**tasiRNAdb: a database of ta-siRNA regulatory pathways**

Changqing Zhang¹*, Guangping Li², Shinong Zhu¹, Shuo Zhang³ and Jinggui Fang³

¹College of Horticulture, Jinling Institute of Technology, Nanjing 210038, China, ²College of Forest Resources and Environment, Nanjing Forestry University, Nanjing 210037, China and ³College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China

Associate Editor: Ivo Hofacker

# ABSTRACT

**Summary:** In plants, many trans-acting small interfering RNA (ta-siRNA) regulatory pathways have been identified as significant components of the gene networks involved in development, metabolism, responses to biotic and abiotic stresses and DNA methylation at the TAS locus. To obtain a more comprehensive understanding on the nature of ta-siRNA regulatory pathways, we developed a freely accessible resource, tasiRNAdb, to serve as a repository for the sequences of ta-siRNA regulatory pathway-related microRNAs, TASs, ta-siRNAs and ta-siRNA targets, and for the cascading relations among them. With 583 pathways from 18 species, tasiRNAdb is the largest resource for known ta-siRNA regulatory pathways currently available. tasiRNAdb also provides a tool named TasExpAnalysis that was developed to map user-submitted small RNA and degradome libraries to a stored/input TAS and to perform sRNA phasing analysis and TAS cleavage analysis.

**Availability:** The database of plant ta-siRNA regulatory pathways is available at http://bioinfo.jit.edu.cn/tasiRNADatabase/

**Contact:** zhang_chq2002@sohu.com

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

Received on September 9, 2013; revised on November 30, 2013; accepted on December 19, 2013

# 1 INTRODUCTION

In plants a classical "phase-initiator→TAS transcript→ta-siRNA→target gene" cascade model for the ta-siRNA regulatory function has been well determined (Allen et al., 2005; Gasculi et al., 2005; Yoshikawa et al., 2005). Briefly, a phase-initiator (usually a miRNA, but sometimes a siRNA) cleaves the transcript of ta-siRNAs-producing locus (TAS) by binding argonaute (AGO) proteins 1/7. Then, RDR6 (RNA-dependent RNA-polymerase 6)-dependent conversion of the resulting cleaved fragments into dsRNA and subsequent cleavage by dicer-like 4 (DCL4) at every ~21 nucleotides (nt) relative to the phase-initiator cleavage site generates a ta-siRNA. The ta-siRNA further cleaves its target RNA to regulate gene expression by binding AGO1 proteins. The ta-siRNA regulatory pathways have been identified as important elements in the gene networks involved in development, metabolism, response to biotic and abiotic stresses and DNA methylation at the TAS locus (Allen et al., 2005; Rajagopalan et al., 2006; Wu et al., 2012). Recently, the cascade system has also been used as a valuable tool in functional genomics studies (Feligges and Weigel, 2009).

*To whom correspondence should be addressed.

2 DESCRIPTION OF THE DATABASE

2.1 Data acquisition

We obtained a list of publications from the NCBI PubMed database using the keywords ""ta-sirnas"[All Fields] OR ""trans-acting sirnas"[All Fields] OR ""tasiRNA"[All Fields] OR ""transacting sirna"[All Fields]". We manually extracted each ta-siRNA regulatory pathway from each publication, and added these ta-siRNA regulatory pathways to tasiRNAdb.

2.2 The tasiRNAdb interface and data processing

The tasiRNAdb currently contains 583 ta-siRNA regulatory pathways from 18 plant species (Supplementary Table S1), which cover 120 pairs of phased initiators and their targeted sites, 77 TASs, 457 pairs of ta-siRNAs and their targeted sites and 296 ta-siRNAs target genes. ‘Home’ page and ‘tasiRNAdb Table’ page are two entry points to these pathways.

The database is implemented on a Unix platform. All user interfaces and scripts are written in CGI-Perl, JavaScript and HTML and are integrated with an Apache web server. In the present version, users can submit new ta-siRNA regulatory pathway(s) to tasiRNAdb via ‘Submit new pathway’ page, and view the submitted content via ‘Check submitted new pathway’ page. After the submitted pathway passes administrator check in tasiRNAdb, it will be processed by a background script and automatically added to tasiRNAdb database. All the archived ta-siRNA regulatory pathways can be displayed on ‘tasiRNAdb table’ page, where the...
information for each pathway is shown in six sections: species, phase initiator, TAS, ta-siRNA, targeted gene and reference(s), and the basic information for each section is further hyperlinked to a dynamic page (Supplementary Fig. S1). To only show the data that meet the criteria that users specify, the contents of the table can also be sorted and/or filtered by species, phase initiator, TAS, tasiRNA, targeted gene, and references. 'tasiRNAdbSearch' and 'TasExpAnalysis' are two online tools developed to use the archived data. They can respectively search the tasiRNAdb and evaluate whether a TAS of interest is transcribed and cleaved to produce ta-siRNA in user-sequenced sRNA dataset.

