

Target-align: a tool for plant microRNA target identification

Fuliang Xie and Baohong Zhang*

Department of Biology, East Carolina University, Greenville, NC 27858, USA

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: MicroRNAs (miRNAs) are important regulatory molecules. A critical step in elucidating miRNA function is identifying potential miRNA targets. However, few reliable tools have been developed for identifying miRNA targets in plants.

Results: Here, we developed a Smith–Waterman-like alignment tool in order to accurately predict miRNA targets. Dynamic programming was used to build a score matrix based on the complementarity of nucleotides in order to trace the optimal local alignments. Important parameters, such as maximum mismatches and maximum consecutive mismatches between miRNAs and their targets, were also used for filtering the optimal local alignments. Almost all of the parameters in this alignment tool can be adjusted by users. Compared to other target prediction tools, Target-align exhibits strong sensitivity and accuracy for identifying miRNA targets. More importantly, Target-align can identify multi-target sites as well potential for non-cleaved targets sites by change the default settings. Windows, web and command-line versions were developed to better serve different users.

Availability: <http://www.leonxie.com/targetAlign.php>.

Contact: zhangb@ecu.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 8, 2010; revised on September 11, 2010; accepted on September 29, 2010

1 INTRODUCTION

MicroRNAs (miRNAs) are an extensive class of small regulatory molecules that post-transcriptionally regulate gene expression by either cleaving mRNA transcripts or by repressing protein translation. Currently, a total of 16 978 miRNAs have been deposited in the miRNA database, miRBase (Griffiths-Jones *et al.*, 2008). Of these, 2600 are from plants. Although increasing evidences have demonstrated that miRNAs play important roles in plant growth and development, the functions of the majority of these miRNAs remains unknown mainly due to a lack of knowledge of miRNA targets.

Although many computational tools have been developed for predicting animal miRNA targets, few tools are available for identifying plant miRNA targets (Zhang *et al.*, 2006). There are currently three tools [miRU (Zhang, 2005), Helper tools (Moxon *et al.*, 2008) and TAPIR (Bonnet *et al.*, 2010)] available for predicting plant miRNA targets; however, these programs have a limited set of options that allow the user to adjust the search criteria. Previous studies demonstrated that most plant miRNAs

cleave targets by perfectly or near-perfectly binding to their target (Rhoades *et al.*, 2002; Schwab *et al.*, 2005). Thus, all the three currently available tools predict plant miRNA targets based on very strictly limited criteria. However, recent studies show that some miRNAs may inhibit translation by non-perfectly binding to target mRNAs with more potential target sites (Brodersen *et al.*, 2008; Montgomery *et al.*, 2008; Yu and Wang, 2010). Thus, a new computational program needs to be developed for identifying plant miRNAs with more flexible criteria. In this study, we developed an alignment tool named Target-align to integrate all potential factors influencing plant miRNA target prediction. Target-align is user-friendly and allows users to set a variety of parameters including alignment score setting, maximum score, number of consecutive mismatches, base site restrictions, number of G:U wobbles and number of gaps.

2 METHODS

The Target-align algorithm is similar to the Smith–Waterman algorithm (Smith and Waterman, 1981). Briefly, the algorithm first calculates a score for a specific miRNA and their potential targets (please refer details in the Supplementary Material). Then, the specific criteria are used to identify potential plant miRNA targets. When determining the optimal alignments and filtering the results, following special characteristics for plant miRNAs and their targets were considered according to previously established methods: ≤ 4 mismatches between the miRNA and its targets, ≤ 2 consecutive mismatches in the alignment region, no mismatches between bases 10–11; ≤ 1 mismatch from bases 1–9, ≤ 6 G:U wobbles, and ≤ 4 -nt gaps. A G:U wobble is only considered to be a 0.5 mismatch when calculating the total number of mismatches. In order to allow variance on the analysis of target genes, all of the above parameters were created so that they can be set by individual users.

3 RESULTS

3.1 Features of Target-align

Target-align was developed in C#, based on the .NET framework 3.0 of Microsoft (www.microsoft.com). The program should be run under the .NET framework 3.0 or on a higher .NET framework version. The interface of Target-align is very user-friendly. All of the parameters can be adjusted by individual users. Users can input both miRNA and mRNA sequences in order to do single target prediction, or they can upload files of miRNA and mRNA sequences and perform target prediction in batch form. However, all of the sequences in the files are required to be in FASTA format. Depending on the user's preference for considering G:U wobble base-pairing, Target-align allows users to turn on or turn off the G:U wobble calculation. Once G:U wobble base-pairing is canceled, the G:U wobble score, maximum number of G:U wobbles, and mismatch

*To whom correspondence should be addressed.

