**Protocol for the comparative prediction of plant miRNA target prediction methods**

Gaurav Sablok1,\*

Universitat Potsdam, Germany

\*Corresponding author: Gaurav Sablok (gauravcodepro@gmail.com[)](mailto:sablokg@gmail.com)).

**Summary:**

Next generation sequencing has opened up new avenues for the identification of the microRNAs and their corresponding roles in abiotic and biotic stress responses. Recently, plethora of evidences suggest the canonical action of the miRNAs-mRNAs interactions to regulate the plant system biology at the post-transcriptional level thus leading to the gain or loss of genetic adaptation in plants. Mostly importantly identification of miRNA targets have always been the first challenge to identify and prioritize candidate genes as targets for miRNAs. In this protocol chapter, we present the detailed protocol on the identification of the miRNA targets using 6 comparative tools that have been developed for the miRNA target identification, which involves transcript and degradome based target prediction.

**Key words:** *microRNA, target prediction, Arabidopsis thaliana, TAPIR, psRobot, psRNATarget, TarHunter, Cleaveland*.

**Introduction:**

A broader view of the of the genetic adaptation of plants to the environmental surroundings and to understand the fluctuations at the transcriptional and the translational level has always been a challenge. With the advent of the next generation sequencing technologies, substantial knowledge has been acquired at the transcriptional level and post-transcriptional level to understand plant regulome. Post-transcriptional mechanism has been regulated by small 20-22 nt long sequences called as microRNAs, which play a critical role in the endonucleolytic post-transcriptional cleavage of the mRNAs. Irrespective of their association with the *ARGONATUTE* family and their diversity, miRNAs target prediction has always been the challenging task and several aspects such as local alignments, and exploration of the miRNA-mRNA duplexes have been considered for the accurate prediction of the miRNA targets in plants.

As compared to humans, plant miRNAs target predictions rules have been relatively easy to explore since the first definitive rules for the plant miRNA target predictions (Rhoades et al. 2002; Meyers et al. 2008) and have been subsequently revisited and revised for the seed length (Axtell, 2013) and have been reviewed (Dai et al. 2011). Considering the pre-definite rules and the perfect complementary between the miRNAs and the targets, several tools both as web-based and stand-alone such as TAPIR (Bonnet et al. 2010), SoMART (Li et al. 2012), psRNATarget (Dai et al. 2018), TargetFinder (Fahlgren et al. 207), psRobot (Wu et al. 2012), and TarHunter (Ma et al. 2018).

Most of these target prediction approaches uses FASTA as a search engine and local alignment approaches to infer the targets with exception such as psRNATarget, which used a scoring schema based on the seed length to allow for the target predictions. Alternative approaches, which uses cleaved target site interactions (degradome sequencing) can be predicted using the CleaveLand (Addo-Quaye et al. 2009) or by a combination of the transcript and degradome with associated expression filtered through the lasso-regression approaches (Zhang et al. 2017). In this protocol chapter, we provide protocol for the miRNA target prediction in plants using both the transcript based and the degradome based sequencing approaches.

**1. Required target prediction tools and associated dependencies:**

For the proper execution of the protocol for miRNA target predictions, following target predictions tools (Table 1) need to be installed.

Table 1: Summary of the plant miRNA Target Prediction Tools

|  |  |  |
| --- | --- | --- |
| miRNA Target Prediction Tool | Availability | Reference |
| TargetFinder | https://github.com/carringtonlab/TargetFinder | Fahlgren and Carrington, 2010 |
| psRNATarget | https://plantgrn.noble.org/psRNATarget/ | Dai et al. 2018 |
| TAPIR | http://bioinformatics.psb.ugent.be/webtools/tapir/ | Bonnet et al. 2010 |
| TarHunter | https://github.com/XMaBio/TarHunterL | Ma et al. 2018 |
| TarHunterL | https://github.com/XMaBio/TarHunterL | Ma et al. 2018 |
| psRobot | http://omicslab.genetics.ac.cn/psRobot/ | Wu et al. 2012 |

For the installation of the above tools, several dependencies scuh as CDHIT available from (<http://weizhongli-lab.org/cd-hit/>), GNU Parallel available from (<http://www.gnu.org/software/parallel/>), MCL available from (<http://micans.org/mcl/>), RNAhybrid available from (<http://bibiserv.techfak.uni-bielefeld.de/rnahybrid>), Perl (v5.8+) with Bioperl version 1.007002\_1, FASTA34 binaries available from (<http://fasta.bioch.virginia.edu/fasta_www2/fasta_down.shtml>), FASTA35 binaries available from (<http://fasta.bioch.virginia.edu/fasta_www2/fasta_down.shtml>), FASTA36 binaries available from (<http://fasta.bioch.virginia.edu/fasta_www2/fasta_down.shtml>), MFOLD version 3.5 available from <http://unafold.rna.albany.edu/?q=mfold/download-mfold> also need to be installed.