2.3 The TasExpAnalysis tool

For researchers who have generated sRNAs sequence libraries, a major challenge is to evaluate whether an interesting TAS among the sRNA reads is cleaved by a phased-initiator and sliced by DCL4 at every ~21 nt relative to the phase-initiator cleavage site to produce functional ta-siRNAs. To meet this need, we developed the TasExpAnalysis tool (Supplementary Fig. S2) and integrated it into tasiRNAdb.

TasExpAnalysis maps the sRNA and degradome reads submitted by the user to a selected/input TAS, and checks whether the 2’ end of the RNA degradome fragment overlaps the TAS cleavage start position. When an overlap is detected, TasExpAnalysis searches the sRNA dataset to find a phased initiator. Next, the numbers of phased/non-phased positions with a sRNA hit on the TAS are counted (phased positions refer to those arranged in ~21-nt increments relative to the cleavage start position; non-phased positions are all other positions that are not phased positions or the start position). The P value of the detected configuration is calculated based on a random hypergeometric distribution using the following equation (Zhang et al., 2012, 2013):

\[
P(k_1) = \frac{\left( L \times 2 - 1 \right) - \left( \frac{k_1}{k_2} \times 2 - 1 \right) - \left( \frac{k_1}{k_3} \times 2 - 1 \right) \times (2 \times s)}{k_2 \left( L \times 2 - 1 \right) - \left( \frac{k_1}{k_2} \times 2 - 1 \right) \times s} \]

where \( L \) is the length of TAS; \( k_1 \) is the number of phased positions having sRNA hits; \( k_2 \) is the number of non-phased positions having sRNA hits; and \( s \) is the maximum allowed offset from the phase position.

The TasExpAnalysis counts the number of unique sRNA and sRNA reads on different phasing registers and the number of different length reads in the user-submitted sRNA dataset. All the detected results are reported in a result page (Supplementary Fig. S2).

3 DISCUSSION

tasiRNAdb is an integrated database comprising a large number of known ta-siRNA cascade pathways. Our aim in developing tasiRNAdb was to collect known plant ta-siRNA cascade pathways and integrate them in such a way that the data could be freely accessed and mined by users. tasiRNAdb is different from other similar resources because it includes all plant species for which such information is available. For example, ASRP currently collects Arabidopsis TAS genes and aims mainly to provide a public resource for genome-wide sRNA data from Arabidopsis (Backman et al., 2008). MPSS aims to provide a signature-based transcriptional resource for analyses of mRNA and sRNA (Nakano et al., 2006) and currently contains Arabidopsis TASs and rice TAs. CSRDB is a database of rice and maize sRNAs, which aims to identify microRNAs and other functional sRNAs, map microRNA transcription units, during development and as a function of biotic and abiotic stress (Johnson et al., 2007). Our tasiRNAdb currently contains 583 known ta-siRNA cascade pathways from 18 plant species, comprising 120 pairs of phased initiators and their targeted sites, 77 ta-siRNA-generating loci, 457 pairs of ta-siRNAs and their targeted sites and 296 ta-siRNAs target genes. Consequently, tasiRNAdb provides more ta-siRNAs cascade pathway information than ASRP, MPSS or CSRDB. Moreover, web-based tools for evaluating whether a known/input TAS is cleaved by phased-initiator and sliced by DCL4 at 21-nt intervals to produce functional ta-siRNAs in user-sequenced sRNA datasets are unavailable. We have developed TasExpAnalysis to meet this need. We evaluated the performance of TasExpAnalysis with an RDR sRNA MPSS library (Lu et al., 2006) and default parameters. All six atTASs previously reported in the sample, including atTAS1a, atTAS1b, atTAS1c, atTAS2, atTAS3a and atTAS3b, were successfully identified by TasExpAnalysis (Supplementary Table S2). They all have a significant P value of <0.005. In addition, the other two known atTASs stored in tasiRNAdb, including atTAS3b and atTAS3c, were also identified by TasExpAnalysis. TasExpAnalysis could be a useful tool for characterizing sRNAs from deep sequencing data.

Funding: This work was supported by the National Science Foundation of China (31171273).

Conflict of Interest: none declared.

REFERENCES