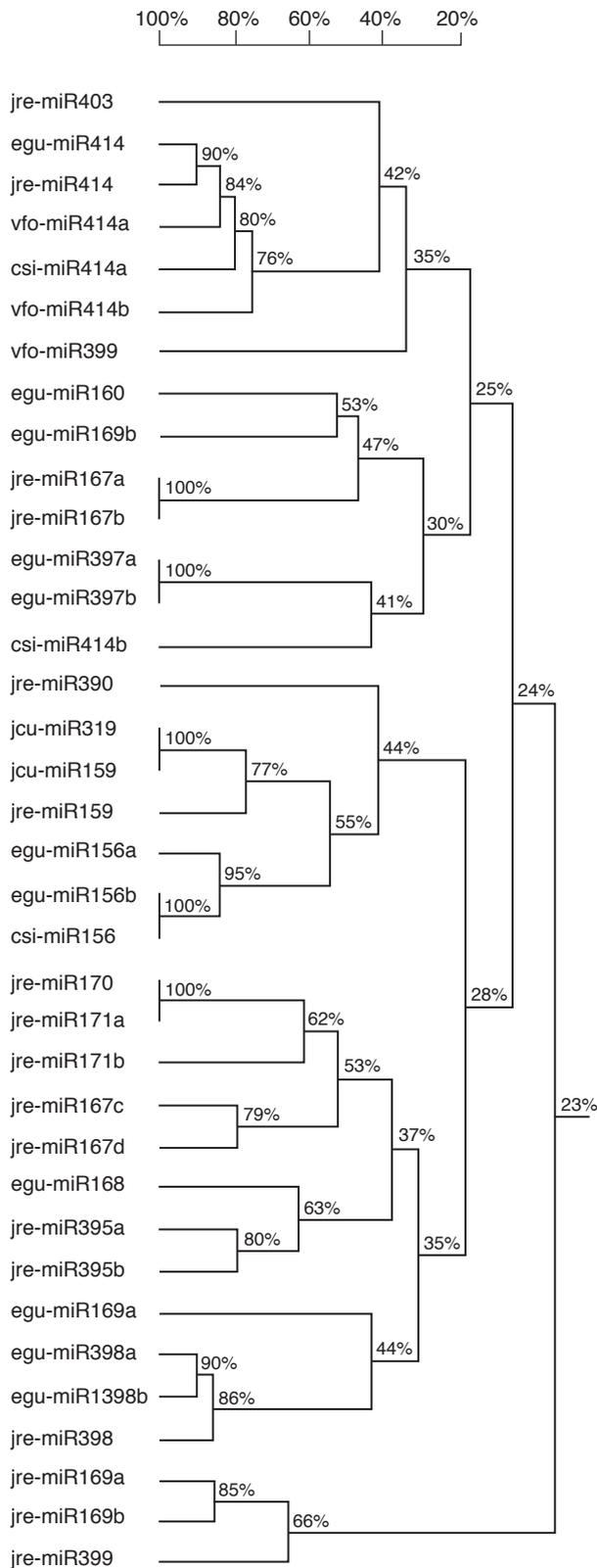
a miRNA family shared the same (or similar) gene sequences at certain positions within the miRNA sequence (Figure 3).

Target Prediction for Woody Oil Crop miRNA

Targets of the miRNA in plants were predicted by homology searches. The psRNATarget was employed to detect complementarity between the newly identified miRNA and predicted target mRNA in *E. guineensis*, *C. sinensis*, *J. regia*, *J. curcas*, *V. fordii*, *O. europaea*, *A. thaliana*, and *O. sativa* by homology searches (Zhang *et al.*, 2005; Dai *et al.*, 2009). Putative targets for these novel candidate miRNA were predicted based on the total number of mismatched nucleotides between the miRNA and the alignment structures of any potential targets, using the mRNA database for six woody oil crops available at NCBI; the DFCI *Arabidopsis* gene indices (<http://compbio.dfci.harvard.edu/tgi/>, Release 15.0), and the DFCI rice gene indices (<http://compbio.dfci.harvard.edu/tgi/>, Release 19.0) as a reference set. The total number of permitted mismatches at complementary sites between the miRNA sequence and any potential mRNA targets in woody oil crops, *Arabidopsis* and *Oryza* was limited to a maximum of three, and no gaps were allowed at complementary sites (Amiteye *et al.*, 2011). Since the protein-coding nucleotide (mRNA) database of woody oil crops has not yet been fully annotated, we applied an additional search (BLASTX) of NCBI proteins from all plant species. Eighty-eight miRNA:target pairs were predicted for 11 miRNA in the *E. guineensis*, giving an average ratio of miRNA to target of 8 (88/11), which is greater than the reported ratios of 6.59 for

