BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 14 2014, pages 2066–2067
doi:10.1093/bioinformatics/btu147

Genome analysis

Advance Access publication March 14, 2014

comTAR: a web tool for the prediction and characterization of conserved microRNA targets in plants

Uciel Chorostecki and Javier F. Palatnik*

IBR (Instituto de Biología Molecular y Celular de Rosario), CONICET and Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, 2000 Rosario, Argentina

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: MicroRNAs (miRNAs) are major regulators of gene expression in plants and animals. They recognize their target messenger RNAs (mRNAs) by sequence complementarity and guide them to cleavage or translational arrest. So far, the prediction of plant miRNA–target pairs generally relies on the use of empirical parameters deduced from known miRNA–target interactions.

Results: We developed comTAR, a web tool for the prediction of miRNA targets that is mainly based on the conservation of the potential regulation in different species. We used data generated from a pipeline applied to transcript datasets of 33 angiosperms that was used to build a database of potential miRNA targets of different plant species. The database contains information describing each miRNA–target pair, their function and evolutionary conservation, while the results are displayed in a user-friendly interface. The tool also allows the search using new miRNAs.

Availability and implementation: The Web site is free to all users, with no login requirements, at http://rnabiology.ibr-conicet.gov.ar/comtar.

Contact: palatnik@ibr-conicet.gov.ar or chorostecki@ibr-conicet.gov.ar

Received on September 20, 2013; revised on February 12, 2014; accepted on March 11, 2014

1 INTRODUCTION

microRNAs (MiRNAs) are small RNAs processed from larger precursors with foldback structures by ribonuclease type III enzymes (Bologna *et al.*, 2012). The mature miRNAs are assembled into ARGONAUTE complexes and recognize target messenger RNAs (mRNAs) by partial base complementarity to control their translation and stability (Axtell, 2013). Ancient miRNAs conserved across angiosperms with important biological roles have been identified, some of them even found in gymnosperms, ferns, lycopods and mosses (Cuperus *et al.*, 2011). The prediction of miRNA targets in plants usually relies on the total number and position of mismatches and the minimum free energy (MFE) hybridization (Allen *et al.*, 2005; Dai and Zhao, 2011; Jones-Rhoades and Bartel, 2004; Rhoades *et al.*, 2002; Schwab *et al.*, 2005; Wang *et al.*, 2004). Usually, miRNA-binding sites are conserved during evolution, and their conservation between *Arabidopsis thaliana* and rice has been exploited to identify miRNAs and targets (Jones-Rhoades and Bartel, 2004). More recently, it has been shown that the conservation of the target sites using partial genomic information can be used to identify new targets for conserved miRNAs in plants (Chorostecki *et al.*, 2012). Here, we present comTAR, a web-based application for the identification of miRNA targets in plants based on sequence conservation during evolution. The tool allows users to analyze the variations of known miRNA targets during evolution, and to predict new miRNA targets.

* To whom correspondence should be addressed.

2 METHODS

2.1 MiRNA and transcript sequences

Because miRNA sequences can vary in different species, especially positions 1, 20 and 21 (Chorostecki *et al.*, 2012), we used sequences 2–19 (18 nt) for the search. If there were still variations between different miRNAs of the same family, we used the most common sequence considering the genomes of *A.thaliana*, poplar and rice (miRNA consensus sequence), as described earlier (Chorostecki *et al.*, 2012). The user can also search with new small RNA sequences. Plant transcript sequence data were extracted from libraries of the Phytozome project (http://www.phytozome.net/), formed by nucleotide FASTA format files of spliced mRNA transcripts (UTR, exons), with or without alternative splice variants.

2.2 Target search

Target search was performed using PatMatch (Yan *et al.*, 2005), which allows mismatches (insertions, deletions and substitutions). We searched for potential targets with four mismatches (substitutions and insertions) to the 18 nt miRNA consensus, while G:U wobbles and bulges were also considered as mismatches.

2.3 Additional features

We have integrated third-party tools and in-house scripts to make the search for targets more powerful and useful.

- To perform the alignment of the miRNA–target pair, we developed an implementation of the Needleman–Wunsch dynamic programming algorithm (Needleman and Wunsch, 1970) in Perl (http://www.perl.org/).
- RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/) (Krüger and Rehmsmeier, 2006) was integrated by developing in-house scrpits to find the MFE hybridization of the miRNA–target duplex for each candidate.
- Candidate sequences were labeled with the best *A.thaliana* TAIR10 hit locus identifier (ID) as a TAG, using the annotation files of Phytozome. Genes from different species having the same TAG were grouped together.

