# IDENTIFICATION OF CONSERVED MICRORNAS AND TRANSCRIPT TARGETS IN
## *Vernicia fordii* SEEDS BY HIGH-THROUGHPUT SEQUENCING

Vanessa Galli[1,2], Rafael da Silva Messias[2], Sérgio Delmar dos Anjos e Silva[2], Rogério Margis[1]

## INTRODUCTION

Crop plants offer substantial potential to provide environmental friendly, renewable and sustainable chemical feedstock that could replace petroleum-based products in industrial applications. The tung tree (*Vernicia fordii*) belongs to the Euphorbiaceae family and presents multipurpose and highly valuable oil. Tung oil shows a very high content (82%) of the conjugated α-eleostearic acid ($C18:3^{\Delta 9c,11t,13t}$), an unusual fatty acid that is widely used in inks, dyes, coatings, and a variety of other bio-based products. Moreover, the tung oil is now being considered for use in biodiesel production (VANHERCKE et al., 2013).

However, significantly more knowledge of genes and transcriptional regulation involved in the seed oil biosynthesis and other agronomic traits are needed in order to breed tung tree for higher fruit yield and for modified oil properties to be used as biodiesel. MicroRNAs (miRNAs) are small non-coding regulatory RNAs which act in post-transcriptional gene regulation (FILIPOWICZ et al., 2008). Genes encoding miRNAs are typically transcribed as long polyadenylated transcripts, which show an imperfect stem-loop structure that is recognized and processed by DICER-Like1 (DCL1) into a pre-miRNA, which is further processed to generate the mature miRNAs with 21 nucleotides (nt) in length. The mature miRNA bound by an argonaute protein recognize the mRNA target by perfect or nearly perfect match to the coding sequence. Then, the posttranscriptional silencing occurs through cleavage or translation inhibition of the mRNA targets. This regulation plays critical roles in a wide variety of biological processes (VOINNET, 2009).

Identification of miRNAs and their target genes is essential for a full understanding of gene expression in plants. Therefore, the aim of this study was to perform a deep sequencing of small

[1]Centro de Biotecnologia, PPGBCM, Laboratório de Genomas e População de Plantas, prédio 43431, Universidade Federal do Rio Grande do Sul - UFRGS, P.O. Box 15005, CEP 91501-970, Porto Alegre, RS, Brazil.
[2]Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, P.O. Box 403, CEP 96010-971, Pelotas, RS, Brazil.

RNAs (sRNA) from mature seeds and combine with bioinformatic analyses to identify conserved miRNAs in tung seeds.

## MATERIAL AND METHODS

For RNA isolation, fruits from *V. fordii* plants were collected at mature stage (approximately 100 days after flower opening). Seeds were dissected and immediately frozen in liquid nitrogen and stored at -80 °C. Total RNA was isolated using Trizol reagent (Invitrogen), according to the manufactures' protocol. The RNA quality was evaluated by electrophoresis on a 1% agarose gel and the RNA concentration was determined by Nanodrop (Nanodrop Technologies). The total RNA (> 10 μg) was sent to Fasteris SA for processing and sequencing a small RNA (sRNA) library and a mRNA-seq library by Illumina HiSeq2000, as previously described (GALLI et al., 2013).

Sequences from the sRNA library were cleaned up from low quality reads, sequences shorter than 18 nt and longer than 25 nt, and from non-coding RNAs (GALLI et al., 2013). mRNA-seq data was *de novo* assembled into contigs using CLC Genome Workbench version 4.0.2 with default parameters. To identify phylogenetically conserved miRNAs, sRNA sequences were mapped to a set of all conserved non-redundant Viridiplantae obtained from miRBase database (Release 19, August 2012) using Bowtie v 0.12.7 (LANGMEAD et al., 2009). The frequency of identified miRNAs was obtained by aligning the conserved precursors identified in this study and the sRNA library using Bowtie v 0.12.7 with the default parameters. The prediction of target genes of conserved miRNAs was performed with psRNAtarget (DAI and ZHAO, 2011) by aligning sRNA sequences against *V. fordii* assembled unigenes. Annotation of predicted targets was performed using BLASTx against the NR and the Swiss-Prot/Uniprot protein databases.

## RESULTS AND DISCUSSION

The sRNA sequence resulted in a total of 25,452,206 reads. After removing low quality sequences and non-coding RNAs, 13,387,276 reads remained. Interestingly, most of the remained reads shows 24 nt length (64%), followed by 21 nt length (13%) (Figure1), the typical size range for Dicer-like (DCL) derived products.
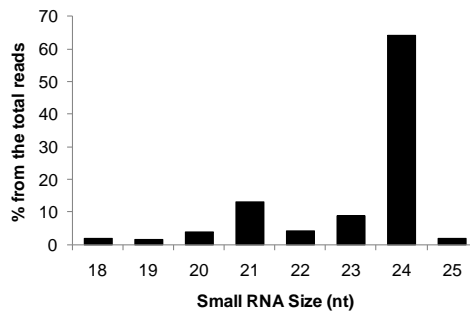
Figure 1. Characterization of sRNA library from *V. fordii* mature seeds: reads distribution according to their size.