J. regia and 7.33 for *V. fordii* but less than the 43.33 and 12 ratios reported for *C. sinensis* and *J. curcas*, respectively (Li and Mao, 2007). After removal of duplicated candidates, or ones that exceeded the selection criteria, 275 miRNA target genes for 27 woody oil crop miRNA were produced (data not shown). A summary for 97 of these targets for which annotation of a reasonable quality is available is provided in Table S1 which is available via <http://jopr.mpob.gov.my/wp-content/uploads/2018/03/FigureS1.pdf>, together with 40 targets that encode known proteins. The remaining miRNA targets encode unknown proteins (not shown). At least one miRNA target has been identified for each of the woody oil crop miRNA queried in the NCBI mRNA database of *Arabidopsis*, *Oryza* and all woody oil crop species, except for the following miRNA searches: Elagu-miR156a, Camsi-miR156, Jugre-miR167a, Jugre-miR167c, Jugre-miR167d, Elagu-miR169b, Jatcu-miR319, Elagu-miR397b, and Verfo-miR414a (Table 2). Of the 137 identified target genes, only five matches were detected in *J. curcas*, three matches in *J. regia*, three matches in *E. guineensis*, two matches in *J. regia*, one match in *V. fordii*, one match in *O. europaea*, three matches in *O. sativa* and the remainder were found in *Arabidopsis*. The miR414 has the highest number of predicted target genes (169), whilst miR403 has only one predicted target gene.

The target genes of miRNA are diverse in both sequence and function, with the majority of them not only associated with transcription factors, but also involved in several biological processes, such as plant development, metabolism and environmental stress response. However, the functions of some target genes remain largely unknown. It is noted



Note: An unrooted phylogenetic tree based on the nucleotide sequence identities of 36 newly identified miRNA in woody oil crops was generated using the Neighbour-Joining algorithm of the MEGA4.0 (Tamura *et al.*, 2007) program package, and bootstrap support values were calculated with 1000 iterations. Numbers on the nodes indicate nucleotide sequence homologies between clades.

Figure 2. Phylogenetic relationship of the 36 newly identified miRNA (microribonucleic acids).

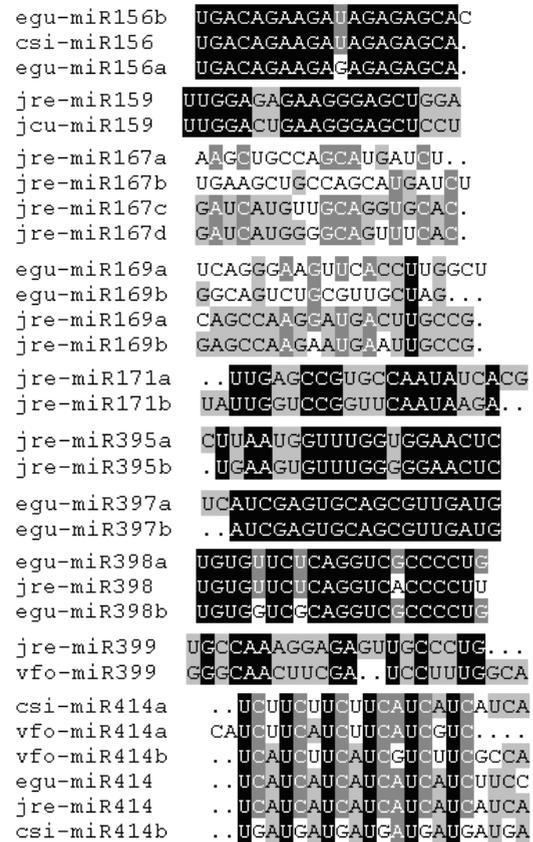


Figure 3. Sequence alignments of 25 newly identified woody oil crop mature miRNA (microribonucleic acids) that can be categorised into 10 miRNA families. Black boxes represent sequence identity, and dark grey boxes represent sequence similarity.

that in some cases a given miRNA may have multiple different mRNA targets, here the targets have been divided into several gene families (Table S1). For example, for the 73 detected targets of miR414, eight were found to be transposable element genes but the others can be grouped into different families. In other cases, a given gene might be targeted by multiple miRNA. An example of this can be seen for miR170 and miR171, which both target *SCARECROW-LIKE TRANSCRIPTION FACTOR 6 (SCL6)*. The *SCL6* transcription factor has been demonstrated to play a role in developmental patterning (Llave *et al.*, 2002). This study suggests that *SCL6* may also be a potential target of both jre-miR170 and jre-miR171, newly identified miRNA in *J. regia*. Another example can be seen for miR395 and miR414, which simultaneously target the 40S ribosomal protein gene family.

