OXFORD

Sequence analysis

miTRATA: a web-based tool for *mi*croRNA *Tr*uncation *a*nd *T*ailing *A*nalysis

Parth Patel[1,2], S. Deepthi Ramachandruni[2], Atul Kakrana[1,2], Mayumi Nakano[3], Blake C. Meyers[2,3,*]

[1]Center for Bioinformatics and Computational Biology and [2]Delaware Biotechnology Institute and [3]Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19711, USA

*To whom correspondence should be addressed

Associate Editor: Ivo Hofacker

Abstract

Summary: We describe miTRATA, the first web-based tool for **mi**croRNA **Tr**uncation **a**nd **T**ailing **A**nalysis—the analysis of 3′ modifications of microRNAs including the loss or gain of nucleotides relative to the canonical sequence. miTRATA is implemented in Python (version 3) and employs parallel processing modules to enhance its scalability when analyzing multiple small RNA (sRNA) sequencing datasets. It utilizes miRBase, currently version 21, as a source of known microRNAs for analysis. miTRATA notifies user(s) via email to download as well as visualize the results online. miTRATA's strengths lie in (i) its biologist-focused web interface, (ii) improved scalability via parallel processing and (iii) its uniqueness as a webtool to perform microRNA truncation and tailing analysis.

Availability and implementation: miTRATA is developed in Python and PHP. It is available as a web-based application from https://wasabi.dbi.udel.edu/~apps/ta/.

Contact: meyers@dbi.udel.edu

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

1 Introduction

Revolutionary changes in next-generation sequencing and its ever-decreasing cost have driven a growing need for bioinformatics tools and algorithms. Molecular analysis of plant microRNAs (miRNAs) has demonstrated their 3′ protection by action of the methyltransferase *HUA ENHANCER1* (*HEN1*), which adds a 2′-O-methyl group to the 3′ terminal ribose of miRNAs and other small RNAs (Zhai *et al.*, 2013). Such 3′ modification protects plant miRNAs from uridylation [addition of uridine (U) nucleotides] (Zhai and Meyers, 2013). We and others have previously demonstrated via sequencing studies and comparative analysis that miRNAs in a *hen1* mutant (and to a much less extent in wild type *Arabidopsis*) demonstrate an array of 3′ modifications, include loss of nucleotides (truncation) and an addition of one to several non-templated nucleotides, predominantly uridines (U) (Li *et al.*, 2005). Previous reports described varying degrees of 3′ modifications that differ among miRNA families yet are conserved for a given miRNA across plant species (Zhai *et al.*, 2013).

With advancements in highly productive forward- and reverse-genetics approaches, additional genes important to 3′ modifications of small RNAs are being characterized. Deep sequencing in these mutants is providing insights into the mechanisms by which these genes function, or at least fail to function in the mutants. For example, analysis of the *hen1/heso1* double mutant revealed a reduction in 3′ uridylation relative to *hen1* (Zhao *et al.*, 2012). Recent genetic analysis identified a *HESO1* ortholog, *URT1*, with a similar but non-redundant role as a nucleotidyl transferase that uridylates miRNAs (Tu *et al.*, 2015). Thus both genetic analysis and sequencing have demonstrated the complexity of 3′ truncation and tailing (Zhai *et al.*, 2013). Interestingly, these 3′ end modifications can alter the nature of interaction with a target, while the pairing of a miRNA with its targets has been proposed to play critical role in the

extent of 3′ modifications (Zhai *et al.*, 2013). Further insights into these miRNA-target interactions and the influence of 3′ modifications could be gained via a new generation of computational tools (Brousse *et al.*, 2014; Kakrana *et al.*, 2014) In analysis of 3′ ends from sRNA sequencing data, computational tools separate the template miRNA from the non-templated additions, while also assessing truncation. Yet, there are no online, web-based tools for these measurements—they're run from a command line. Hence, we aimed to provide biologists with an automated, parallelized, web-based tool for these analyses in plants and animals (miTRATA, for microRNA Truncation and Tailing Analysis).