**2. File preparation for target predictions:**

For the present protocol, we have selected few microRNA families of *Arabidopsis thaliana* namely miR156, miR157 and miR158 families. The corresponding miRNAs families were downloaded from the miRBase and were selected for target prediction using 6 target predictions methods, which are focussed on the prediction of targets using the transcript binding sites. To ease the access of the miRNAs target predictions, a corresponding FASTA file of the miRNAs was prepared, which reflects the miRNAs and the corresponding family classification such as

***CODE:*** $head -n 6 selected\_Arabidopsis\_miRNAs.fasta

>ath-miR156a-5p MIMAT0000166 Arabidopsis thaliana miR156a-5p

UGACAGAAGAGAGUGAGCAC

>ath-miR156a-3p MIMAT0031865 Arabidopsis thaliana miR156a-3p

GCUCACUGCUCUUUCUGUCAGA

To integrate the target predictions in the loop shell, we used the following command to sort the miRNA sequences on one line:

***CODE:*** $sed '/^>/d' selected\_Arabidopsis\_miRNAs.fasta > seleted\_miRNAs\_target\_finder.fasta and the formatted file for the target find has now the following attributes:

***CODE:*** $head -n 6 seleted\_miRNAs\_target\_finder.fasta

UGACAGAAGAGAGUGAGCAC

GCUCACUGCUCUUUCUGUCAGA

UGACAGAAGAGAGUGAGCAC

which represents one single miRNA sequence on one line. For target prediction, corresponding targets input file was downloaded from Phytozome version 12 available from https://phytozome.jgi.doe.gov/pz/portal.html

**3. Target predictions using the TargetFinder:**

TargetFinder has been widely used for the prediction of the miRNAs targets in plants and predicts the miRNA target predictions using the Smith-Waterman alignment approaches (Fahlgren et al. 2007, Fahlgren and Carrington, 2010). For the prediction of the targets using the TargetFinder, there are 2 options, one is targetfinder.pl, which requires the input as miRNA sequence and the target database and the second option is the targetfinder\_threads.pl, which allows for FASTA formatted files both the miRNAs and the targets.

***NOTE:*** Before invoking the targetfinder, please export the PATH of the ssearch35\_t aligner to the targetfinder DIR like

export PATH=/FASTA35\_installation/bin:$PATH

Option 1 using the TargetFinder.pl, which requires the sequence and the target file. To do this in a for loop, the following loop can be invoked:

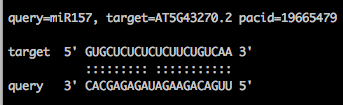
***CODE:*** cat seleted\_miRNAs\_target\_finder.fasta | while read line; do perl targetfinder.pl -s $line -d Athaliana\_167\_TAIR10.fasta -p table > $line.target.txt; done. Predicted targets for the corresponding set of miRNAs will be displayed in table format:



**Figure 1: Table format of the TargetFinder**

Option 2 using the TargetFinder\_threads.pl, which requires the FASTA file of the miRNAs and the corresponding Target file and can be invoked:

***CODE:*** $perl targetfinder\_threads.pl -f selected\_Arabidopsis\_miRNAs.fasta -d Athaliana\_167\_TAIR10.fasta -o selected\_miRNAs\_TAIR10.txt -c 4 -t 5 -p classic. Predicted targets for the corresponding set of miRNAs will be displayed in classic format:



**Figure 2: Classic format of the TargetFinder.**

***NOTE:*** There are 2 customizable parameters, which user can optimize and customized according to the needs;

1. -c, which represents the prediction score cut-off value and the default value is 4, which works fine with most of the target predictions.

2. -p Visualization of the small RNA-target pairs, which can be classic, gff, json, and table format.