Based on sequence homology, we identified 839,846 reads represented by 138 conserved plant miRNAs, which were distributed in 34 conserved miRNAs families. Overall, MIR166 family was the most abundant conserved miRNA family present in *V. fordii* seeds, followed by MIR156. These abundant miRNA families are largely found in Viridiplantae, indicating a fundamental role in plant life maintenance. These were also the largest families, with 20 and 14 members, respectively. The expression varied significantly among different miRNA families (from 1 to 759,212 reads) (Figure 2).
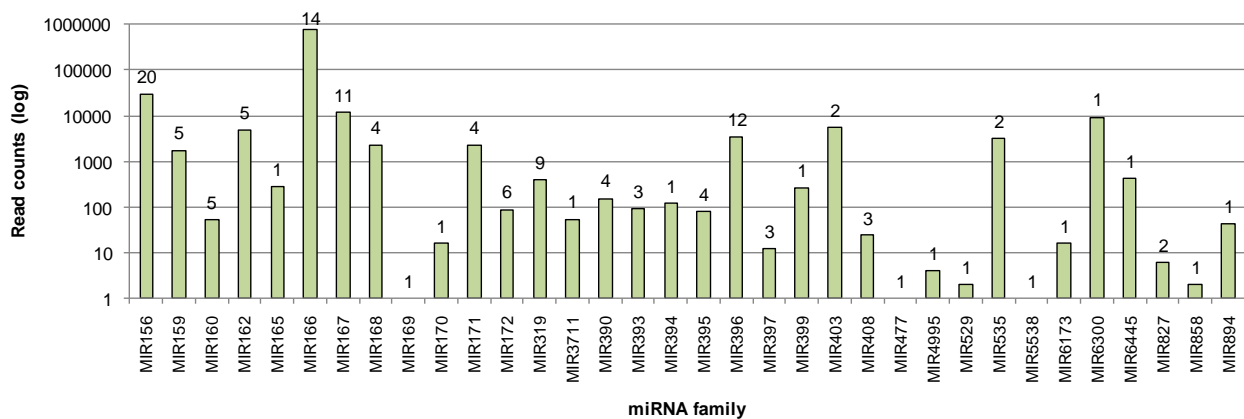


Figure 2. Number of identified *V. fordii* miRNAs in each conserved miRNA family. The values above the bars indicate the number of members identified for each conserved miRNA family.

A total of 358 sequences were predicted as potential targets of 79 plant-known miRNAs distributed in 13 miRNA families. Among the conserved miRNA targets, we found transcription

factors: squamosa promoter binding protein (SBP)-like (SPL) genes (target of MIR156 and MIR535), MYB33, MYB101 and GAMYB (target of MIR159), among others; and genes involved in the silencing machinery: Dicer-like-1and Argonaute 2 (target of MIR162 and MIR403, respectively). Moreover, several candidate miRNA targets of *V. fordii* are involved with response to biotic or abiotic stimulus.

Finally, we have identified conserved miRNA in tung seed with the following predicted targets involved in the lipid metabolism: glycerol-3-phosphate acyltransferase (target of MIR156_5p), which incorporate activated fatty acids to the glycerol backbone to form triacylglycerol; enoyl reductase (target ofMIR156_5p), responsible for the elongation of malonyl-CoA to long-chain acyl groups; long chain acyl-synthetase (target of MIR168_3p), involved in the activation and transport of long-chain acyl groups to endoplasmic reticulum, and diacylglycerol acyltransferase (target of MIR403a_3p), which catalyzes the final step of lipid synthesis (BEISSON et al., 2013).

## CONCLUSIONS

The present study reported the identification of conserved miRNAs and their target genes in *V. fordii* mature seeds through deep sequencing and bioinformatics. This study contributes for elucidation of miRNA mediated regulatory systems related to the control seed development and other physiological processes.

## REFERENCES

BEISSON YL, SHORROSH B, BEISSON F, ANDERSSON MX, ARONDEL V, et al. **Acyl-Lipid Metabolism**. The Arabidopsis Book v.11, e0161, 2013.
DAI X, ZHAO PX. psRNATarget: a plant small RNA target analysis server. **Nucleic Acids Res**, v. 39, p. 155 – 159, 2011.
GALLI V, GUZMAN F, OLIVEIRA LFV, LOSS-MORAIS G, et al. Identifying MicroRNAs and Transcript Targets in Jatropha Seeds. **PLosOne**, v. 9, n. 2, e83727, 2013.
LANGMEAD B, TRAPNELL C, POP M, SALZBERG SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biol**, v. 10:R25, 2009.
VANHERCKE T, CRAIG CW, STYMNE S, SINGH SP, GREEN AG. Metabolic engineering of plant oils and waxes for use as industrial feedstocks. **Plant Biotech J**, v. 11, p. 197 – 210, 2013.
VOINNET O (2009) Origin, biogenesis and activity of plant microRNAs. Cell 136:669-687.
ZAMPESE E, FASOLATO C, KIPANYULA MJ, BORTOLOZZI M, POZZAN T, PIZZO P. Presenilin 2 modulates endoplasmic reticulum (ER)-mitochondria interactions and Ca2+ cross-talk. **PNAS**, v.15, p.2777 – 2782, 2013.