BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 20 2012, pages 2561–2568
doi:10.1093/bioinformatics/bts500

Genome analysis

Advance Access publication August 22, 2012

Identification of *trans*-acting siRNAs and their regulatory cascades in grapevine

Changqing Zhang[1,*], Guangping Li[2], Jin Wang[3] and Jinggui Fang[4]

[1]College of Horticulture, Jinling Institute of Technology, Nanjing 210038, China, [2]College of Forest Resources and Environment, Nanjing Forestry University, Nanjing 210037, China, [3]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Science, Nanjing University, Nanjing 210093, China and [4]College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China

Associate editor: Ivo Hofacker

ABSTRACT

Motivation: *Trans*-acting small interfering RNAs (ta-siRNAs) play an essential role in the regulation of plant gene expression, but relevant reports are still limited. Bioinformatic analyses indicate that many ta-siRNA-producing loci (TASs) are present in plants, implying the existence of as yet undiscovered ta-siRNAs and related regulatory pathways. To expand our knowledge of these plant gene regulators, we applied high-throughput computational and experimental methods to grapevine (*Vitis vinifera* L.).

Results: Based on bioinformatic predictions, we identified 49 TASs from 49 055 small RNA clusters. Using RNA degradome analysis, we experimentally validated 5 TASs, 22 ta-siRNAs and 37 ta-siRNA targets. The *cis*-activities of ta-siRNAs were also confirmed, which suggested an inactive mechanism of TAS transcription, and a produced mechanism of multiple forms of small RNA from same TAS. We examined the conservation of newly identified ta-siRNA regulatory cascades and found that while the cascade related to vviTAS3 was conserved, cascades related to vviTAS7, vviTAS8, vviTAS9 and vviTAS10 were grape-specific. These results broaden the known scope of ta-siRNA regulation.

Contact:zhang_chq2002@sohu.com

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

Received on April 17, 2012; revised on August 2, 2012; accepted on August 6, 2012

1 INTRODUCTION

Endogenous small RNAs (sRNAs) are key post-transcriptional and transcriptional regulators of plant gene expression (Molnar *et al.*, 2007). They include microRNAs (miRNAs), heterochromatic small interfering RNAs (siRNAs), natural antisense siRNAs (nat-siRNAs) and *trans*-acting siRNAs (ta-siRNAs).

miRNAs, which are generated by cleavage of single-stranded RNA stem–loop structures by DICER-Like 1, play pivotal roles in development and stress responses by targeting the transcripts of transcription factors that mediate the transition between various developmental stages or regulate responses to biotic or abiotic stimulation (Meng *et al.*, 2011). Heterochromatic siRNAs trigger DNA methylation and histone modification and are involved in heterochromatin formation of repeated genomic regions and the silencing of other genomic regions, including the control of transposon movement (Katiyar-Agarwal and Jin, 2010). nat-siRNAs consist of two types of siRNAs, i.e. 24-nt nat-siRNAs and 21-nt nat-siRNAs (Urano *et al.*, 2010). They are processed from pairs of natural *cis*-antisense transcripts by cleavage with DICER-Like 2 and DICER-Like 1, respectively. The 24-nt nat-siRNAs play a key role in environmental stress response and in reproduction by targeting unlinked transcripts and guiding cleavage to produce 21-nt nat-siRNAs whose function is still not clear.

ta-siRNAs play an essential role in regulating plant development, metabolism and responses to biotic and abiotic stresses (Chen *et al.*, 2007; Hu *et al.*, 2011). They are produced from ta-siRNA-producing locus (TAS) genes (Allen *et al.*, 2005). A single-stranded RNA is first transcribed from a ta-siRNA locus and then cleaved by a phase initiator (a miRNA or, in some cases, a ta-siRNA). RDR6-dependent conversion of the resulting fragments into double-stranded RNA, and subsequent cleavage by DICER-Like 4, generates 21-nt phased ta-siRNAs. These ta-siRNAs then combine with argonaute proteins to form RISC complexes and cleave targeted mRNAs.