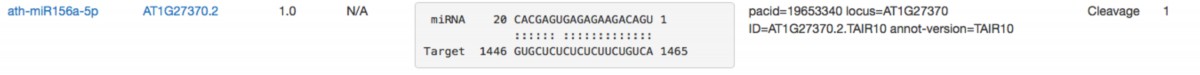
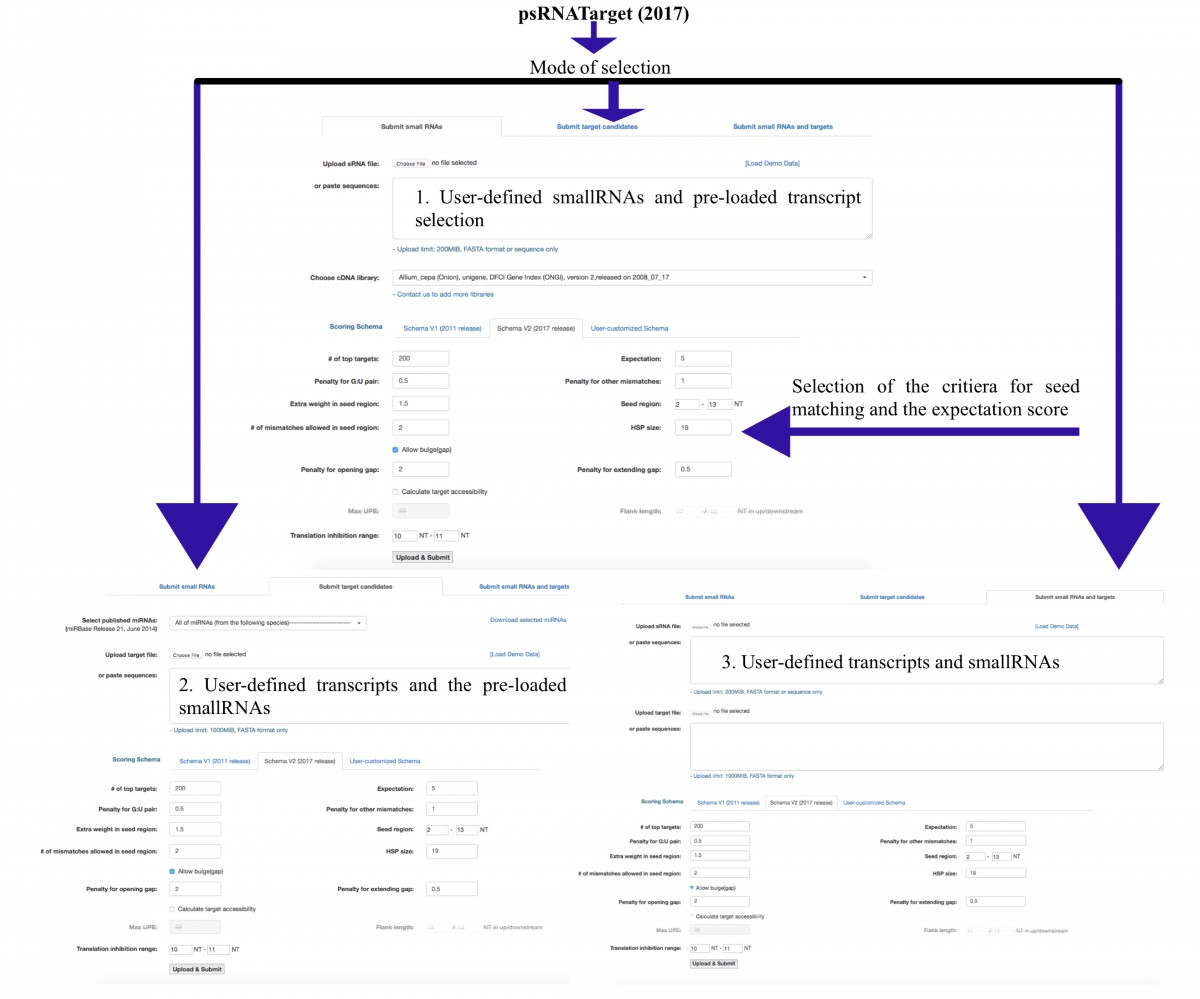
***NOTE:*** For insertion into the database or the alignment files, either the classic format or the table format can be used.

