miRNA Identification from *Jatropha curcas* L. whole genome shotgun assembly sequences

**CURRENT STATUS:** UNDER REVIEW

**ExRNA**  
**BMC**

Nutan Prakash Vishwakarma  
pranutan@gmail.com **Corresponding Author**

**DOI:**  
10.21203/rs.3.rs-22674/v1

**SUBJECT AREAS**  
Epigenetics & Genomics

**KEYWORDS**  
*Jatropha curcas* L., miRNA Identification, target genes
Abstract
miRNAs are 20–25 nucleotide in length, endogenous in origin, and noncoding RNA molecules found in all the eukaryotic organism and takes part in the regulation of gene expression. *Jatropha curcas* L. oil has drawn attention as one of the most potential biofuels that can be a substitute of never-ending fuel crisis. Six miRNA candidates, each belonging to different miRNA families were predicted after screening of 41,790 whole genome shotgun assembly sequences by using the comparative genomics approach. The targets of identified miRNA were also predicted, the predicted target genes are belonging to diverse functions like in metabolism, transport, and genes involved in fatty acid metabolism. The predicted miRNAs show their targets which involve in the fatty acid metabolism, like miR414 targets the gene of acetyl-CoA carboxylase and O-acyltransferase WSD1, miR846 targets gene of FAD omega-3 fatty acid, FAD plastid fatty acid desaturase, miR5658 have the target gene long chain acyl-CoA synthetase 1, and miR407 have the target gene Trigalactosyldiacylglycerol 2, miR2938 GDSL esterase/lipase, glycolipid transfer protein 1.

Introduction
For a long time, researchers have been working on unravelling the role of genes which are responsible for the expression of proteins. Researchers were curious to know the reason that why the non-coding sequences are conserved through evolutionary selection. New light was shed on this problem in 1993 when the first miRNA was identified from the junk DNA of the nematode *C. elegans* named as lin-4 (1). The first plant miRNA was discovered in the year 2002 from the *Arabodopsis thaliana*, after that total of 4014 miRNAs from 52 different plant species have been identified and deposited miRBASE database which is available in public domain and freely accessible (2). miRNAs are 20–25 nucleotide in length, endogenous in origin, small, and noncoding RNA molecules found in all the eukaryotic organism. miRNA takes part in the regulation of gene expression at the post-transcriptional level by leading to the target degradation or translational repression with the sequence specific interactions to complementary sites of target mRNA (3). The transcription of MIR genes takes place similar to mRNA coding genes by the enzyme RNA polymerase II (4) (5). Once MIR genes are transcribed, they produce long pri-miRNA transcripts with
characteristic features similar to mRNAs. The stabilization of pri-miRNA takes place through the addition of a 5’ 7-methylguanosine cap, and a 3’ poly A tail (6). Pri-miRNA transcripts fold back due to their internal sequence complementarity into an imperfect stem-loop structure and are recognized by the members of a ribonuclease enzyme Dicer-like1 (DCL) family which belongs to RNAse III endoribonuclease protein family. Inside the nucleus the pri-miRNA sliced in two steps by the DCL1 complex. In the first step the base of the stem is trimmed and remaining structure is called as pre-miRNA. In the second step the loop of pre-miRNA is removed and the duplex of miRNA:miRNA* is left for further processing (2). The enzyme methytransferase HEN1, 2′-O-methylates the 3’ terminus of miRNA:miRNA* duplexes (7). 2′-O-methylation of miRNA-3p/5p duplex is essential to protect the 3’ termini of unwound mature miRNA from the action of exonucleases, such as SDN proteins (8).

Previously it was assumed that after the maturation of miRNA:miRNA*, only miRNA was loaded to RISC and proceeds for binding with the target mRNA. But now it is reported that miRNA* also may load on the RISC and can regulate the target genes. So considering this point now the miRNA : miRNA* duplex is named as miRNA-5p and miRNA-3p (9). The miRNA strand of the miRNA 5p/3p is loaded into the RNA inducing silencing complex (RISC) led by a protein AGO thereby directing miRNA towards its complementary target mRNA sequence (10). In the RNA induced silencing complex (RISC), miRNAs bind to target mRNA, based on the perfect or near-perfect complementarity between the miRNA and the mRNA either transcript cleavage or translation repression takes place.