To date, several 'miRNA → ta-siRNA → target gene' cleavage cascades have been identified as important elements of gene regulatory networks. TAS1a, TAS1b, TAS1c and TAS2 are four ta-siRNA loci targeted by miR173, and their 3′-cleavage products produce six ta-siRNAs that target at least two transcripts with unknown function and two transcripts of pentatricopeptide repeat genes (Allen *et al.*, 2005). TAS3, another ta-siRNA locus, is targeted by miRNA390. Its 5′-cleavage product produces two ta-siRNAs that target transcripts of auxin response factors ARF2, ARF3 and ARF4. A fourth ta-siRNA locus, TAS4, is targeted by miRNA828 (Rajagopalan *et al.*, 2006). Its 3′-cleavage product produces one ta-siRNA that targets three MYB gene transcripts. Another TAS gene, also named TAS4, has been identified in moss. It produces ta-siRNAs targeting a putative transcription factor similar to AP2/EREBP, suggesting a function in plant development (Talmor-Neiman *et al.*, 2006).

It has been reported that ta-siRNAs from TAS3 regulate leaf morphology and juvenile-to-adult vegetative-stage transitions; ta-siRNAs from TAS4 regulate anthocyanin, flavonoid and phenylpropanoid biosynthesis and may also be involved in regulating leaf senescence (Allen *et al.*, 2005; Hsieh *et al.*, 2009; Luo *et al.*, 2012). Computational analysis indicates that many TASs

*To whom correspondence should be addressed.

exist in plants (Chen *et al.*, 2007; Lu *et al.*, 2006), suggesting they may play a more fundamental role than currently understood. Identification and functional characterization of new ta-siRNAs is therefore needed. Grapevine (*Vitis vinifera* L.) is a model system for studying fruit-bearing agricultural crop plants and, based on fruit consumption and wine yield, is also of great economic importance (Jaillon *et al.*, 2007). Many biological processes, including those involving ta-siRNAs generated from TAS3 and TAS4, greatly affect grape yield and quality. The identification of grapevine ta-siRNAs and their regulatory cascades is thus important in efforts to improve grape cultivation and breeding. Recent publication of deep sequencing sRNA libraries, especially a degradome library generated from cleaved mRNA fragments, opens the door to the identification of grapevine ta-siRNAs and their cascades (Pantaleo *et al.*, 2010; Wang *et al.*, 2011).

2 METHODS

2.1 Sources of sRNA libraries

In this study, we used a total of 35 sRNA datasets, which included a degradome library and 34 sRNA libraries. The degradome library was used for evaluating cleaved positions of phase initiators and ta-siRNAs, while the 34 sRNA libraries were used to ascertain their conservation. For various reasons, such as the specific plant tissue, mutant line or sequencing technology used, a given sRNA might be absent from a particular library; consequently, we used as many different sRNA libraries as possible to minimize the chance of incorrectly concluding an absent sRNA was not conserved in a given species. All datasets were downloaded from the Gene Expression Omnibus (GEO, see text for GEO accession numbers) except for three obtained from the Tomato Functional Genomics Database. The genomic sequence and annotation for grapevine were downloaded from NCBI (National Center for Biotechnology Information, updated April 7, 2009).

2.2 Identification of sRNA clusters

A genomic region was considered to be an sRNA cluster when it contained at least three sRNA hits with a maximum separation distance of 100 nt (Johnson *et al.*, 2009). Each sRNA cluster was classified according to the sRNA length that was most predominant. In other words, if at least 50% of the sRNAs were 21 nt long, it was categorized as a 21-nt cluster, and similarly, if at least 50% were 24 nt long, it was grouped as a 24-nt cluster. If neither 21- nor 24-nt sRNAs represented at least 50% of the identified sRNAs, the cluster was classified differently.

BLASTn from the NCBI standalone BLAST (Basic Local Alignment Search Tool) package was used to search for sRNA hits in the grapevine genomic sequence. Parameters were set to standard settings with no filtering, and a perfect match was required.

2.3 Identification of grapevine TASs and their phase initiators

To identify TAS loci, we modified and used a method based on Chen et al. (2007) and Dai and Zhao (2008). First, assuming that each 5′-end of an sRNA in an sRNA cluster was the cleavage start site, the number of phased/non-phased positions was counted on either side of this start position (phased positions refer to those arranged in 21-nt increments relative to the cleavage start position; non-phased positions are all those other than phased positions and the start position). Second, the *P* value of each detected configuration was calculated on the basis of a random hypergeometric distribution (Zhang *et al.*, 2011) using the following equation:

$$P(k_1) = \sum_{x=k_1}^{\min((k_1+k_2),\ (\frac{L}{21}\times 2-1))} \frac{\binom{L\times 2-1-(\frac{L}{21}\times 2-1)}{k_2}\binom{\frac{L}{21}\times 2-1}{k_1}}{\binom{L\times 2-1}{k_1+k_2}},$$

where *L*, which should be a multiple of 21, is the length of the detected configuration, K_1 is the number of phased positions having sRNA hits and K_2 is the number of non-phased positions having sRNA hits. Finally, detected configurations with $P<0.001$, proposed by Chen et al. (2007) and corresponding to the criterion that a cluster with fewer than two phased sRNAs was <5% likely to be determined true in our preliminary analysis, were identified as TASs.