**4. Target predictions using the psRNATarget**

psRNATarget (Dai et al. 2018) is an online web-based miRNA target prediction resource (<http://plantgrn.noble.org/psRNATarget/analysis?function=1>), which allows for the identification of the miRNAs targets based on the application of the pre-defined scoring scheme to define the complementary matching between the miRNAs and the corresponding targets. Additionally, it also allows for the prediction of the miRNA targets based on the target site accessibility and thus increase the sensitivity of target prediction and reduces false positive. The current version of the psRNATarget improves the miRNA target interactions by implementation of the new scoring scheme, which allows for the identification of the canonical and the non-canonical targets at higher recall rates (Dai et al. 2018) by extending the seed region from 2-13 bp with 2 mismatches allowed with in the seed region (Meyers et al. 2008; Axtell, 2013).

psRNATarget offers easy to use graphical user interface, which allows for the prediction of the miRNA targets based on 3 specific patterns (Figure 3): 1. User-defined smallRNAs and pre-loaded transcripts, which allows for the comprehensive search of the miRNA’s targets for the species for which well annotated transcripts are present; 2. User defined transcripts and pre-loaded smallRNAs, which allows for the identification of new targets using the previously defined smallRNAs and 3. User-defined smallRNAs and transcripts, which allows for the identification of the targets in non-model species, thus extending the scope of the psRNATarget to target prediction in non-model species.

Figure 3. Graphical User interface-based prediction of psRNATarget.



***NOTE:*** psRNATarget based target prediction searches can be optimized by considering two parameters: 1. Seed matching and 2. Expectation score, which defines the rate of the false positive.

**5. Target predictions using the TAPIR**

TAPIR (Bonnet et al. 2010) was developed as among the first of the web-based approaches as well as stand-along packages that provides the miRNA target predictions based on the alignment-based searches using the FASTA search module as well as the miRNA-mRNA duplexes. For the predictions of the miRNA targets using the TAPIR, two search modes can be used: 1) a fast search method, which detects the targets interactions based on the minimal free energy (MFE) and 2) Precise search, which involves the RNAHybrid to check for the structure optimization.

***NOTE:*** Prior to the search of the targets using the TAPIR, export the TAPIR directory PATH to the directory, where the installation has been done like

***CODE:*** export PATH="${PATH}:/protocol/chapter/tapir/"

To start the search of the miRNA targets using the FASTA fast search for the miRNAs present in the file: selected\_Arabidopsis\_miRNAs.fasta;

***CODE:*** $./tapir\_fasta --mir\_file selected\_Arabidopsis\_miRNAs.fasta --target\_file Athaliana\_167\_TAIR10.fasta --score 4 --mfe 0.8.

***NOTE:*** For the compilation of the TAPIR, FASTA34 libraries are needed, which can be obtained from (http://fasta.bioch.virginia.edu/fasta\_www2/fasta\_down.shtml) and PATH of the FASTA34 directory should be exported to the TAPIR directory:

***CODE:*** export PATH=/Users/protocol/FASTA34/:$PATH

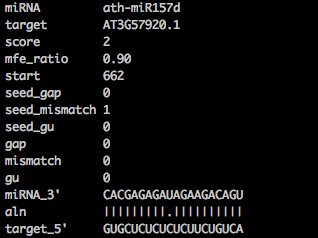


Figure 3. Target prediction output of TAPIR, using ***CODE:*** /tapir\_fasta --mir\_file selected\_Arabidopsis\_miRNAs.fasta --target\_file Athaliana\_167\_TAIR10.fasta --score 4 --mfe 0.8

***NOTE:*** score cut-off in TAPIR is a measure of the number of mismatches, gaps and the number of G:U pairs in case of the miRNA-mRNA duplexes.

**6. Target predictions using the psRobot**

psRobot has been developed to address the high throughput identification of the smallRNAs with stem-loop precursors and also allows for the prediction of the microRNA targets based on the previously defined criteria for the miRNA target predictions (Rhoades et al. 2002; Meyers et al. 2008). The unique features of psRobot allows for the both the transcript based and the degradome based target prediction coupled with the expression datasets and allows for the target site multiplicity and target site conservation-based approach to define the targets for the given set of miRNAs.