The role of plant miRNA have been established in the regulation of several developmental processes such as leaf morphogenesis and polarity, root development, vascular development, phase transition from somatic to reproductive, flower and seed development (11). The role of miRNAs are also well established in diverse responses to stresses such as abiotic stresses like drought, salt, cold, oxidative, nutrient deficiency and biotic stresses (12). Most of the protein coding genes which are targetted by miRNAs, coding for transcription factor, which act in key biological processes like responding to abiotic and biotic stresses, performing metabolic functions, and regulating developmental processes.

*Jatropha curcas* L. is a monoecious and perennial shrub which is extensively grown throughout the tropical countries. It grows in any unfavorable agro-climatic conditions due to its low moisture
demand, low fertility requirement, tolerance to excessive temperature, pest and disease resistance. Due to ever inflating fuel prices and increasing concern over the greenhouse gas emissions such as CO₂, has led researchers to search for potential renewable biofuels. *Jatropha curcas L.* oil has drawn attention as one of the most potential biofuels that can be a substitute of never-ending fuel crisis. The crude seed oil of *Jatropha curcas* have the potential to be converted into biodiesel which can qualify the European and United State standards (13). A major constraint to high quality oil yield from *Jatropha curcas* is the lack of knowledge about its genetic regulation. To elucidate the role of miRNAs in the *Jatropha curcas* many studies has been conducted for the miRNA’s identification in *Jatropha curcas*. The first of its kind for the miRNA identification study was reported by using small RNA cloning methodology but researchers did not provide the precursor sequences (14). In the second study, the conserved miRNAs of *Jatropha curcas* were predicted by using the available Expressed sequence tags and Genome survey sequences in the publicly available databases by using comparative genomics approach (15). In the third study, conserved and novel miRNAs from *J. curcas* were identified through the deep sequencing of small RNAs (sRNA) from mature seeds (16). In the latest study, conserved and novel miRNAs from *J. curcas* were identified through the deep sequencing of small RNAs (sRNA) from the leaves tissue of *Jatropha curcas* (17). In a strategy to predict the conserved miRNAs from the different plant species by exploiting the whole genome shot gun sequences many reports have been published. At present, the miRNA prediction from the available whole genome shotgun assembly sequences of *J. curcas* is not reported from any researcher. This is an in-silico approach for the screening of potential conserved miRNAs by using homolog search of model plant *Arabidopsis thaliana* miRNAs against the genome of *Jatropha curcas L.* To identify the role of *J. curcas* miRNAs, the target genes of predicted miRNA is also identified.

**Results**

**Identification of conserved miRNAs**

Six miRNA candidates belonging to six different miRNA families were predicted after screening of 41,790 whole genome shotgun sequences (WGS). The homology search was performed to identify input sequences which contains mature miRNA sequences. In this step, the 318 mature miRNAs of
plant *Arabidopsis thaliana* from miRBase was used as source of mature miRNAs for the identification process. The identification methods include sequence scans with BLAST and the number of allowed mismatches as equal to or less than four nt between a source mature miRNA and its homolog in an input sequence of the genome sequence. A total of 2798 homologous sequences were detected and were selected as input for the primary miRNA folding. A miRNA gene is non-coding in nature and by folding it makes stem-loop precursor in its secondary structure. The pri-miRNA folding module of C-mii removes mRNA, tRNA, and rRNA sequences from input sequences, and predicts the secondary structure of pri-miRNAs. The sequences do not fold were excluded and remaining 2388 sequences were selected for the pre-mRNA prediction. The pre-miRNA folding module of C-mii, after removal of stem-loop sequences which match with mRNA and other types of non-coding RNAs, re-predicts the secondary structure of the extracted sequences.

