BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Web Server

Volume 6(6)

miRTour: Plant miRNA and target prediction tool

Ivan Milev, Galina Yahubyan , Ivan Minkov, Vesselin Baev*

Department of Plant Physiology and Molecular Biology, University of Plovdiv, 24 Tsar Assen St, 4000 Plovdiv, Bulgaria; Vesselin Baev - Email: vebaev@gmail.com; *Corresponding author

Received April 05, 2011; Accepted April 25, 2011; Published June 06, 2011

Abstract:
MicroRNAs (miRNAs) are important negative regulators of gene expression in plant and animals, which are endogenously produced from their own genes. Computational comparative approach based on evolutionary conservation of mature miRNAs has revealed a number of orthologs of known miRNAs in different plant species. The homology-based plant miRNA discovery, followed by target prediction, comprises several steps, which have been done so far manually. Here, we present the bioinformatics pipeline miRTour which automates all the steps of miRNA similarity search, miRNA precursor selection, target prediction and annotation, each of them performed with the same set of input sequences.

Keywords: microRNAs, microRNAs targets, plant miRNA homologs

Availability: miRTour is provided as a web-based service at http://bio2server.bioinfo.uni-plovdiv.bg/miRTour/

Background:

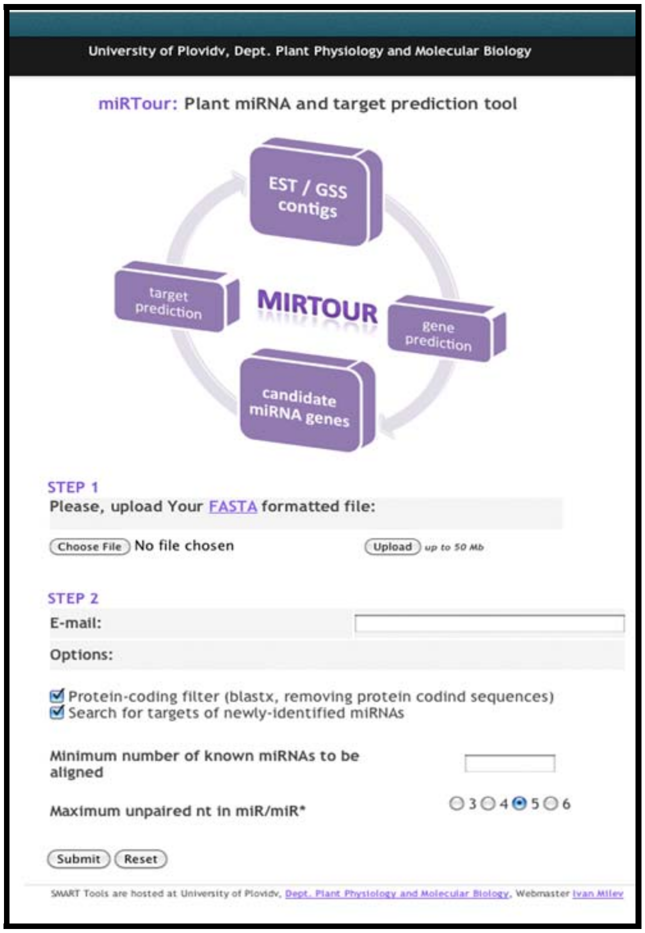
MicroRNAs (miRNAs) are important components of the regulatory networks in plant and animals. Plant miRNAs have been well characterized in several model systems for plant genomics such as Arabidopsis, Brachypodium, Oryza and Populus **[1-6]**. They are single-stranded 20–22 nt small RNAs that are endogenously produced from their own genes **[2]**. MiRNA genes in plants are systematically grouped into families that yield almost identical miRNAs. Currently, 2952 miRNA genes across 43 plant species were identified and annotated in miRBase v16 **[7]**. Here, we present the computational approach miRTour for homology-based discovery of plant miRNA and their targets from sequencing datasets (EST, GSS, SRA, etc.). The program automates all the steps of miRNA similarity search, miRNA precursor selection, target prediction and annotation, each of them performed with the same set of input sequences. The miRTour software has a user friendly web-interface which provides a comprehensive pipeline for identification of plant miRNA homologs.

Implementation:

We developed a web-based application for identification of new homologs of plant miRNAs in EST/GSS/SRA or assembled contig dataset. The program uses a local database that consists of all plant mature miRNA sequences from miRBase v16 (http://www.mirbase.org). The first step in our algorithm is mapping the plant mature miRNA sequences onto the ESTs/GSS dataset provided by the user. For the short sequence mapping, we employ the Gassst (Global Alignment Short Sequence Search Tool) v1.27, with no more than 4 mismatches. **[8]**. In the next step, the miRTour filters out the protein-coding ESTs/GSSs with miRNA hits applying BLASTX with E value 10^{-6} with *Arabidopsis* and *Orysa* protein sequences (A local copy of the BLAST tool was obtained from the NCBI FTP server, ftp://ftp.ncbi. nih.gov/blast). Next, putative miRNA foldback precursors are extracted from the non-coding ESTs/GSSs that contain matched known miRNAs. For selection of new pre-miRNA and miRNA candidates, the following criteria are taken into consideration: (1) MFEI (minimum free energy index) for the predicted pre-miRNA secondary structure to be greater than 0.7; (2) the number of the mismatches in miRNA/miRNA* must be less than 5; (3) the number of consecutive mismatches in miR/miR* must be less than 4. The secondary structure of the putative precursor sequence is predicted and visualized with the RNAfold algorithm from Vienna RNA package, on which bases the described criteria are evaluated **[9]**. Using ClustalW, the miRTour creates a colour alignment of the predicted precursor sequence with sequences of known plant mature miRNAs. To find the protein-coding targets of newly identified miRNAs in the same initial EST/GSS dataset, we have wrapped the software for target searching TargetFinder v1.6 in our script **[10]**. The annotation of the protein-coding target sequences is performed against all known *Arabidopsis* and *Orysa* protein sequences by BLASTX.

Software Input:

The miRTour is freely available at http://bio2server.bioinfo.uni-plovdiv.bg/miRTour/. It has a simple and intuitive web-based user interface **(Figure 1)**. In order to run a job on the miRTour, the following data are required as an input: **Step 1:** First, the user has to upload a file with EST/GSS/contig sequences (up to 50MB) in multi FASTA format. **Step 2:** The user has to input an e-mail address. This step is obligatory since the batch process of a large dataset requires more server-side time. The user can turn on/off both, the BLASTX filter and the target search module, if needed. Other options that the user should specify are the minimum number of plant homologs that the miRTour must align and the maximum number of unpaired bases in miR/miR* (the default parameter is 5).



University of Plovidv, Dept. Plant Physiology and Molecular Biology

miRTour: Plant miRNA and target prediction tool

STEP 1

Please, upload Your FASTA formatted file:

Choose File No file chosen

Upload up to 50 Mb

STEP 2

E-mail:

Options:

☑ Protein-coding filter (blastx, removing protein codind sequences)
☑ Search for targets of newly-identified miRNAs

Minimum number of known miRNAs to be aligned

Maximum unpaired nt in miR/miR* ○ 3 ○ 4 ● 5 ○ 6

Submit Reset

SMART Tools are hosted at University of Plovidv, Dept. Plant Physiology and Molecular Biology, Webmaster Ivan Milev

Figure 1: The miRTour software web-interface.

Software Output:

The output results are sent automatically by e-mail in HTML and PDF formats. The result data will be kept on the server for 5 days from the moment of submission. For each query dataset submitted to the miRTour, the output consists of sections with putative pre-miRNA and mature miRNA sequences; homology alignment of the newly-identified miRNA precursors to known plant miRNAs; details for secondary structure of the precursor - MFEI, CG% and number of mismatches in miR/miR*; predicted target sequences for the new mature miRNAs.

Conclusion:

High-throughput parallel sequencing technologies have made possible the production and accumulation of large genome and transcriptome datasets (EST/GSS contigs). Homology-based discovery of miRNAs in plant species, along with target prediction step, were a comprehensive assignment, which have been done so far manually. The developed miRTour automates those tasks and provides a pipeline for identification of new miRNAs and prediction of their targets from a single dataset of sequences.

Caveat and future development:

Due to server limitation, large datasets (more than 50MB) cannot be submitted for analysis by the miRTour. As the future development of our tool, we plan to extend our server capacity in a way that will allow larger file uploading. Another stage of development is to upgrade the miRTour to work with animal miRNAs, but it will take a longer time since the prediction of animal miRNA targets is more complex task compared to plant miRNA target prediction.

References:

[1] Baev V *et al*. *Genomics* 2011 **97**: 282 [PMID: 21371551]
[2] Bartel DP. *Cell* 2004 **116**: 281 [PMID: 14744438]
[3] Klevebring D *et al*. *BMC Genomics.* 2009 **10**: 620 [PMID: 20021695]
[4] Sunkar R *et al. Plant Cell* 2005 **17**: 1397 [PMID: 15805478]
[5] Sunkar R & Jagadeeswaran G. *BMC Plant Biol.* 2008 **8**: 37 [PMID: 18416839]
[6] Sunkar R *et al. BMC Plant Biol.* 2008 **8**: 25 [PMID: 18312648]
[7] Griffiths-Jones S *et al. Nucleic Acids Res.* 2008 **36**: D154 [PMID: 17991681]
[8] Rizk G & Lavenier D. *Bioinformatics* 2010 **26**: 2534 [PMID: 20739310]
[9] Hofacker IL. *Nucleic Acids Res.* 2003 **31**: 3429 [PMID: 12824340]
[10] Fahlgren N *et al*. *PLoS One.* 2007 **2**: e219 [PMID: 17299599]

Edited by P Kangueane

Citation: Milev *et al.* Bioinformation 6(6): 248-249 (2011)