For each identified TAS, belonging to 21-mer clusters, 24-mer clusters or other clusters, we searched for potential phase initiators in the sRNA library. First, we retrieved the genomic region containing the TAS and an additional 100 nt either to its left or right, whichever was on the same side as the cleavage start position. We then searched for complementary regions between the sRNA and the retrieved genomic sequence using psRNATarget (Dai and Zhao, 2011) with default parameters. Potential phase initiators were identified using the criteria of a 10th sRNA nucleotide position corresponding to the TAS cleavage start position. Overlapping loci with same phase were combined and the one with the best *P* value was chosen as representative. As a final step, we validated the TAS cleavage start position by mapping an RNA degradome with GEO accession number GSE18406 to the grapevine genome using BLASTn with standard settings and no filtering. If the 5′-end of any RNA degradome fragment overlapped the TAS cleavage start position, the start position was considered to be validated.

2.4 Identification of ta-siRNA targets

Prediction of ta-siRNA targets was carried out by searching for complementary regions between ta-siRNA and grapevine transcripts using BLASTn with standard settings and no filtering, and allowing a maximum of four mismatched base pairs. The validation of the ta-siRNA cleavage site was performed using the RNA degradome, with the requirement that it corresponded to ta-siRNA position 10.

2.5 Conservation of TASs, phase initiators and ta-siRNAs

TAS conservation was detected by performing a megablast search against NCBI standard sets and selected nucleotide (nr/nt) databases. The top 100 hits were used for TAS conservation analysis. Conservation of phase initiators and ta-siRNAs was ascertained from the number of sequences in 34 sRNA libraries.

3 RESULTS

3.1 Detection of sRNA clusters

An sRNA cluster is a genomic region mapped by many sRNAs where potential TASs may be present. In this study, we mapped an sRNA library from grapevine stems and leaves (GEO accession number GSM458927) to a grapevine genomic sequence using BLAST, searched for perfect matches and identified the sRNA clusters.

Using the criterion that an sRNA cluster should contain at least three sRNAs with a maximum separation distance of 100 nt (Johnson *et al.*, 2009), 49 055 sRNA clusters were identified (Table 1). There were 2373 21-nt clusters primarily made up of 21-nt sRNAs and representing approximately 5% of the total number, and 25 917 24-nt clusters with 24-nt sRNAs in the

Table 1. The number of sRNA clusters overlapped to gene or exon regions in different grapevine tissues

	Stems and leaves (%)	Tendrils (%)	Inflorescences (%)	Berries (%)
21-nt clusters				
Total	2373 (5)	2671 (4)	2076 (2)	2054 (3)
Overlap to gene	1000 (42)	1166 (44)	929 (45)	895 (44)
Overlap to exon	674 (67)	944 (81)	768 (83)	696 (78)
24-nt clusters				
Total	25 917 (53)	45 418 (61)	80 543 (74)	50 244 (67)
Overlap to gene	4103 (16)	7631 (17)	11 710 (15)	7648 (15)
Overlap to exon	356 (9)	601 (8)	1949 (17)	871 (11)
Other clusters				
Total	20 765 (42)	26 919 (36)	26 162 (24)	23 206 (31)
Overlap to gene	4982 (24)	5038 (19)	5062 (19)	4891 (21)
Overlap to exon	430 (9)	813 (16)	853 (17)	598 (12)
Total	49 055	75 008	108 781	75 504

majority, which made up about 53% of the total. Cluster size ranged from 20 to 3070 nt, with an average of 139.4 nt, and the number of sRNAs in a given cluster ranged from 3 to 417, with an average of 9.6.