2 Algorithm

We adapted for web access the 'truncation and tailing' algorithm that we previously developed (Zhai and Meyers, 2013). We re-implemented it in Python (version 3). **Step 1:** miTRATA's web interface collects input parameters such as adapter-cleaned sequence files ('tag count' format, or sequence plus abundance), the genome to be used (aka the 'Bowtie index'), a list of mature miRNA(s), etc., from the user. One of the most substantial changes to the algorithm is the addition of parallel processing to increase the scalability of our webtool especially when dealing with multiple high-throughput sRNA sequencing datasets and large genomes like maize and Norway spruce. **Step 2:** for an individual sRNA sequence file (the 'tag count' file), all non-genome-matched reads are first chopped iteratively, one nucleotide at a time, from their 3′ end, until the remaining 'head' sequence perfectly matches to the genome. This way, we divide individual non-genome-matched reads into two parts: the 5′ genome-matched component (5GMC) or head, and a 3′ tail. The modified workflow can be seen in the Supplementary Figure S1. Previously, this process of finding the 5GMC and tail was performed serially for each tag count file and was time consuming; parallel processing the individual tag count files finds the 5GMC and tail simultaneously. **Step 3:** the 5GMC is aligned to known, mature miRNAs from miRBase (Kozomara and Griffiths-Jones, 2014) to determine the miRNA and the extent of truncation and tailing. **Step 4:** in this last step, images are generated to represent the truncation and tailing results for each miRNA. Finally, miTRATA notifies the user, sending a link to download the results or view them online.
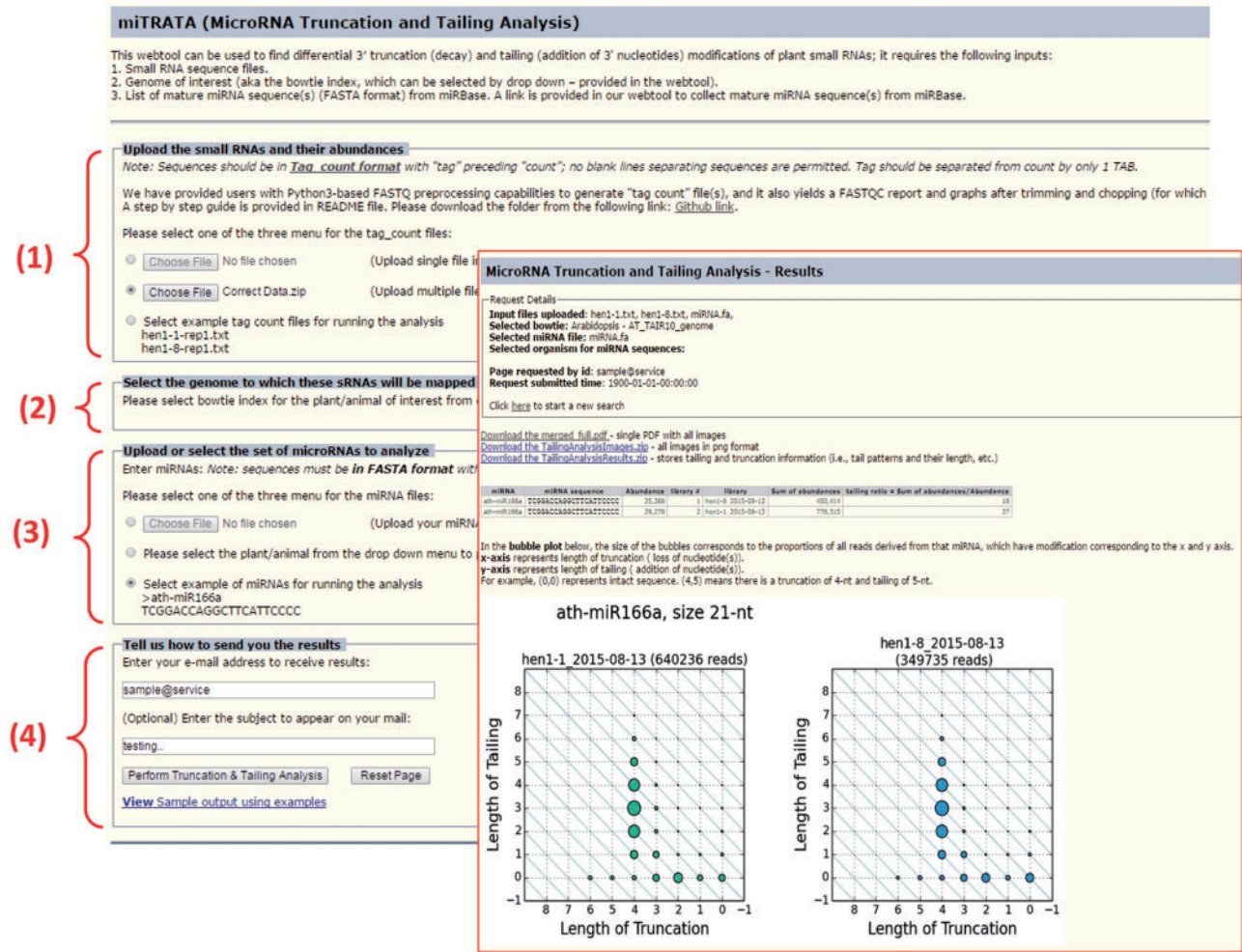
Fig. 1. The miTRATA web tool. The user enters: (**1**) first, the input data or parameters such as the small RNA sequence files; (**2**) the genome of interest (the Bowtie index to be used); (**3**) the list of miRNAs to analyze; (**4**) their contact information. When the user selects 'Perform Truncation and Tailing Analysis', the files and settings are uploaded and the analysis is performed. The results page (the inset red box) displays the output, including a summary of truncation and tailing results in a tabular form, the graphed data, and links to download the results

3 Web tool design

miTRATA is available online at **https://wasabi.dbi.udel.edu/~apps/ta/**. We designed this webtool so that it has necessary settings to aid advanced as well as intermediate users. This webtool requires the following inputs:

1. **Small RNA sequence file(s):** User can choose to upload small RNAs in 'tag count' formatted files in two different ways: (i) a single file in '.txt' or compressed '.gz' format, (ii) multiple files in a compressed archive ('.zip' or '.tar.gz'). Alternatively, the user can also select an example file to run the analysis and see expected results. We have also provided users with FASTQ preprocessing capabilities to generate these 'tag count' files (see Section 4); this step will remove adapters and poor quality sequences.
2. **Genome sequence:** The Bowtie index of the genome of interest can be selected from a drop-down menu of pre-loaded genomes. The default genome is set to *Arabidopsis* for the example.
3. **Mature miRNA sequence(s):** The user has two options to select miRNAs for comparison: (i) an advanced user can upload a set of mature miRNAs in FASTA format. (ii) Mature miRNA sequences from miRBase (currently version 21) can be selected from a drop-down menu listing chosen species present in miRBase. Alternatively, the user can choose to run the analysis with an example miRNA.

When the user selects 'Perform Truncation and Tailing Analysis', all the settings and required files are uploaded; the user's email is required for notification once the analysis is complete. The results are displayed on a webpage for visualization, and links allow a download of the results (Fig. 1). Download links on a webpage contain three folders, as follows: (i) 'merged_full.pdf'—a single pdf with all images. (ii) TailingAnalysisImages.zip—all images in .png format. (iii) TailingAnalysisResults.zip—stores tailing and truncation information (i.e. tail patterns and their lengths).

4 Discussion

New algorithms are emerging in the study of 3′ modifications of sRNAs. For example, EnD-Seq (Welch *et al.*, 2015) is an experimental protocol for high-throughput sequencing of RNAs in which the 3′ modifications of the captured RNA is sequenced separately from the 5′ end. This technique is partnered with an algorithm, AppEnD (Welch *et al.*, 2015) that describes these 3′ end modifications from the EnD-Seq-derived data. Our approach differs from that of Welch *et al.* (2015) in that our algorithm uses standard sRNA sequencing data, it characterizes the extent of 3′ truncation (a common 3′ modification of sRNAs), and it generates a graphical output for the results of each miRNA or sRNA of interest. Another tool may be a better comparison to ours: Tailor (Chou *et al.*, 2015) is specialized for sRNA 3′ analyses, and it characterizes both biological alterations (such as RNA editing events) and technical confounding factors (such as sequencing errors). Although poor quality nucleotides will be removed by FASTQ processing, our approach does not characterize sequence polymorphisms generated by RNA editing; however,

we have emphasized truncation, as this 3′ modification is a common source of variation in sRNAs, and a unique graphical output. Finally, to make our tool less prone to the inclusion of sequencing errors and make more easily accessible, we provided users with a standalone and Python-based FASTQ processing script to produce the 'tag count' formatted files required as input to our webtool. The processing script performs trimming and chopping, taking the raw FASTQ file(s) (single or paired end) and a set of user-defined parameters that include adapter sequences that may vary from default Illumina adapters. The parameter file also determines whether the FASTQ processing yields a FASTQC report, and whether it generates the graphs after trimming and chopping (for which the genome sequence must be provided). These details are described in a README file. This processing script is available from the web page or can be downloaded from (https://github.com/atulkakrana/helper.github.git). In summary, we have built a biologist-friendly web tool that provides easy access to our previously-described algorithm for 3′ analysis of sRNAs (Zhai *et al.*, 2013), generating reports detailing the complexity of truncation and tailing modifications; this tool should be particularly useful for characterizing the impact of mutations in the pathways for sRNA stabilization or turnover.

References

Brousse,C. *et al.* (2014) A non-canonical plant microRNA target site. *Nucleic Acids Res.*, **42**, 5270–5279.

Chou,M.-T. *et al.* (2015) Tailor: a computational framework for detecting non-templated tailing of small silencing RNAs. *Nucleic Acids Res.*, **43**, e63.

Kakrana,A. *et al.* (2014) sPARTA: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets,including new miRNA target-identification software. *Nucleic Acids Res.*, **43**, e139.

Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, 68–73.

Li,J. *et al.* (2005) Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. *Curr. Biol.*, **15**, 1501–1507.

Tu,B. *et al.* (2015) Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in Arabidopsis. *PLoS Genet.*, **11**, e1005119.

Welch,J.D. *et al.* (2015) EnD-Seq and AppEnD: sequencing 3′ ends to identify nontemplated tails and degradation intermediates. *RNA*, **21**, 1375–1389.

Zhai,J. and Meyers,B. (2013) Deep sequencing from *hen1* mutants to identify small RNA 3′ modifications. *Cold Spring Harb Symp Quant Biol.*, **77**, 213–219.

Zhai,J. *et al.* (2013). Plant microRNAs display differential 3′ truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. *Plant Cell*, **25**, 2417–2428.

Zhao,Y. *et al.* (2012) The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.*, **22**, 689–694.