J. Hortl. Sci.
Vol. 10(1):90-93, 2015

Short communication

A guide to *in silico* identification of miRNAs and their targets

V. Radhika, Kanupriya, R. Rashmi and C. Aswath
Division of Biotechnology
ICAR-Indian Institute of Horticultural Research
Hessaraghatta Lake Post, Bengaluru – 560 089, India
E-mail: vr@iihr.ernet.in

ABSTRACT

MicroRNAs (miRNA) are non-coding RNA molecules that play a critical role in gene regulation including translational repression in animals and mRNA cleavage in plants. MicroRNAs control various cellular, metabolic and physiological processes in living organisms. In this paper, we provide an overview on the significance of miRNA, nomenclature, their biogenesis and the pipelines for prediction of miRNA and their targets. These tools are important for identification of conserved miRNAs in crops where miRNAs have not been previously discovered. The newly-identified miRNAs and their targets play an important role in understanding regulation of growth, development and gene silencing in various life forms.

Key words: miRNA, bioinformatics, miRNA targets, structure, software tools

MicroRNAs (miRNAs) are non-coding RNAs (19-22 nt) molecules that are derived from one arm of the precursor miRNA sequences. These are produced from the non-coding portion of DNA and are generally transcribed as independent units. In plants, miRNAs bind to protein-coding regions of mRNAs and cause mRNA degradation (Llave *et al*, 2002) and translational repression at the 'seed region' (i.e., 2-8 nts at 5' end of a mature miRNA). In plants, miRNAs are processed from transcripts that can fold into a stable hairpin (Llave *et al*, 2002). Several miRNA sequences have been found to be highly conserved in different species, and, pre-miRNAs have a unique secondary structure, which helps identify them through *in silico* approaches.

Nomenclature

Predicted miRNAs are named as per MiRBase guidelines (Griffiths–jones, 2006). Name of the microRNA consists of the prefix 'mir', followed by a dash. For example, osa-miR444 is a miRNA where 'osa' indicates the name of the species, *Oriza sativa*, 'miR' indicates mature sequences, and '444' indicates the order of its discovery. Sometimes, both miR444a and miR444b are present. Here '444a' indicates that it was discovered before miR444b. Sometimes, microRNAs are denoted as miR-444-5p or miR-444-3p, which indicates the origin of microRNAs from the 3' and 5' end, respectively.

Biogenesis of microRNAs

MicroRNA genes are found in the intergenic regions of DNA sequences. The process of miRNA biogenesis starts from the nucleus, and is completed in the cytoplasm. In the nucleus, sequences that contain mature miRNA sequences are transcribed by RNA polymerase II (polu II) into a primary RNA. The primary miRNA is then processed in the nucleus by endonuclease into a precursor miRNA sequence, containing 60-100nts long stem loop structure. The pre-miRNA is then cleaved into a miRNA:miRNA* duplex by a Dicer-like enzyme (DCL-1) in the nucleus, and, these sequences are exported from nucleus to the cytoplasm**.** In the cytoplasm, one of these strands of precursor miRNA produces the mature miRNA, which is approximately 22nts. This gets associated with the RNA-induced silencing complex (RISC) to interact with its mRNA targets.

Source of sequences for miRNA prediction

MicroiRNA can be predicted from different sequences, viz., expressed sequence tags (ESTs) (Reddy *et al*, 2012), genomic survey sequences (GSS), new generation sequences (NGS) (Kanupriya *et al*, 2013), or unigenes. These sequences can be generated or extracted from any public repository database and used for the prediction of miRNA. Known miRNA sequences are available in miRBase database (http://www.mirbase.org).

Prediction of conserved miRNAs

BLASTX/ BLASTN tools help identify query sequences that contain miRNA homologs. Precursor miRNA sequences are extracted from sequences containing miRNA homologs by taking into consideration 50 nucleotides upstream and 50 nucleotides downstream from the mature miRNA position.

Secondary structure and mature microRNA prediction

Secondary structure of these pre-miRNA sequences is then predicted and the minimum free energy is computed. Identification of mature miRNA depends on the following parameters (Reddy *et al*, 2012):

1. RNA sequences should fold into a complete stemloop hairpin
2. Length of mature miRNAs should be between 19 and 21 nts
3. Predicted miRNAs should have ≦ 2 nt mismatches
4. Minimum free-energy of the secondary structure should be ≧ 18 kcal $mole^{-1}$
5. A+U content should be in the range of 30-70%

Target prediction

MiRNA target genes control biological, metabolic and physiological processes in plants and, hence, identification of their targets is important. They help understand the role and functional importance of miRNAs. It has been shown that one miRNA can target more than one regulatory gene. Functional characterization of a miRNA target is essential for providing a biological insight into each miRNA-mediated pathway (Reddy *et al*, 2012). In plants, miRNAs are important in regulating plant growth and development**.** A flow-chart depicting various steps in the prediction of miRNA and their targets is presented in Fig. 1.