Egu-miR156 has been predicted to target *Squamosa* promoter-binding protein-like (SPL) genes (Table S1). It has also been reported that miR156 targets the SBP genes in other plant species (Dhandapani *et al.*, 2011). Targeting of the SBP transcription factor by miR156 is thought to regulate its two paralogous SPL genes SPL9 and SPL15 for shoot maturation and is important in leaf

development (Kantar *et al.*, 2010). In *Arabidopsis*, there are more than 10 SPL genes that are targeted by miR156 (Guo *et al.*, 2007). Since egu-miR156 targets SPL, this would suggest that Squamosa may also be regulated by miR156 in *E. guineensis*.

Some studies have indicated that miR159 is a highly conserved miRNA with complementarity to a highly conserved motif in the coding region of GAMYB-like genes, which have been found to play an important role in plant flowering under short days, anther development and seed germination via repression of GAMYB-Like genes (Zuo *et al.*, 2011). Palatnik *et al.* (2003) showed that overexpression of MYB33 resistant to miRNA resulted in leaves with upward curl in *Arabidopsis*. Similarly, Millar and Gubler (2005) demonstrated that miR159 regulated MYB33 expression by transforming *Arabidopsis* plants with the MYB33 miRNA target site mutated. When this motif was absent, aberrant MYB33 gene expression occurred, and the plants showed abnormal curl-shaped leaf and pleiotropic developmental defects. We predicted three MYB genes to be regulated by *J. curcas* miR159, but also found two new miR159 target genes (TCP) that are not related to MYB (Buxdorf *et al.*, 2010). TCP encode a novel protein that is dissimilar to the GAMYB family of transcription factors. These targets, which represent and as yet unknown and novel function for this miRNA, may also be regulated by miR159 in *J. curcas*.

Previous studies have demonstrated that several auxin response factors (ARF) are targeted by miRNA. ARF play important roles in plant seed germination, flower development, and signalling pathways. Wang *et al.* (2005) showed that ARF10 and ARF16, regulated by miR160 in *Arabidopsis*, are indispensable for root cap formation. It has been reported that ARF17, regulated by miR160, is important for *Arabidopsis* growth and development (Xing *et al.*, 2011). In this study, ARF10, ARF16 and ARF17 were predicted to be the potential targets of miR160, suggesting that ARF10, 16 and 17 may also be regulated by miR160 in *E. guineensis*.

It is known that several targets of the miRNA miR168 and miR403 are members of the argonaute gene family, which regulate their own biosynthetic pathway. The miR168 targets mRNA encoding argonaute-like protein (AGO), which is one of the important proteins in the regulation of miRNA biogenesis. In *Arabidopsis*, AGO1 participates in the regulation of the metabolic pathway of the miRNA precursor. The AGO2 mRNA is a target of miR403 (Allen *et al.*, 2005; Lobbes *et al.*, 2006), which is involved in miRNA biogenesis. We recently predicted a target gene which codes AGO proteins, and this gene is the target of miRNA 403 (Griffiths-Jones *et al.*, 2006).

The NF-YA transcription factor genes have been reported to be targets of miR169 and are involved

in plant tolerance to environmental stresses. AtNF-YA5 has been reported to be involved in drought resistance in *Arabidopsis* (Li *et al.*, 2008). In this report, NF-YA1, NF-YA3, NF-YA5, NF-YA8 and NF-YA9 were predicted to be targets of jre-miR169. Furthermore, *ABA Repressor1 (ABR1)* was also predicted to be a target of the same miRNA family, which would provide evidence that miR169 and its target genes in *J. regia* play a role in the responses of plants to drought stress. It is known that walnut has a high resistance to drought.

Whilst miR390 has been reported to target *TAS3* and produce tasiRNA that also regulate another subset of ARF, namely *ARF3* and *ARF4* in *Arabidopsis*, there is no evidence of miR390 conservation in *J. regia* since miR390 was found to likely target other genes.