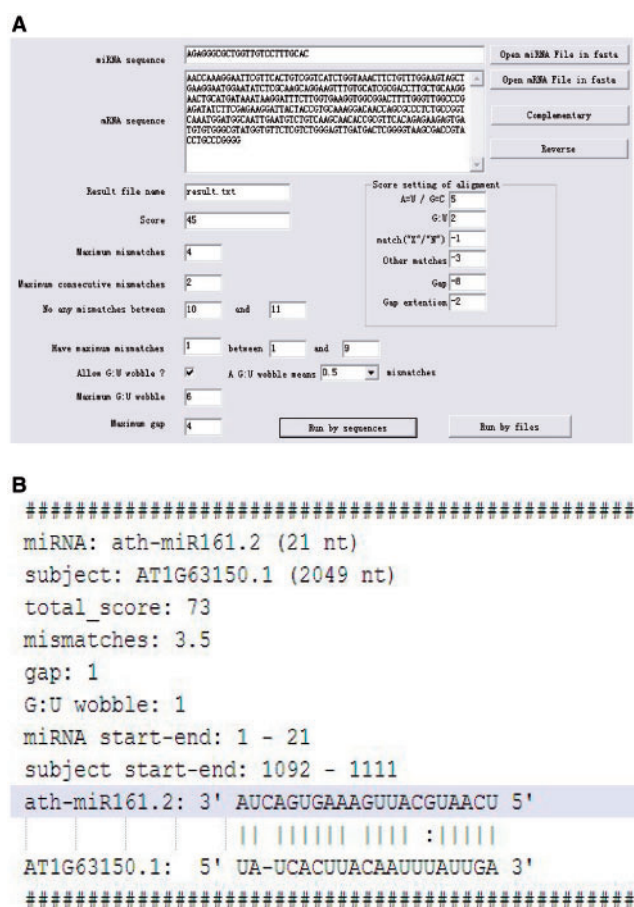


Fig. 1. Target-align interface (A) and search results (B).

point values of a G:U wobble base-pairs are invalid. Moreover, Target-align is available as a web version and command-line version. Users can open the tool in their program with a command-line as long as related parameters are assigned.

3.2 Output of alignment results

The output report includes the length of the miRNA and query sequence, total score, number of mismatches, number of gaps, number of G:U wobbles, and alignment results (Fig. 1). For traditional alignment tools like BLAST, WATER and FASTA, target sequences are often viewed incompletely due to end mismatches that are not considered. This leads to difficulty in calculating the exact number of mismatches between the miRNA and its potential target genes. Target-align prints the full query (miRNA) sequence and the target site sequence is extended to match the ends of the query sequence if there are mismatches at the ends.

3.3 Validation of Target-align

To test and compare the Target-align program to existing tools for plant miRNA target prediction, we tested Target-align with 102 validated miRNA-target pairs, as reported in TAPIR (Bonnet *et al.*, 2010). When Target-align was used to search against the

whole *Arabidopsis* transcriptome (TAIR annotation version 9, <http://www.arabidopsis.org/>), 95 of the 102 miRNA-target pairs (93.14%) from the reference set were detected with a false positive rate of 57.8%. By adjusting the maximum number of mismatches allowed (from 1 to 2) between bases 1 and 9, four additional miRNA-target pairs were detected; however, this change also increased the false positive rate from 57.8% to 84.0%. Finally, only three miRNA-target pairs were missed. According to TAPIR statistic results for sensitivity and specificity, the average amount of true positives obtained for FASTA and RNAhybrid were 91.83% and 93.14%, respectively, whereas the average amount of false positives for FASTA and RNAhybrid were 81.47% and 88.97%, respectively. Sm target and miRU were also tested by Bonnet and his co-workers (Bonnet *et al.*, 2010). One independent study showed that the average rate of true positives for miRU was 78.1% with a corresponding average false positive rate of 83.23% (Zhang, 2005). Another study determined that the average rate of true positives obtained by Sm target was 82.4 % with a 68.3% report of false positives (Moxon *et al.*, 2008). For Target-align, however, the true positive rate is 93.14% with a 57.8% false positive report. Therefore, Target-align exhibits better sensitivity and specificity. This is because more factors are considered by Target-align. By placing restrictions and parameters on the miRNA binding site, many false positive miRNA-target pairs were avoided.

Target-align can also identify multi-target sites (one miRNA may have multiple binding locations in a single target mRNA) as well potential for non-cleaved targets sites by changing the default settings, which are usually missed using other tools.

ACKNOWLEDGEMENT

We appreciate the suggestions from Robert Sheridan from MSKCC.

Funding: USDA (2010-01708) and NCBC (2009-BRG-1202) in part.

Conflict of Interest: none declared.

REFERENCES

- Bonnet,E. *et al.* (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, **26**, 1566–1568.
- Brodersen,P. *et al.* (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, **320**, 1185–1190.
- Griffiths-Jones,S. *et al.* (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
- Montgomery,T.A. *et al.* (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, **133**, 128–141.
- Moxon,S. *et al.* (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.*, **18**, 1602–1609.
- Moxon,S. *et al.* (2008) A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics*, **24**, 2252–2253.
- 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.
- Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
- Yu,B. and Wang,H. (2010) Translational inhibition by microRNAs in plants. *Prog. Mol. Subcell. Biol.*, **50**, 41–57.
- Zhang,B.H. *et al.* (2006) Computational identification of microRNAs and their targets. *Comp. Biol. Chem.*, **30**, 395–407.
- Zhang,Y.J. (2005) miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.*, **33**, W701–W704.