All the predicted secondary hairpin structure of pre-miRNA shows highly negative MFE values which shows their stability. From the value of MFE the MFEI value was calculated which is usually consider to distinguish miRNAs from other coding and non-coding RNAs (26). 11 number of distinct sequences were successfully folded, but from the folded sequences only highly negative MFEI (≥ -0.85) having transcripts were considered for the study. The range variation of MFEI was observed to be in the range of -0.93 to -1.11 kcal/mol (Table 1). A total of six potential conserved miRNAs predicted from whole genome shotgun sequences which share a high degree of sequence identity with known miRNAs and fulfil the other criteria for miRNAs are listed in Table 1 and their predicted secondary structures are shown in Fig. 2. Jatropha pre-miRNAs sequences showed great variability in their length from 38 to 88nt (Table 1). jcu-mir2938 exhibited the shortest precursor length of 38nt, whereas jcu-mir407 exhibited the longest precursor length of 88nt. After inspecting the homology and secondary structure prediction results, total six miRNAs were predicted which belong to six different miRNA family. All the predicted mature miRNA sequences were of 21nt in length and are in the stem portions of the hairpin structures, as shown in Fig. 2.

Potential target genes for newly predicted conserved miRNAs miRNAs regulates expression of specific gene by binding to mRNA transcripts to inhibit translation,
promote RNA degradation, or both (27). For the understanding of the functions of the predicted conserved miRNAs, the target genes were predicted using the psRNATarget program with default parameters (http://plantgrn.noble.org/) against the *Jatropha curcas* mRNA transcripts. A total of 712 targets genes were predicted for the six predicted miRNAs based on the complementarity with their target sequences. The predicted targets were belonging to a number of gene families that involved in regulation of metabolism, hormone biosynthesis, transcription factor, development and oil synthesis (Supplementary file). The miRNA, miR414 exhibited the highest 233 target genes followed by miR407, miR2938, miR846, and miR5658 with 115, 97, 96, 89, and 82 targets, respectively. Majority of the identified target mRNAs were found to be transcriptional factors, while other were involved in metabolism and development and response to biotic and abiotic stresses.

**Discussion**

It has been shown that majority of the plant metabolic processes and responses to environment are controlled by miRNAs. Identification of miRNAs and their targets in Jatropha will help not only in understanding the mechanism of control of cellular processes but also help in controlling the traits. In this study, jatropha genome sequence is used for the prediction of miRNAs and their targets. A total of six conserved miRNAs was identified which shows similar characteristics to the miRNAs identified in other plants. The predicted miRNAs play role by regulating genes which are mostly involved in biotic, abiotic stress responses and plant developmental pathways. The predicted target genes are belonging to diverse functions like in metabolism, transport, and genes involved in fatty acid metabolism. The predicted miRNAs show their targets which involve in the fatty acid metabolism like miR414 targets the gene of acetyl-CoA carboxylase and O-acyltransferase WSD1, miR846 targets gene of FAD omega-3 fatty acid, FAD plastid fatty acid desaturase, miR5658 have the target gene long chain acyl-CoA synthetase 1, and miR407 have the target gene Trigalactosyldiacylglycerol 2, miR2938 GDSL esterase/lipase, glycolipid transfer protein 1. Further validation and confirmation of the predicted targets is needed which would help to gain insight into the roles these conserved miRNAs play in the Jatropha oil synthesis. However, in-silico based prediction of other types of targets specifically involved at the phase of translation or post translation and which may not be identified
experimentally might prove to be useful. This result may apparently be in-silico mined but this
prediction of conserved miRNAs has resulted in significant enrichment of the repertoire of *Jatropha
curcas* miRNAs which further provides rich insights into miRNAs regulation of genes.

**Materials And Methods**

Data retrieval of *Jatropha curcas* genome and plant miRNAs

From the publicly available Genbank database, a complete set of 41790 whole genome shotgun
assembly sequences with with the project accession AFEW00000000 of *Jatropha curcas* L., were
downloaded (18). A total of 318 mature miRNAs of *Arabidopsis thaliana* belonging to group
viridiplantae were retrieved from the database miRBase (Release 19.0) available on
http://miRNA.sanger.ac.uk (19).