The miR395 and miR399 are triggered to respond to nutrient deficiency (Fujii *et al.*, 2005; Chiou *et al.*, 2006). It has been shown that expression of miR395 is strongly induced under sulphate depletion conditions. The miR395 targets ATP-sulfurylases (APS) that are involved in inorganic sulphate assimilation (Jones-Rhoades and Bartel, 2004). Under phosphate starvation in *Arabidopsis*, miR399 targets the ubiquitin-conjugating E2 enzyme implicated in Pi acquisition from the soil (Chiou *et al.*, 2006; Jones-Rhoades and Bartel, 2004). Hence, miR399 is of great importance in the regulation of plant phosphate homeostasis. Our results may thus predict the existence of an APS and E2 function in *J. regia*.

Laccases are enzymes that are involved in lignin metabolism. Laccase has also been predicted to be the target of miR397 in *Arabidopsis* and rice (Sunkar and Zhu, 2004; Luo *et al.*, 2006). Egu-miR397 was predicted to target laccase genes, suggesting that *laccase2* could also be regulated by miR397 in *E. guineensis*. In addition, another gene, which encodes a serine-type endopeptidase, was predicted to be an egu-miR397 target. Therefore, egu-miR397 may perform additional functions in oil palm through the targeting of additional genes.

The miR398 family targets copper/zinc superoxide dismutase genes (CSD). In previous reports (Li *et al.*, 2010; Dugas and Bartel, 2008; Beauclair *et al.*, 2010), miR398 induction was seen to be inversely correlated with the expression of two copper superoxide dismutase genes in *Arabidopsis* and rice. Notably, down-regulation of miR398 under oxidative stress conditions was shown to up-regulate Cu/Zn-superoxide dismutases CSD1 and CSD2 in *Arabidopsis*. In contrast, the up-regulation of miR398 under copper deficiency conditions was implicated in negatively regulating expression of CSD1 and CSD2 genes. In this work, CSD was predicted to be the potential target of both egu-miR398a and jre-miR398. These results illustrate that CSD may

also be regulated by miR398 in *E. guineensis* and *J. regia*.

There were 169 potential targets of miR414 detected in *E. guineensis*, *C. sinensis*, *J. regia*, and *V. fordii* (Table S1). Amongst these targets, four are related to transcription factors. Others are mainly involved in signal transduction, responses to environmental conditions, metabolism and developmental processes. Nucleosomes are the fundamental repeating subunits of all eukaryotic chromatin. Hence, nucleosome assembly is a fundamental biological process, for which nucleosome assembly protein (NAP) is required. In this study, miR414, which was perfectly complementary to oil palm NAP-1 mRNA, was detected, and furthermore, the miRNA binding site was seen to be conserved in both monocotyledonous and dicotyledonous plant species, including *Arabidopsis*. This suggests that miR414 may have control over the regulation of nucleosome assembly in *E. guineensis*.

Characterisation of Putative miRNA by Semi-quantitative RT-PCR

To confirm the computationally identified miRNA we used semi-quantitative RT-PCR to analyse the expression levels of 11 Egu-miRNA in leaf samples from 13-year old oil palm, with *U6* as the control for normalisation. The semi-quantitative RT-PCR results showed that 11~60-bp fragments and a 76-bp fragment were amplified (Figure 4), which suggested that the designed primers had good specificity. Additionally, the results implied that the 11 mature ~18–21-bp miRNA were detectable in leaves. A densitometric

analysis of bands in Figure 4, in which the 11 Egu-miRNA expression levels were normalised against that of *U6* expressed in leaves showed that the expressions of miR156a, miR156b, miR160, miR168, miR397a, miR397b, miR398a and miR398b increased compared with the control, whereas the expressions of miR169a, miR169b and miR414 decreased in the tender leaves. The highest increased rate of expression level was over two-fold for miR160, and the changes in the regulatory rates of the other 10 miRNA were within two-fold.

To investigate the effects of miRNA identified in leaves on oil palm fruit development, the miRNA expressions in fruitlet tissues were assayed by stem-loop semi-quantitative RT-PCR. A densitometric analysis of the bands in Figure 5, in which the 11 Egu-miRNA expression levels were normalised against that of *U6* expressed in fruits indicated that miR168 had the highest expression level in oil palm fruit, followed by miR398a. In addition, miR168 and miR398a exhibited increased expression levels of more than two-fold, and miR397a and miR398b exhibited increased expressions of 1.5- to two-fold, respectively, in fruit tissue as compared with the control *U6*. However, the expressions of miR156b, miR160, miR169a, miR169b, miR397b and miR414 decreased.