Functional annotation of miRNA targets

Identification of biological information of the coding portion of a sequence is an important aspect. MicroRNAs play an important role in regulating gene expression in a variety of manners, including translational repression, mRNA cleavage and deadenylation, in both plants and animals. The role of individual miRNAs in an organism, namely biochemical, biological, metabolic, gene expression, and physiological function, can be predicted using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) tools.

Tools for miRNA and target prediction

Various tools, both online and offline, are available for predicting miRNA, their secondary structures and targets (Table 1); miRAuto (Lee *et al*, 2013) is a comprehensible tool for miRNA prediction from small RNA sequencing data in plant species. miRAuto software analyzes the expression

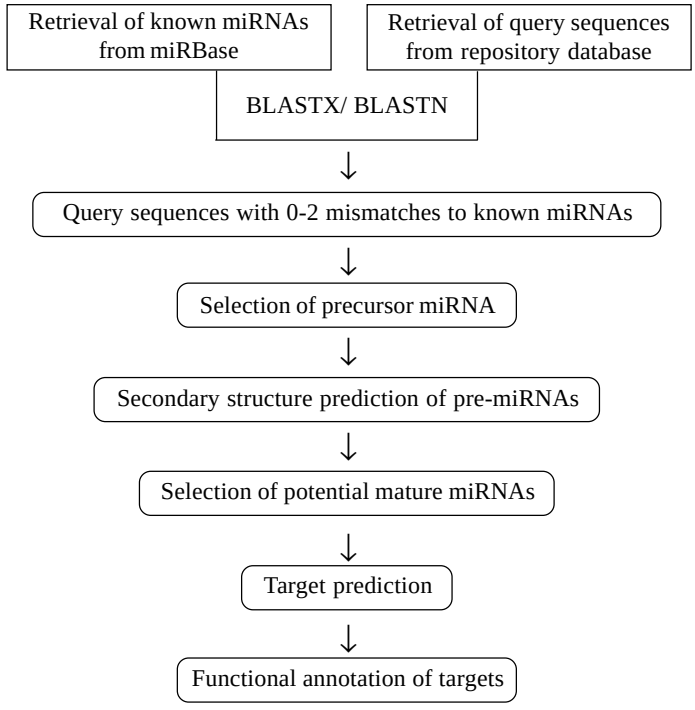

Fig. 1. Flowchart for the prediction of miRNA and their targets

Table 1. Tools for miRNA analysis

Tool	Website
Tools for miRNA and Target prediction.	
miRAuto	http://nature.snu.ac.kr/software/miRAuto.htm
MaturePred	http://nclab.hit.edu.cn/maturepred/
miRPara	http://www.whiov.ac.cn/bioinformatics/mirpara
miRDeep	*www.australianprostatecentre.org/research/software/mirdeep-star*
MicroPC	http://www.biotec.or.th/isl/micropc
C-mii	http://www.biotec.or.th/isl/c-mii/documentation.Php
miRTour	http://bio2server.bioinfo.uni-plovdiv.bg/miRTour/
psRNATarget	http://plantgrn.noble.org/psRNATarget/
TAPIR	http://bioinformatics.psb.ugent.be/webtools/tapir
Tools for structure prediction	
RNAfold	*subtiliswiki.net/wiki/index.php/RNAfold_WebServer*
UNAfold	http://www.bioinfo.rpi.edu/applications/hybrid/download.php
MFold	http://www.bioinfo.rpi.edu/applications/mfold
Tools for functional annotation	
GO	*www.geneontology.org*
KEGG	*www.genome.jp/kegg*

of 5' -end position of compared RNAs in reference sequences, to candidate miRNAs, for the possibility of presence of miRNA fragments. MaturePred tool, based on machine learning method, is used for accurately predicting plant miRNAs. Using this tool, we can extract the position, structure and energy related information from real/ pseudo miRNA:miRNA* duplex; miRPara (Wu *et al*, 2011) is based on SVM, and predicts mature miRNA coding regions from genome-scale sequences. In this tool, sequences are classified from miRBase into animal, plant and overall categories, and it uses a support vector machine to train the three models based on an initial set of 77 parameters related to physical properties of the pre-miRNA and its miRNAs; miRDeep is a non-comparative computational method developed for identification of miRNAs from a pool of sequenced RNA transcripts, obtained by deep-sequencing experiments (An *et al*, 2013). This method at first aligns the transcript reads to genomic locations, and selects genomic sequences from locations that can form hairpin secondary structures.