Employed Bioinformatics Software

For the prediction of miRNAs of *Jatropha curcas* L. a comprehensive tool C-mii version 1.11 was used
(20). For the prediction of secondary structures of pri-miRNA and pre-miRNAs, UNAFold program was
used (21). psRNATarget server based on homology search method was used for the target prediction
of conserved miRNAs (22). For the annotation of identified target program Blast2GO was used (23).

Potential miRNA prediction

Whole genome sequences assemblies were aligned with *Arabidopsis thaliana* miRNA sequences using
the BLAST program and, the parameters settings were adjusted to cut-off e-values as 0 (24).
Sequences having equal or less than four nt mismatches between the aligned WGS and miRNA
sequences were selected for subsequent scanning for possible miRNA sequences. To eliminate the
small RNAs (tRNA and rRNA) the matched sequences were conducted homology search against the
Rfam database using BLAST by setting the cut-off e-value 1e-8 and the sequences matched with the
Rfam were remove. Further the sequences were searched for homology using BLASTX against Uniprot
and TrEMBL and eliminating protein coding sequences to retain only non-protein coding sequences
(25). The remained sequences were entered as input in the program UNAFold for folding of primary
and precursor miRNAs. For pri-miRNA secondary structure folding, default UNAFold parameters were
used. The default parameter, as used in pri-miRNA folding was also applied for the pre-miRNA folding
in the C-mii software. Precursor miRNAs, was extracted from primary-miRNAs with 2 nucleotides at 3′ overhang by the C-mii tool. The selected sequences were considered as potential conserved miRNA candidate if the sequence follow certain criteria like: folding of sequence into a stem-loop hairpin structure, the mature miRNA localized in any one arm of the stem-loop structure, without any loop or break and less than six mismatch between miRNA-5p or miRNA-3p sequences, higher negative MFEI values of precursors than other types of RNAs, and miRNA-5p: miRNA-3p sequences form a duplex with two nucleotides 3′ overhang in stem-loop hairpin secondary structures (3).

Prediction of potential jcu-miRNAs targets and their functional annotation

The targets of conserved miRNAs were predicted through homolog search by subjecting conserved miRNA sequences as query against the J. curcas, protein-coding sequences using psRNATarget server at their default parameter (22). For the functional annotation of predicted targets, the most widely used tool, BLAST2GO was used (23). The workflow followed for prediction of miRNAs and their targets is shown in Fig. 1.

| miRNA   | Mature miRNA | LM | LPM | Strand | NM | GC%  | MFE (kcal/mol) | MFEI (kcal/mol) |
|---------|--------------|----|-----|--------|----|------|----------------|-----------------|
| jcu-miR407 | TTTAAATC ATTTAAAT TTGTT | 21 | 88  | +      | 4  | 21.59| -21.1         | -1.11           |
| jcu-miR414 | TCTTCATC ATCACAT ATGCTA | 21 | 49  | +      | 4  | 34.69| -15.9         | -0.93           |
| jcu-miR846 | TAGAATTT TAGTTGTT GATGAA | 21 | 84  | +      | 3  | 16.66| -15.3         | -1.09           |
| jcu-miR1886 | AATTTAAG ATTTTATT ATAAAT | 21 | 63  | +      | 4  | 9.52 | -6.5          | -1.08           |
| jcu-miR2938 | GAGATTG GAGAGAG TCCCAT | 21 | 38  | +      | 4  | 39.47 | -15.4       | -1.02           |
| jcu-miR5658 | ATTTTTT GCACTAAT CAGGT | 21 | 58  | -      | 4  | 34.48 | -22.4       | -1.11           |

LM = Length of miRNA, LPM = Length of pre- miRNA, NM = number of mismatch (nt) with homologous miRNAs

| Abbreviations |
|---------------|
| miRNAs: microRNAs, AGO: Argonaute, BLAST: Basic Local Alignment Search Tool, DCL1: DICER-LIKE1, DNA: Deoxyribonucleic acid, HEN1: HUA ENHANCER1, MFE: Minimal folding free energy, MFEI: Minimal folding free energy index, mRNA: messenger RNA, tRNA: transfer RNA, pri-miRNA: Primary microRNA, pre-miRNA: Precursor microRNA, pol II: RNA Polymerase II, RNA: Ribonucleic acid, sRNA: Small RNA, rRNA: ribosomal RNA, RISC: RNA Induced Silencing Complex |
Declarations
Acknowledgements

None.