Validation of miRNA Expression Using Quantitative RT-PCR

To validate the miRNA tested by semi-quantitative PCR analysis in tender leaves, we examined the expression profiles of 11 conserved miRNA in leaf tissues using stem-loop real time quantitative RT-PCR. As shown in Figure 6, among



Figure 4. The expression profiles of mature Egu-miRNA in leaves from oil palm. Agarose gel electrophoresis detection of the reverse transcription-polymerase chain reaction (RT-PCR) products of different Egu-miRNA in leaves. CK: negative control; 1: miR156a; 2: miR156b; 3: miR160; 4: miR168; 5: miR169a; 6: miR169b; 7: miR397a; 8: miR397b; 9: miR398a; 10: miR398b; 11: miR414; 12: *U6*; M: 100-bp ladder marker.



Figure 5. The expression profiles of mature Egu-miRNA in fruits from oil palm. Agarose gel electrophoresis detection of the reverse transcription-polymerase chain reaction (RT-PCR) products of different Egu-miRNA in fruits. M: 100-bp ladder marker; 1: miR414; 2: miR398b; 3: miR398a; 4: miR397b; 5: miR397a; 6: miR169b; 7: miR169a; 8: miR168; 9: miR160; 10: miR156b; 11: miR156a; 12: *U6*; CK: negative control.

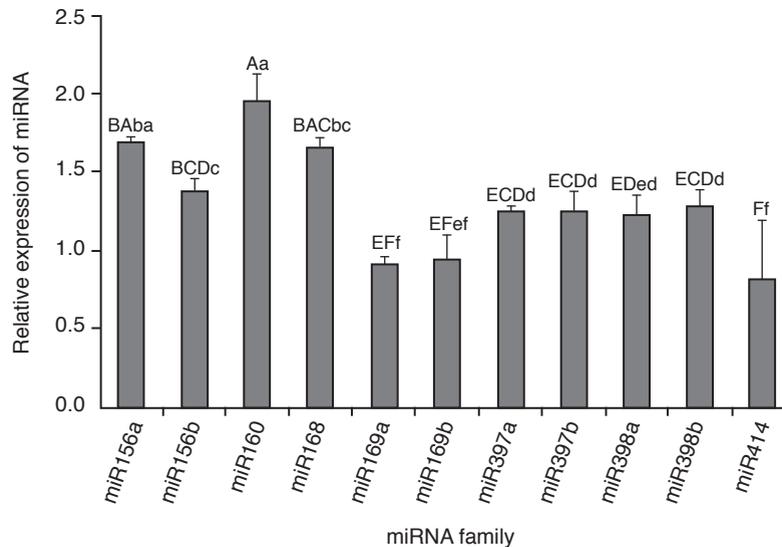


Figure 6. Relative expression profiles of 11 newly identified miRNA (microribonucleic acids) from oil palm leaf tissues. The average (\pm SE) values represent the expression levels of 11 Egu-miRNA normalised to a reference gene from three biological replicates of the same tested tissues.

the 11 Egu-miRNA detected, miR160 displayed the highest expression level, followed by miR156a and miR168. The expression levels of miR160 and miR156a were 1.50- and 1.30-fold the mean value of the 11 miRNA, respectively. On the contrary, miR414, miR169a and miR169b exhibited relatively low expression levels compared with the other miRNA detected. For instance, the expression level of miR414 was less than 0.62-fold the mean value of the 11 miRNA.

A comparison between the semi-quantitative RT-PCR and real-time quantitative RT-PCR showed a significant correlation ($R = 0.861$, $P < 0.001$) over three biological replicates (Figures 4 and 5). This suggested that the results were in agreement with each other and; therefore, the expression results by both approaches were reliable. However, semi-quantitative RT-PCR provided not only detection of miRNA but also its sequence information through subcloning and sequencing, and the detection of miRNA by semi-quantitative RT-PCR was possible even at expression levels near the detection limits of real time qRT-PCR.

