

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.

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 scripts 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.

* To whom correspondence should be addressed.