***CODE:*** Installation of the psRobot on Linux 4.4.0-134-generic #160-Ubuntu x86\_64.

$tar zxvf psRobot\_v1.2.tar.gz

$cd psRobot\_v1.2

$export PATH=/user/mfold/Bin64:$PATH

$./configure

$make && sudo make install

Following the installation of the psRobot, miRNA target predictions for the set of the miRNAs can be defined as:

***CODE:* Strict mode**

$psRobot\_tar -s selected\_Arabidopsis\_miRNAs.fasta -t Athaliana\_167\_TAIR10.fasta -o target.txt -ts 3.0 -fp 1 -tp 31 -gl 1 -p 4 -gn 0

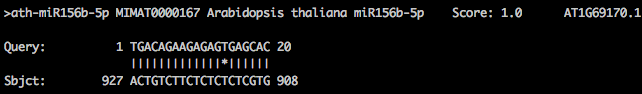
***CODE:* Moderate mode**

$psRobot\_tar -s selected\_Arabidopsis\_miRNAs.fasta -t Athaliana\_167\_TAIR10.fasta -o target.txt -ts 2.5 -fp 2 -tp 17 -gl 30 -p 4 -gn 1

***CODE:* Loose mode**

$psRobot\_tar -s ../selected\_Arabidopsis\_miRNAs.fasta -t ../Athaliana\_167\_TAIR10.fasta -o target.txt -ts 3.0 -fp 2 -tp 17 -gl 1 -p 4 -gn 2

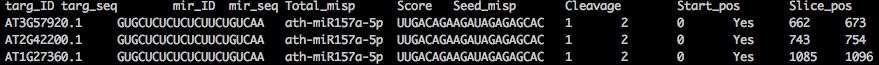
Output of the psRobot is a tab-delimited file showing the alignment of the miRNAs to the corresponding target along with the score value.



**7. Target prediction using the TarHunter and TarHunterL:**

miRNA target predictions using the TarHunter and the TarHunterL (Ma et al. 2018) relies on the cross-species conservation filter, which allows for the identification of the cross-species conserved miRNA targets and also the miRNA targets with non-canonical binding. TarHunterL represents a simplified version of the TarHunter, which predicts the miRNA targets based on the cut-off score and can be invoked as:

***CODE:*** $ perl TarHunterL.pl -q selected\_Arabidopsis\_miRNAs.fasta -b Athaliana\_167\_TAIR10.fasta -o target.txt -T 4 -t 1 -f 4 and the delimited output with the information on the cleavage site and the start and the end coordinates can be visualized using excel.



***NOTE:*** Here -f defines the cut-off score for the target predictions and to keep the score similar across all the target predictions algorithm, a synonymous number of the cut-off threshold 4 has been used.

On the contrary, TarHunter provides several options, which entails the detection of the species-specific miRNAs and also the presence of the conserved miRNAs across the species.

**References**

1. Addo-Quaye C, Miller W, Axtell MJ (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics 25:130-1.
2. Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol. 64:137-59.
3. Bonnet E, He Y, Billiau K, Van de Peer Y (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. Bioinformatics 26:1566-8.
4. Dai X, Zhuang Z, Zhao PX (2011) Computational analysis of miRNA targets in plants: current status and challenges. Brief Bioinform. 12:115-21.
5. Dai X, Zhuang Z, Zhao PX (2018) psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 46(W1): W49-W54.
6. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC (2007) High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. PLoS ONE 2: e219.
7. Fahlgren N., Carrington J.C. (2010) miRNA Target Prediction in Plants. In: Meyers B., Green P. (eds) Plant MicroRNAs. Methods in Molecular Biology (Methods and Protocols), vol 592. Humana Press
8. Li F, Orban R, Baker B (2012) SoMART: a web server for plant miRNA, tasiRNA and target gene analysis. Plant J. 70:891-901.
9. Ma X, Liu C, Gu L, Mo B, Cao X, Chen X (2018) TarHunter, a tool for predicting conserved microRNA targets and target mimics in plants. Bioinformatics 34:1574-1576.
10. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu JK (2008) Criteria for annotation of plant MicroRNAs. Plant Cell. 20:3186-90.
11. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. Cell. 110:513-20.
12. Wu HJ, Ma YK, Chen T, Wang M, Wang XJ (2012) PsRobot: a web-based plant small RNA meta-analysis toolbox. Nucleic Acids Res.40:W22-8.
13. Zhang L, Qin C, Mei J, Chen X, Wu Z, Luo X, Cheng J, Tang X, Hu K and Li SC (2017) Identification of MicroRNA Targets of Capsicum spp. Using MiRTrans—a Trans-Omics Approach. Front. Plant Sci. 8:495