Differentially Expressed miRNA between Leaf and Fruit Tissues

The miRNA expression profiles of two oil palm tissues were determined by semi-quantitative RT-PCR analysis. The results showed that eight Egu-miRNA were significantly differentially expressed, with two being up-regulated (fold change > 2) and six being down-regulated (fold change < 0.5) in fruit as compared with their levels in leaf. Specifically, the expressions of miR168 and miR398a increased, and the expressions of miR156b, miR160, miR169a, miR169b, miR397b and miR414 decreased in fruit

tissues relative to leaf tissues. Thus, the top two significantly up-regulated miRNA might play important roles in fruit. There were greater effects of sample type on the identification of miRNA, with more identified in leaf tissue samples.

CONCLUSION

Various plant species miRNA have been identified in NCBI EST and GSS sequence databases using *in silico* computer-based homology searches. Thirty-six new woody oil crop mature and pre-miRNA were analysed at the sequence and structural levels. Their potential targets were also predicted computationally and 275 target mRNA from other taxa were detected. To verify the existence of the computationally identified miRNA, we measured all of the newly detected oil palm miRNA in leaf tissue employing both semi-quantitative RT-PCR and real-time quantitative RT-PCR. Additionally, these novel miRNA transcripts were measured in fruit tissue for further validation by semi-quantitative RT-PCR. A correlation analysis of the data of both the two PCR assays and miRNA expression levels in the tissues indicated that the semi-quantitative RT-PCR assays can be used to detect the expression of miRNA in different organ tissues of oil palm. Markedly different expression levels of miRNA transcripts were detected between the leaf and the fruit tissues of *E. guineensis*.

ACKNOWLEDGEMENT

We thank Guangxia Yang of the Beijing Genomics Institute (BGI), P R China, for the functional

annotations. This work was supported by the National Natural Science Foundation of China (Project No. 3/270713).

REFERENCES

- ADAI, A; JOHNSON, C; MLOTSHWA, S; ARCHER-EVANS, S; MANOCHA, V; VANCE, V and SUNDARESAN, V (2005). Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res.*, 15: 7891.
- ALLEN, E; XIE, Z; GUSTAFSON, A M and CARRINGTON, J C (2005). microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell*, 121: 207-221.
- AMITEYE, S; CORRAL, J M; VOGEL, H and SHARBEL, T F (2011). Analysis of conserved microRNAs in floral tissues of sexual and apomictic *Boechera* species. *BMC Genomics*, 12: 500.
- BARTEL, D P (2004). microRNAs: genomics, biogenesis, mechanism and function. *Cell*, 116: 281-297.
- BEAUCLAIR, L; YU, A and BOUCHÉ, N (2010). microRNA-directed cleavage and translational repression of the copper chaperone for superoxide dismutase mRNA in *Arabidopsis*. *The Plant J.*, 62: 454-462.
- BUXDORF, K; HENDELMAN, A; STAV, R; LAPIDOT, M; ORI, N and ARAZI, T (2010). Identification and characterization of a novel miR159 target not related to MYB in tomato. *Planta*, 232: 1009-1022.
- CAMACHO, C; COULOURIS, G; AVAGYAN, V; MA, N; PAPADOPOULOS, J; BEALER, K and MADDEN, T L (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421.
- CHEN, C; RIDZON, D A; BROOMER, A J; ZHOU, Z; LEE, D H; NGUYEN, J T; BARBISIN, M; XU, N L; MAHUVAKAR, V R; ANDERSEN, M R; LAO, K Q; LIVAK, K J and GUEGLER, K J (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33: e179-e180.
- CHIOU, T J; AUNG, K; LIN, S I; WU, C C; CHIANG, S F and SU, C (2006). Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. *Plant Cell*, 18: 412-421.
- DAI, Y; SUI, W; LAN, H; YAN, Q; HUANG, H and HUANG, Y (2009). Comprehensive analysis of microRNA expression patterns in renal biopsies of lupus nephritis patients. *Rheumatol Int.*, 29: 749-754.
- DHANDAPANI, V; RAMCHIARY, N; PAUL, P; KIM, J; CHOI, S H; LEE, J; HUR, Y and LIM, Y P (2011). Identification of potential microRNAs and their targets in *Brassica rapa* L. *Mol. Cells*, 32: 21-37.
- DUGAS, D V and BARTEL, B (2008). Sucrose induction of *Arabidopsis* miR398 represses two Cu/Zn superoxide dismutases. *Plant Mol. Biol.*, 67: 403-417.
- FAHLGREN, N; HOWELL, M D; KASSCHAU, K D; CHAPMAN, E J; SULLIVAN, C M; CUMBIE, J S; GIVAN, S A; LAW, T F; GRANT, S R; DANGL, J L and CARRINGTON, J C (2007). High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS ONE*, 2: e219.
- FILIPOWICZ, W; BHATTACHARYA, S N and SONENBERG, N (2008). Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nat. Rev. Genet.*, 9: 102-114.
- FRAZIER, T P; XIE, F; FREISTAEDTER, A; BURKLEW, C E and ZHANG, B (2010). Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). *Planta*, 232: 1289-1308.
- FUJII, H; CHIOU, T J; LIN, S I; AUNG, K and ZHU, J K (2005). A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr. Biol.*, 15: 2038-2043.
- GRIFFITHS-JONES, S (2006). miRBase: The microRNA sequence database. *Methods Mol. Biol.*, 342: 129-138.
- GRIFFITHS-JONES, S; GROCOCK, R J; VAN DONGEN, S; BATEMAN, A and ENRIGHT, A J (2006). miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, 34: D140-D144.
- GRIFFITHS-JONES, S; SAINI, H K; VAN, D S and ENRIGHT, A J (2008). miRBase: Tools for microRNA genomics. *Nucleic Acids Res.*, 36: D154-D158.
- GUO, Q; XIANG, A; YANG, Q and YANG, Z (2007). Bioinformatic identification of microRNAs and their target genes from *Solanum tuberosum* expressed sequence tags. *Chinese Science Bulletin*, 52(17): 2380-2389.
- HUANG, Y; NIU, A; GAO, Y; FU, L and LI, W (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, 26: 680-682.