When we counted the number of 21- and 24-nt clusters that overlapped with annotated gene loci and exons, we found that 42% of the 21-nt clusters were associated with annotated gene regions, with 67% of these further associated with exon regions. For the 24-nt clusters, in contrast, only 16% overlapped with annotated gene regions, with 9% also overlapping with the exon regions. The remaining clusters (i.e. all those except the 21- and 24-nt clusters) had a larger gene overlap compared with the 24-nt clusters, and a smaller gene overlap compared with 21-nt clusters. To further examine this phenomenon, we applied the same method and criteria to sRNA libraries from grapevine tendrils (GEO accession number GSM458928), inflorescences (GSM458929) and berries (GSM458930), with similar outcomes (Table 1). These results indicate that 21-nt clusters tend to be biased toward gene loci, especially exon regions, whereas 24-nt clusters are generally absent from those regions.

3.2 Identification of grapevine TASs

A strategy for identifying TASs based on the detection of sRNAs arranged in 21-nt increments and the evaluation of statistical significance (P value) has been proposed by several authors (Chen *et al.*, 2007; Dai and Zhao, 2008). Using this approach, three previously known TASs were successfully detected in Arabidopsis. Although this strategy has proven successful, it is limited by a 231-bp TAS length requirement. In this article, we present an improved method allowing for a variable TAS length.

In this study, we used a cutoff of $P < 0.001$ to predict potential TASs. We also predicted potential phase initiators, based on an application of psRNATarget with default parameters and an assumption that the 10th nucleotide position on the sRNA serving as the phase initiator corresponds to the cleavage start position of its targeted TAS. Using this approach, 49 TASs and their corresponding phase initiators were identified (Supplementary Table 1). Their lengths ranged from 42 to 1281 bp, with P values varying from 9.5×10^{-4} to 1.2×10^{-10}. When we analyzed whether the TASs are overlapped with coding or non-coding regions, we found 22 were entirely contained in coding regions, 1 was partly contained and 29 had no overlap with coding regions, suggesting that TAS distribution was not obviously correlated to coding or non-coding regions. Using an RNA degradome dataset (GSE18406) produced from grape leaves and having a 5′-end that can be used to determine the cleavage site of long RNAs, including mRNAs and other cDNAs, we subsequently confirmed five vviTASs (Table 2).

Of these five TASs, vviTAS3 is located in 3′UTR of grapevine LOC100244732 gene and is similar to the sequence that has been reported in Arabidopsis (Allen *et al.*, 2005). On vviTAS3, nine sRNAs were arranged in 21-nt increments (Fig. 1A) and 62 sRNAs mapped to the middle of the 21-nt interval ($P = 9.18 \times 10^{-4}$). When we validated the cleavage sites based on the RNA degradome library, we found 51 distinct degradation fragments that mapped to the vviTAS3 transcript and 3 whose 5′-ends overlapped with the phased positions (i.e. positions arranged in 21-nt increments relative to the cleavage start position); 48 fragments, representing 42 non-phased positions, overlapped with non-phased positions (i.e. those other than phased positions and the start position). A previously reported miRNA, miRNA390, is known to target TAS3 and direct the production of ta-siRNAs in Arabidopsis (Allen *et al.*, 2005). We also found this miRNA in the sRNA leaf library, and the LOC100244732 cleavage start position corresponds to the 10th nucleotide position of vvimiRNA390. When the vvimiRNA390/target-transcript pair was compared with that in Arabidopsis, a completely identical pair was found, showing that the interaction of miRNA390 and TAS3 is conserved in both grapevine and Arabidopsis. Four additional new sRNAs, designated as vviTAS3-primoRNA2, vviTAS3-primoRNA3, vviTAS3-primoRNA4 and vviTAS3-primoRNA5 in our study, were also found. They are complementary to the vvimiRNA390 target site with no more than two mismatches; their 10th positions overlap the start position used to determine phased position, indicating their function as siRNAs. Two distinct degradation fragments also overlapped on their 5′-ends to the cleavage start position, validating the cleavage of the phase initiators.

