

MTide: an integrated tool for the identification of miRNA–target interaction in plants

Zhao Zhang¹, Li Jiang¹, Jingjing Wang¹, Peizhen Gu² and Ming Chen^{1,*}

¹Department of Bioinformatics, State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China and ²Department of Control Science and Engineering, Zhejiang University, Hangzhou 310058, China

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: Small RNA sequencing and degradome sequencing (also known as parallel analysis of RNA ends) have provided rich information on the microRNA (miRNA) and its cleaved mRNA targets on a genome-wide scale in plants, but no computational tools have been developed to effectively and conveniently deconvolute the miRNA–target interaction (MTI).

Results: A freely available package, MTide, was developed by combining modified miRDeep2 and CleaveLand4 with some other useful scripts to explore MTI in a comprehensive way. By searching for targets of a complete miRNAs, we can facilitate large-scale identification of miRNA targets, allowing us to discover regulatory interaction networks.

Availability and implementation: <http://bis.zju.edu.cn/MTide>

Contact: mchen@zju.edu.cn

Received on June 15, 2014; revised on September 16, 2014; accepted on September 18, 2014

1 INTRODUCTION

microRNAs (miRNAs) are a kind of small endogenous non-coding RNAs that negatively control gene expression at the post-transcriptional level by degrading target mRNAs in plants (Bartel, 2009). To make the functions of miRNA totally clear, it is essential to exactly identify the conserved and species-specific miRNAs and their target genes. Small RNA sequencing and degradome sequencing have been widely used in plants to identify miRNAs and target genes on a genome-wide scale. Many tools have been developed to handle small RNA sequencing and degradome data separately, but no specialized tool exists for a one-stop survey of this kind of interaction.

miRNAs are cleaved from stem-loop structured precursors called pre-miRNAs by Dicer-like 1 (Kurihara and Watanabe, 2004). Maturation of miRNAs releases RNA fragments derived from different parts of pre-miRNAs with asymmetric abundance. miRDeep2 (Friedlander *et al.*, 2012), a new version of miRDeep, uses the position and frequency of sequenced reads along with the secondary structure of pre-miRNAs and uses a probabilistic model of miRNA biogenesis in animals to score the possible miRNA sequence. Compared with animals, miRNA precursors are much longer with more variable lengths and more miRNAs belong to paralogous families with multiple

members in plants (Yang and Li, 2011). Therefore, it cannot be applied to the plant systems straightforward (Thakur *et al.*, 2011).

CleaveLand (Addo-Quaye *et al.*, 2009) was developed specifically to analyze degradome data. However, because of the algorithms implemented in it, it is impractical to analyze all possible miRNA–target interactions (MTIs) in a reasonable timescale for miRNAs identified from small RNA sequencing data. Moreover, it only reports targets cleaved between the 10th and 11th nucleotides, but cleavages between 9th and 10th or 11th and 12th nt have also been reported (Allen *et al.*, 2005; Jagadeeswaran *et al.*, 2009).

Here, we present MTide, an integrated tool for exploring MTI in plants. It consists of a modified miRDeep2 with plant-specific parameters, a modified CleaveLand4 with more targets reports and some other useful scripts. To control the running timescale in a low level, we add multiple threads support for miRDeep2, CleaveLand4 and other time-consuming parts.

2 FEATURES

Based on deep sampling of small RNA libraries and degradome fragments of target mRNA by next-generation sequencing, MTide enables users to explore expression patterns of known miRNA genes, discover novel ones and identify target mRNA of these miRNAs. Figure 1 illustrates the workflow of MTide. It is composed of four modules, which are listed below:

- Module 1: miRNA identification

First, the small RNA sequencing reads are processed by removing 5' and 3' adaptors, discarding reads that are <18 nt or >30 nt and parsing them into FASTA format with their copy number recorded. Second, the cleaned reads are mapped to other non-coding libraries, such as RFam, RepBase and CDS sequences, to remove mRNA, rRNA, tRNA, snRNA, snoRNA and other non-coding RNAs. Third, the preprocessed reads are mapped to genome sequences using Bowtie (Langmead *et al.*, 2009) with not more than two mismatches and 15 valid alignments allowed. The mapping result file will be transformed to 'arf' format, which was introduced by miRDeep2. Finally, the re-parameterized miRDeep2 is invoked. The length of sequences for predicting RNA secondary structure is set to 250, and a plant-specific

*To whom correspondence should be addressed.