- JONES-RHOADES, M W and BARTEL, D P (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, 14(6): 787-799.
- KANTAR, M; UNVER, T and BUDAK, H (2010). Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. *Funct. Integr. Genomics*, 10: 493-507.
- KIM, K I and VAN DE WIEL, M A (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics*, 9(1): 114.
- LEUNG, A K L and SHARP, P A (2010). MicroRNA functions in stress responses. *Mol. Cell*, 40: 205-215.
- LI, A L and MAO, L (2007). Evolution of plant microRNA gene families. *Cell Res.*, 17: 212-218.
- LI, W X; OONO, Y; ZHU, J; HE, X J, WU, J M; IIDA, K; LU, X Y; CUI, X; JIN, H and ZHU, J K (2008). The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and post-transcriptionally to promote drought resistance. *Plant Cell*, 20(8): 2238-2251.
- LI, Y F; ZHENG, Y; ADDO-QUAYE, C; ZHANG, L; SAINI, A; JAGADEESWARAN, G; AXTELL, M J; ZHANG, W X and SUNKAR, R (2010). Transcriptome-wide identification of microRNA targets in rice. *Plant J.*, 62: 742-759.
- LIANG, X F; HUANG, S Q; GUO, K; XIANG, A L; ZHU, Y Y; NIE, L and YANG, Z M (2007). Computational identification of novel microRNAs and targets in *Brassica napus*. *FEBS Lett.*, 581: 1464-1474.
- LLAVE, C; XIE, Z; KASSCHAU, K and CARRINGTON, J (2002). Cleavage of scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science*, 297: 2053-2056.
- LOBBES, D; RALLAPALLI, G; SCHMIDT, D D; MARTIN, C and CLARKE, J (2006). SERRATE: A new player on the plant microRNA scene. *EMBO Reports*, 7: 1052-1058.
- LOW, E T; ROSLI, R; JAYANTHI, N; MOHD-AMIN, A H; AZIZI, N; CHAN, K L; MAQBOOL, N J; MACLEAN, P; BRAUNING, R; MCCULLOCH, A; MORAGA, R; ONG-ABDULLAH, M and SINGH, R (2014). Analyses of hypomethylated oil palm gene space. *PLoS ONE*, 9(1): e86728.
- LU, C; TEJ, S S; LUO, S; HAUDENSCHILD, C D; MEYERS, B C and GREEN, P J (2005a). Elucidation of the small RNA component of the transcriptome. *Science*, 309: 1567-1569.
- LU, S; SUN, Y H; SHI, R; CLARK, C; LI, L and CHIANG, V L (2005b). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell*, 17: 2186-2203.
- LUO, Y C; ZHOU, H; LI Y; CHEN, J Y; YANG, J H, CHEN, Y Q and QU, L H (2006). Rice embryogenic calli express a unique set of microRNAs, suggesting regulatory roles of microRNAs in plant post-embryogenic development. *Febs Letters*, 580: 5111-5116.
- MATHEWS, D H; DISNEY, M D; CHILDS, J L; SCHROEDER, S J; ZUKER, M and TURNER, D H (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. of Sci. USA*, 101: 7287-7292.
- MATHEWS, D H; SABINA, J; ZUKER, M and TURNER, D H (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288: 911-940.
- MEYERS, B C; AXTELL, M J; BARTEL, B; BARTEL, D P; BAULCOMBE, D; BOWMAN, J L; CAO, X; CARRINGTON, J C; CHEN, X; GREEN, P J; GRIFFITHS-JONES, S; JACOBSEN, S E; MALLORY, A C; MARTIENSSSEN, R A; POETHIG, R S; QI, Y; VAUCHERET, H; VOINNET, O; WATANABE, Y; WEIGEL, D and ZHU, J K (2008). Criteria for annotation of plant microRNAs. *Plant Cell*, 20: 3186-3190.
- MILLAR, A A and GUBLER, F (2005). The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell*, 17(3): 705-21.
- NASARUDDIN, N M; HARIKRISHNA, K; OTHMAN, R Y; HOON, L S and HARIKRISHNA, J A (2007). Computational prediction of microRNAs from oil palm (*Elaeis guineensis* Jacq.) expressed sequence tags. *Asia Pac. J. Mol. Biol. Biotechnol.*, 15 (3): 107-113.
- PALATNIK, J F; ALLEN, E; WU, X; SCHOMMER, C; SCHWAB, R; CARRINGTON, J C and WEIGEL, D (2003). Control of leaf morphogenesis by microRNAs. *Nature*, 425(6955): 257-263.
- STELLWAG, E J (2008). Identification of soybean microRNAs and their targets. *Planta*, 229: 161-182.
- SUNKAR, R and ZHU, J K (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell*, 16: 2001-2019.