MicroPC (μPC) (Mhuantong *et al*, 2009) is an online tool for predicting and comparing plant miRNAs and their targets. It offers three, main interactive pages for comparing, searching and predicting plant miRNAs. Target-align was proposed for plant miRNA target identification, and developed as both web and command line versions. C-mii (Numark *et al*, 2012) is a stand-alone software package, with graphical user interface for identifying, manipulating and analyzing plant miRNAs and targets. C-mii tool performs sequence-similarity search, secondary-structure folding, automatic stem-loop identification and manipulation, and, functional and gene ontology (GO) annotation. It can be used for plant miRNA and target prediction only; miRTour (Milev *et al*, 2011), based on comparative approach, is used for both miRNA and target prediction. All the steps of miRNA and target prediction like homolog search, miRNA precursor, target prediction and annotation, are performed by the same set of input sequences. psRNATarget (Dai *et al*, 2011) is a plant small RNA target analysis server, which consists of two important functions: (i) Reverse complementary matching between small RNA and target transcript using a proven scoring schema, and (ii) Target-site accessibility evaluation by calculating unpaired energy (UPE) required to 'open' secondary structure around small RNA's target site on mRNA. The psRNA Target incorporates recent discoveries in plant miRNA target recognition. TAPIR (Bonnet *et al*, 2010) is a web server designed for the prediction of plant microRNA targets. The server offers a possibility of searching for plant miRNA targets, using a fast and a precise algorithm.

Tools for miRNA structure prediction

RNAfold (Zuker and Stiegler, 1981) is a tool which reads RNA sequences, calculates their minimum free energy and structure, and, returns the structure in bracket notation and its free energy. UNAFold software (Markham *et al*, 2008) is a collection of several programs that simulate folding, hybridization, and melting pathways for one or two single-stranded nucleic acid sequences. Secondary structure prediction for single-stranded RNA or DNA combines free energy minimization, partition function calculations and stochastic sampling. It is an offline tool. Mfold is a web server for prediction of secondary structure of single-stranded nucleic acids.

MicroRNAs regulate gene expression in a variety of ways such as translational repression, mRNA cleavage and deadenylation in plants. A number of computational tools based on comparative and non-comparative algorithms are available for identification of mature miRNA and their targets. In this study, an algorithm for prediction and analysis of miRNAs through bioinformatics tools has been presented.

REFERENCES

An, J., Lai, J., Lehman, M.L. and Nelson, C.C. 2013. miRDeep: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res.*, **41:**727-37

Bonnet, E., He, Y., Billiau, K. and Peer, Y.V. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, **26:**1566-1568

Dai, X. and Zhao, P.X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, **39:**155-159

Griffiths-Jones, S. 2006. miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**:129–138

Kanupriya, C., Radhika V. and Ravishankar, K.V. 2013. 'Mining of miRNAs in pomegranate (*Punica granatum* L.) by pyrosequencing of part of the genome.' *J. Hort'l. Sci. Biotech.*, **88**:735-742

Lee, J., Kim, D., Park, J.H., Choi, I. and Shin, C. 2013. miRAuto: An automated user-friendly microRNA prediction tool utilizing plant small RNA sequencing data. *Molecules and Cells*, **35:**342-347

Llave, C., Kasschau, K.D., Rector, M. and Carrington, J.C. 2002. Endogeneous and silencing-associated small RNAs in plants. *Pl. Cell*, **14:**1605–1619

Markham, N.R. and Zuker, M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol*., **453:**3-31

Mhuantong, W. and Wichadakul, D. 2009. MicroPC (ìPC): A comprehensive resource for predicting and comparing plant microRNAs. *BMC Genomics*, **10:**366

Milev, I., Yahubyan, G., Minkov, I. and Baev, V. 2011. miRTour: Plant miRNA and target prediction tool. *Bioinformation*, **6:**248-249

Numnark, S., Mhuantong, W., Ingsriswang, S. and Wichadakul, D. 2012. C-mii: a tool for plant miRNA and target identification. *BMC Genomics*, **13:**7-16

Reddy, D.C.L., Radhika, V., Bhardwaj, A., Khandagalek, S. and Aswath, C. 2012. miRNAs in brinjal (*Solanum melongena*) mined through an *in silico* approach. *J. Hort'l. Sci. Biotech.,* **87:**186-192

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. 2002. Prediction of plant microRNA targets. *Cell*, **110:**513–520

Wu, Y., Wei, B., Liu, H., Li, T. and Rayner, S. 2011. MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. *BMC Bioinformatics*, **12:**107

Zuker, M. and Stiegler, P. 1981. Optimal computer folding of large RNA sequences using thermodynamic and auxiliary information. *Nucl Acid Res*, **9:**133-148

(MS Received 07 June 2014, Revised 31 March 2015, Accepted 07 April 2015)