Another TAS, designated as vviTAS7 in this study, was novel. vviTAS7 has a length of 231 nt and like vviTAS3 was also located in a genomic repeat region. A total of 75 distinct sRNAs existed for vviTAS7; 14 were in the 21-nt increment phase and 8 phased sRNAs consisted of four pairs of perfect complementary duplexes (Fig. 1B). The P value for the observed sRNA arrangement on vviTAS7 was 1.36×10^{-7}. When we validated the cleaved phase positions of vviTAS7, we found 62 distinct RNA degradation fragments that mapped to it. Of these, five, representing four-phased positions, overlapped on their 5′-ends with the phased positions; 57, representing 45 non-phased positions, overlapped the non-phased positions. When we attempted to identify the potential phase initiators of vviTAS7, we found that the small RNAs we have designated as vvimiRNA828a and vviTAS7-primoRNA2 are complementary to the phase-initiator target sites and can direct the cleavage at its 10th nucleotide position. This is supported by the fact that

Table 2. vviTASs and their phase initiator(s) validated by RNA degradome analysis

vviTASs			Phase-initiators	
Name	Coordinate	*P*	Name	Sequence
vviTAS3	chr5:359214..359381(−)	9.18×10^{-4}	vvimiRNA390	AAGCUCAGGAGGGAUAGCGCC
			vviTAS3-primoRNA2	UAGCUCAGGAGGGAUAGACAA
			vviTAS3-primoRNA3	CAGCUCAGGAGGGAUAGGCAA
			vviTAS3-primoRNA4	UAGCUCAGGCGGGAUAGACAAG
			vviTAS3-primoRNA5	AAGCUCAGGAGGGAUAGCCCA
vviTAS7	chr14:21607855..21608085(+)	1.36×10^{-7}	vvimiRNA828a	UCUUGCUCAAAUGAGUAUUCCA
			vviTAS7-primoRNA2	UCUUGCUCAUAUGAGUAUUCC
vviTAS8	chr9:3017768..3017977(+)	8.60×10^{-4}	vviTAS8-primoRNA1	UCCGAUGUCAAACACUCUACU
			vviTAS8-primoRNA2	UCUGAUGUCAAACACUCCACU
vviTAS9	chr5:22564335..22564691(+)	2.26×10^{-4}	vviTAS9-primoRNA	AUGGAAGAUAGGGAAUGUAAA
vviTAS10	chr19:5718901..5719446(−)	1.22×10^{-10}	vviTAS10-primoRNA1	UAAUAUGGAGGACUGUGUUCCU
			vviTAS10-primoRNA2	UAAUAUGGAGGACUGUGUCCUU
			vviTAS10-primoRNA3	CAAUGUGAAGAGCUGUCUUUU
			vviTAS10-primoRNA4	UAAUAUGGAGGACUGUGUUCU
			vviTAS10-primoRNA5	UAAUAUGGAGGACUGUGUUCUU
			vviTAS10-primoRNA6	UAAUAUGUAGGACUGUGUUCU