Authors’ contributions All the bioinformatics analysis, and drafting of the manuscript is prepared by Nutan Prakash Vishwakarma.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843–54.

2. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. Genes & Development 2002;1616–26.

3. Carrington JC, Ambros V. Role of MicroRNAs in Plant and Animal Development. Science. 2003;301:336–8.

4. Lee Y, Kim M, Han J, Yeom K, Lee S, Baek SH. et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23:4051–60.

5. Nozawa M, Miura S, Nei M. Origins and evolution of microRNA genes in plant species. Genome Biol Evol. 2012;4:230–9.

6. Xie Z, Allen E, Wilken A, Carrington JC. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. Proc
7. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW. et al. Methylation as a crucial step in plant microRNA biogenesis. Science 2005; 307.

8. Ramachandran V, Chen X. Degradation of microRNAs by a Family of Exoribonucleases in Arabidopsis. Science. 2008;321:1490-2.

9. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39:D152-7.

10. Voinnet O. Origin, biogenesis, and activity of plant microRNAs. Cell. 2009;136:669-87.

11. Chen X. Small RNAs and Their Roles in Plant Development. Annu Rev Cell Dev Biol. 2009;25:21-44.

12. Khraiwesh B, Zhu J, Zhu J. Biochimica et Biophysica Acta Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. BBA-Gene Regul Mech. 2012;1819:137–48.

13. Achten WMJ, Nielsen LR, Aerts R, Lengkeek AG, Kjær ED, Trabucco A. et al. Towards domestication of Jatropha curcas. Knowl Creat Diffus Util. 2010;1:91–107.

14. Wang CM, Liu P, Sun F, Li L, Ye J, Yue GH. Isolation and identification of miRNAs in Jatropha curcas. Int J Biol Sci. 2012;8(3):418–29.

15. GSS Vishwakarma NP, Jadeja VJ. Identification of miRNA encoded by Jatropha curcas from EST and GSS Identification of miRNA encoded by Jatropha curcas from EST and GSS. 2017; 2324.

16. Galli V, Guzman F. de Oliveira LF V, Loss-Morais G, Körbes AP, Silva SD a, et al. Identifying microRNAs and transcript targets in Jatropha seeds. PLoS One 2014; 9: e83727.
17. Prakash VN, Jadeja VJ. Identification and characterization of miRNA transcriptome in Jatropha curcas by high-throughput sequencing. Plant Gene. 2018;15:44–50.

18. Benson DA, Boguski MS, Lipman DJ, Ostell J. GenBank. 1997;25(1):1–6.

19. Griffiths-Jones S, Saini HK. van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucl Acids Res. 2008; 36.

20. Numnark S, Mhuantong W, Ingsriswang S, Wichadakul D. C-mii: a tool for plant miRNA and target identification. BMC Genomics 2012; 13.

21. Markham NR, Zuker M. UNAFold BT-Bioinformatics: Structure, Function and Applications. In: Keith JM, editor. Totowa: Humana Press; 2008. pp. 3–31.

22. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011;39:W155–9.

23. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics 2008; 619832.

24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215:403 – 10.

25. Apweiler R. The universal protein resource (UniProt) in 2010. Nucleic Acids Res. 2009;38:193–7.

26. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson T, Tech T. Conservation and divergence of plant microRNA genes. Plant J. 2006;46:243–59.

27. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet. 2010;11:597–610.

Figures
Figure 1

Flow chart for Genome wide prediction of microRNAs from Jatropha curcas L.
Conserved miRNAs in J. curcas along with mature and precursor sequences and the predicted stem loop structure.

Supplementary Files

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

Targets.xlsx
SupplementaryFile.xls