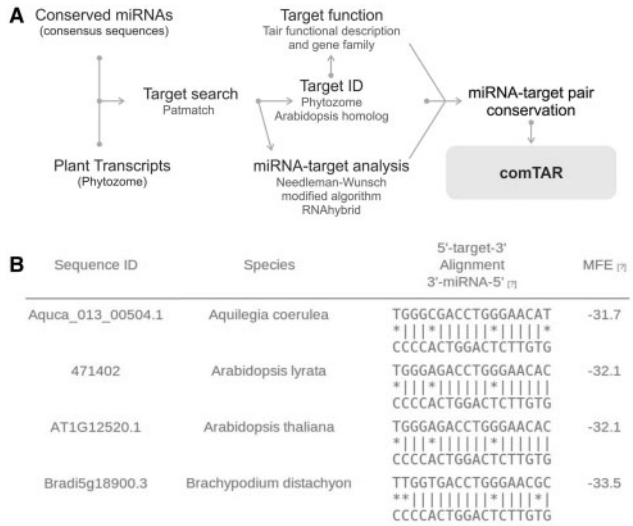

Fig. 1. comTAR (**A**) flowchart describing the database. (**B**) Output of comTAR showing the miR398/SOD1 (At1g12520) pair in different species

- Each *A.thaliana* TAG was indexed with a short description, curator summary and computational description belonging to the TAIR10 functional description file. Targets were also sorted into gene families according to the TAIR10 classification.

2.4 Web tool and data storage

ComTAR was designed as a web application implemented in an open-source PHP framework (Codeigniter) for the Graphical User Interface (GUI), but the analysis is based on a back-end written in Perl language, and for data storage, we chose MySQL (http://www.mysql.com/). The back-end performs the searches, the integration of third-party tools for increasing the specificity and sensitivity and the generation of final results, whereas the front-end is responsible for displaying the results (Fig. 1A). The TAG of the best hit in *A.thaliana* determines the number of species that a hit is present, a parameter that is defined by the user. The user can search for targets of new miRNAs and also has access to graphical display of the miRNA–target alignment, the hybridization energy and a filter for empirical miRNA–target parameters (Fig. 1B). Because plant miRNAs generally regulate genes coding for proteins of the same family, the tool has another feature that allows the search for targets by grouping the results by families rather than by gene-TAG. Furthermore, users can introduce a particular locus TAG (either the *A.thaliana* or Phytozome gene ID), and comTAR identifies the species where this particular gene can be an miRNA target.

3 DISCUSSION

We developed a web tool, comTAR, that allows the characterization of miRNA–target interactions in different plant species. The tool also allows the prediction of new candidates based mainly on the conservation of the miRNA–target pair with a relaxed number of mismatches. An advantage of the tool presented here is that evolutionary conserved miRNA–target interactions might be likely involved in relevant biological processes. The approach requests that miRNA targeting should be able to occur in the context of a minimum set of interacting parameters in different species. Therefore, the sequence of the target itself does not need to be conserved. A previous report using a similar strategy, but using a fewer species and datasets, was successfully used to detect targets previously validated in *A.thaliana* and to predict and experimentally validate new targets in *A.thaliana* (Chorostecki *et al.*, 2012). The web tool can also be used in a reverse way to search for species where a gene of interest is potentially regulated by an miRNA.

4 CONCLUSION

ComTAR allows users to predict and characterize miRNA-regulated genes by focusing on the conservation of the potential targeting. It provides an alternative strategy to other predictions tools based solely on empirical parameters.

ACKNOWLEDGEMENT

The authors thank Edgardo Bresso, Juan Debernardi, Carla Schommer and other members of JP laboratory for continuous help and discussions.

Funding: The Argentinean National Agency of Science and Technology (PICT) and fellowships of CONICET (to U.C.); J.P. is a member of the same institution.

Conflict of Interest: none declared.

REFERENCES

Allen,E. *et al.* (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.

Axtell,M.J. (2013) Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.*, **64**, 137–159.

Bologna,N.G. *et al.* (2012) Processing of plant microRNA precursors. *Brief. Funct. Genomics*, **12**, 37–45.

Chorostecki,U. *et al.* (2012) Identification of new microRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Res.*, **40**, 8893–8904.

Cuperus,J.T. *et al.* (2011) Evolution and functional diversification of miRNA genes. *Plant Cell*, **23**, 431–442.

Dai,X. and Zhao,P.X. (2011) psRNA target: a plant small RNA target analysis server. *Nucleic Acids Res.*, **39**, 155–159.

Jones-Rhoades,M.W. and Bartel,D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, **14**, 787–799.

Krüger,J. and Rehmsmeier,M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.

Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.

Rhoades,M.W. *et al.* (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513–520.

Schwab,R. *et al.* (2005) Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, **8**, 517–527.

Wang,X.-J. *et al.* (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol.*, **5**, R65.

Yan,T. *et al.* (2005) PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Res.*, **33**, W262–W266.