Target-align: a tool for plant microRNA target identification

Fuliang Xie and Baohong Zhang*
Department of Biology, East Carolina University, Greenville, NC 27858, USA

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, 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 changing the default settings.

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). There are 2,600 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 plant 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 test and compare the Target-align program to existing tools, the output report includes the length of the miRNA and query sequence if there are mismatches at the ends. Target-align is available as a web version and command-line version. 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. Sen 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 Sma target was 82.4% with a 68.3% report of false positives (Motson 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.


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