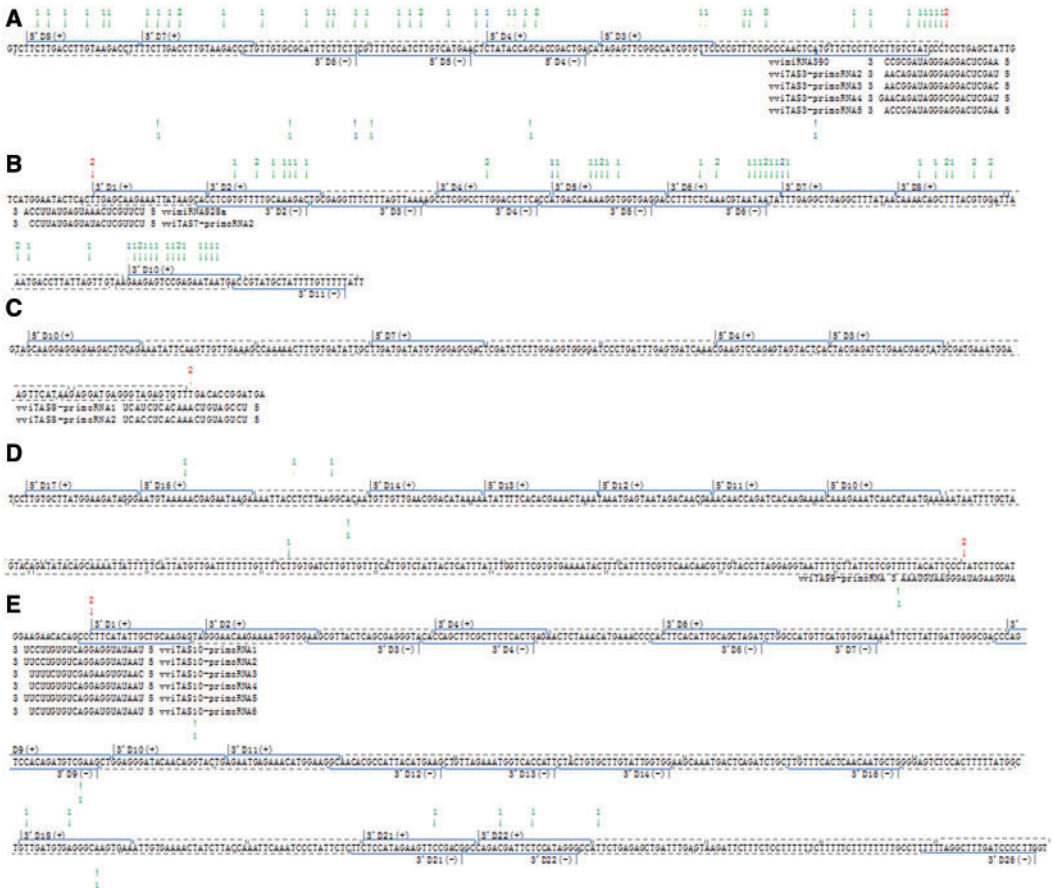

Fig. 1. Distributions of phased regions, RNA degradation fragments and targeting phase-initiator(s) on vviTAS3 (**A**), vviTAS7 (**B**), vviTAS8 (**C**), vviTAS9 (**D**) and vviTAS10 (**E**). Numbers above (below) arrows indicate the number of RNA degradation fragments with the 5′-end overlapped to the position determined on the plus (minus) strand. The red number and arrow indicate the cleavage start position directed by potential phase initiator; the blue number and arrow indicate the phased position and the green number and arrow indicate the non-phased position. Horizontal square brackets indicate the phased regions; solid and dotted lines indicate the regions mapped or unmapped by phased sRNAs, respectively

two distinct degradation fragments overlapped on their 5′-ends to the cleavage start position (Fig. 1B).

vviTAS8, vviTAS9 and vviTAS10 had lengths of 210, 357 and 546 nt, respectively, with P values of 8.60×10^{-4}, 2.26×10^{-4} and 1.22×10^{-10}. Eleven sRNAs mapped onto vviTAS8, with four overlapping to phased positions; no RNA degradation fragments were detected (Fig. 1C). There were 23 sRNAs associated with vviTAS9 and 7 of them overlapped to phased positions (Fig. 1D). When we validated the cleaved phase positions of vviTAS9, we found that six distinct RNA degradation fragments mapped to it, but none mapped to phased positions. For vviTAS10, there were 85 sRNAs; 22 were in the 21-nt increment phase and 12 of the phased sRNAs consisted of six pairs of perfect complementary duplexes; 10 RNA degradation fragments mapped onto vviTAS10, but none to the phased positions (Fig. 1E).

We used Mfold (Zuker, 2003) to determine if the initiators were miRNAs. We found that none of them, except for vvimiRNA390 and vvimiRNA8282a, could form a typical stem–loop structure with their 200-bp flanking sequences, indicating that they might be siRNAs.

The length of phase initiator is a key determinant for triggering ta-siRNA biogenesis (Chen *et al*., 2010; Cuperus *et al*., 2010; Laubinger *et al*., 2010). This is usually 22 nt, but exceptions are known, such as miRNA390, which is only 21 nt long in Arabidopsis. The explanation given for this situation is that there are two miRNA target sites in TAS3 RNA (Chen *et al*., 2010). In this study, we identified vvimiRNA390 and vviTAS3-primoRNA2-5 as phase initiators of vviTAS3. When we analyzed their lengths and targeted sites, we found they were all 21 nt long and also had targets in both sides of vviTAS3 (data not shown). This is consistent with the miRNA390 studies reported previously. vviTAS7 and vviTAS10 both have phase initiators with the usual 22 nt, and other vviTASs have only 21 nt phase initiators. Further analyzing their targeted sites, we did not find two targets on their flanking 100 nt, indicating some other mechanism might exist.

3.3 Identification of ta-siRNAs and their *trans*-acting targets

To find new ta-siRNAs and their downstream regulatory pathways, we assumed each sRNA from vviTASs was a ta-siRNA and identified its *trans*-acting targets by searching for complementary regions in genome-wide transcripts. We allowed a maximum of four mismatched base pairs and required that the *trans*-cleaved position corresponds to the 10th position of the ta-siRNA and overlaps the 5′-end of the RNA degradation fragment mapped onto the transcript. Using these criteria, 22 ta-siRNAs and 37 *trans*-acting targets were located (Supplementary Fig. 1 and Supplementary Table 2).