- TAMURA, K; DUDLEY, J; NEI, M and KUMAR, S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24(8): 1596-1599.
- WANG, J W; WANG, L J; MAO, Y B; CAI, W J; XUE, H W and CHEN, X Y (2005). Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. *The Plant Cell*, 17: 2204-2216.
- WEI, B; CAI, T; ZHANG, R; LI, A; HUO, N; LI, S; GU, Q Y; VOGEL, J; JIA, J; QI, Y and MAO, L (2009). Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.) Beauv. *Funct. Integr. Genomics*, 9: 499-511.
- WILLMANN, M R and POETHIG, R S (2007). Conservation and evolution of miRNA regulatory programs in plant development. *Curr. Opin. Plant Biol.*, 10: 503-511.
- XING, H; PUDAKE, R N; GUO, G; XING, G; HU, Z; ZHANG, Y; SUN, Q and NI, Z (2011). Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. *BMC Genomics*, 12: 178.
- YOSHIKAWA M; PERAGINE A; PARK M Y and POETHIG R S A (2005). Pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. *Genes Dev.*, 19: 2164-2175.
- ZHANG, B H; PAN, X P; WANG, Q L; COBB, G P and ANDERSON, T A (2005). Identification and characterization of new plant microRNAs using EST analysis. *Cell Res.*, 15: 336-360.
- ZHANG, B H; PAN, X P; COX, S B; COBB, G P and ANDERSON, T A (2006a). Evidence that miRNAs are different from other RNAs. *Cell Mol. Life Sci.*, 63: 246-254.
- ZHANG, B H; PAN, X; COBB, G P and ANDERSON, T A (2006b). Plant microRNA: A small regulatory molecule with big impact. *Dev. Biol.*, 289: 3-6.
- ZHANG, B; PAN, X; CANNON, C H; COBB, G P and ANDERSON, T A (2006c). Conservation and divergence of plant microRNA genes. *The Plant J.*, 46: 243-259.
- ZHANG, Y (2005). miRU: An automated plant miRNA target prediction server. *Nucleic Acids Res.*, 33: W701-W704.
- ZHAO, T; LI, G; MI, S; LI, S; HANNON, G J; WANG, X J. and QI, Y A (2007). Complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.*, 21: 1190-1203.
- ZUKER, M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31: 3406-3415.
- ZUO J H; WANG Y X; LIU H P; MA Y Z; JU Z; ZHAI B Q; FU D Q; ZHU Y; LUO, Y B and ZHU, B Z (2011). MicroRNAs in tomato plants. *Sci. China Life Sci.*, 54(7): 599-605.