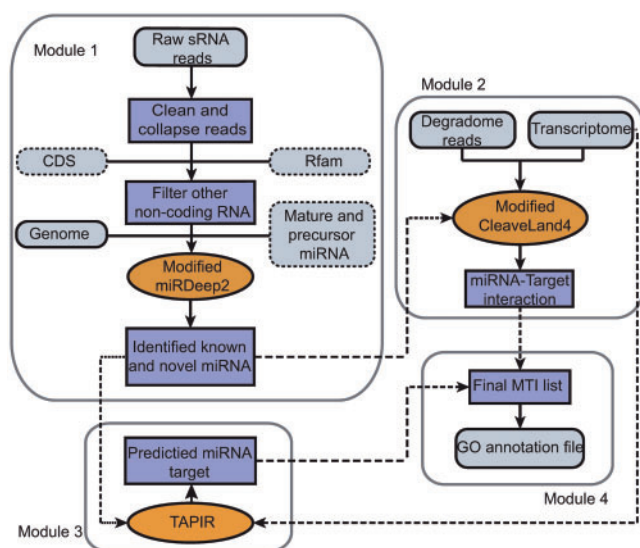


Fig. 1. Diagram of the workflow of MTide

scoring system is added to miRDeep2. The overall steps quantify the known miRNA expression and identify novel ones with stem-looped precursors.

- Module 2: miRNA target identification

To make target searching time efficient, we only consider mRNA having degradome signature while running GSTar.pl, which is a core script in CleaveLand4 package. To find more targets, we add support for cleavages between 9th and 10th or 11th and 12th nt of the miRNA sequences, as suggested by SeqTar (Zheng *et al.*, 2012).

- Module 3: miRNA target prediction

TAPIR is one of the most precise tools for miRNA target prediction in plants (Bonnet *et al.*, 2010). Here, we use the ‘precise search’ option of TAPIR and add multiple threads support for it.

- Module 4: prioritization of the predicted target

The targets of the same miRNA tend to share similar functions. To remove false-positive targets predicted by TAPIR, all the targets are scored based on GO similarity to targets of the same miRNA identified from degradome data.

We also have developed an integrated script including all modules listed above, and users can run this script just by modifying the accompanying configure file. In case of two small RNA sequencing samples, Mtide provides an option for differential expression analysis of known miRNA using DESeq (Anders and Huber, 2010).

3 IMPLEMENTATION

The MTide package was developed in Perl by combining the core algorithm of miRDeep2 and CleaveLand4, cutadapt (Martin, 2011) for adaptor removing, Bowtie for reads mapping and the Vienna RNA package for RNA secondary structure prediction. If users want to prioritize the predicted targets based on GO similarity and do differential expression analysis of miRNA, two R packages, csbl.go (Ovaska *et al.*, 2008) and DESeq, should also be installed. All the scripts can be executed sequentially or by an integrated script MTide.pl in a command-line environment. The package has been tested on two Linux platforms, Ubuntu 12.14 and Fedora 8, and should work on similar systems that support Perl and R. The package and user manual can be obtained from <http://bis.zju.edu.cn/MTide>.

Funding: This work was supported by the National Natural Sciences Foundation of China [31371328, 30971743], National Science and Technology Project of China [2008AA10Z125]; the Fundamental Research Funds for the Central Universities, and the Program for Innovative Research Team in University.

Conflict of interest: none declared.

REFERENCES

- Addo-Quaye, C. *et al.* (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, **25**, 130–131.
- Allen, E. *et al.* (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.
- Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Bonnet, E. *et al.* (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, **26**, 1566–1568.
- Friedlander, M.R. *et al.* (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.
- Jagadeeswaran, G. *et al.* (2009) Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families. *New Phytol.*, **184**, 85–98.
- Kurihara, Y. and Watanabe, Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl Acad. Sci. USA*, **101**, 12753–12758.
- Langmead, B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 10–12.
- Ovaska, K. *et al.* (2008) Fast gene ontology based clustering for microarray experiments. *BioData Min.*, **1**, 11.
- Thakur, V. *et al.* (2011) Characterization of statistical features for plant microRNA prediction. *BMC Genomics*, **12**, 108.
- Yang, X. and Li, L. (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, **27**, 2614–2615.
- Zheng, Y. *et al.* (2012) SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Res.*, **40**, e28.