Those ta-siRNAs derived from vviTAS3 mainly target transcripts encoding auxin response factor 4/5, ribosomal protein S1, senescence-associated protein DIN1, ferric reductase, ribosomal protein L10, DNA damage checkpoint protein rad24 and other uncharacterized proteins. The ta-siRNAs derived from vviTAS7 target transcripts encode leucine-rich receptor-like protein kinase (LRPK), adenosine triphosphatase (ATPase), hydroxyproline-rich glycoprotein, histone acetyltransferase, ACBF-like DNA-binding protein, fatty acid desaturase, cellulose synthase catalytic subunit and several uncharacterized proteins. The ta-siRNAs from vviTAS8 target the transcripts encoding snRNP core protein SMX5b and glycerol-3-phosphate acyltransferase. ta-siRNAs from vviTAS9 target transcripts encode chloroplast ribulose-1,5-bisphosphate carboxylase and cytochrome b6-f complex subunit 7. The ta-siRNAs of vviTAS10 mainly target transcripts encoding early-responsive protein to dehydration stress, phosphoribosyl-AMP cyclohydrolase, ATP-binding cassette protein, caspase catalytic subunit p20, malonyl-coenzyme A reductase, 1-phosphatidylinositol-3-phosphate 5-kinase and several uncharacterized proteins. Analyzing these targets, we found that only auxin response factor has been previously reported (Allen *et al*., 2005). The others are novel, which further broadens the known scope of ta-siRNA regulation and indicates that tasi-RNAs play a basic regulatory role in plant growth and development.

3.3 *cis*-activity of ta-siRNAs

Although the *trans*-activity of ta-siRNA is well known, its *cis*-acting function, i.e. cleavage of TAS transcripts, has not been intensely studied. In this study, we examined the *cis*-activity of ta-siRNAs by searching for ta-siRNAs, detecting their degradation products and validating their *cis*-cleaved positions based on the criterion that the *cis*-cleaved position corresponds to the 10th nucleotide position of the ta-siRNA and overlaps with the 5′-end of the RNA degradation fragment mapped onto the TAS.

We observed ta-siRNA *cis*-activity for both vviTAS3 and vviTAS7. As shown in Figure 2, vviTAS3 *cis*-activity was driven by two ta-siRNAs originating from vviTAS3 5′D2(–). The *cis*-cleavage start site was validated by three sRNAs from the RNA degradome. With respect to the results of *cis*-cleavage, five secondary small RNAs—four generated from the 5′ cleavage product and one from the 3′-cleavage product—were found to be arranged on vviTAS3 in 21-nt increments. Three ta-siRNAs, named vviTAS7-tasiRM4a, vviTAS7-tasiRM4b and vviTAS7-tasiRM4c, were involved in vviTAS7 *cis*-activity, and one secondary sRNA was distributed on vviTAS7 in 21-nt increments. These results confirm the *cis*-activity of ta-siRNA, suggest an inactive mechanism of TAS transcription and a produced mechanism of multiple forms of sRNA from the same TAS.

3.4 ta-siRNA regulatory cascades and their conservation

Five different ta-siRNA regulatory cascades were identifiable from the information we uncovered. As illustrated in Figure 3A, the first cascade, related to vviTAS3, starts with cleavage mediated by vvimiRNA390 and vviTAS3-primoRNA2-5. Following cleavage on vviTAS3, six ta-siRNAs are produced that target and *trans*-cleave transcripts of ribosomal protein S1, auxin response factors 4/5, senescence protein DIN1, ferric reductase, ribosomal protein L10, DNA damage checkpoint protein rad24 and several uncharacterized proteins. The ta-siRNAs, which include vviTAS3-tasiRM2a, vviTAS3-tasiRM2b and vviTAS3-tasiRM2c, can also target vviTAS3 itself and direct *cis*-cleavage to produce new sRNAs. In the second cascade, related to vviTAS7 and illustrated in Figure 3B, the process begins with cleavage mediated by vvimiRNA828a and vviTAS7-primoRNA2. Thirteen ta-siRNAs are produced as a

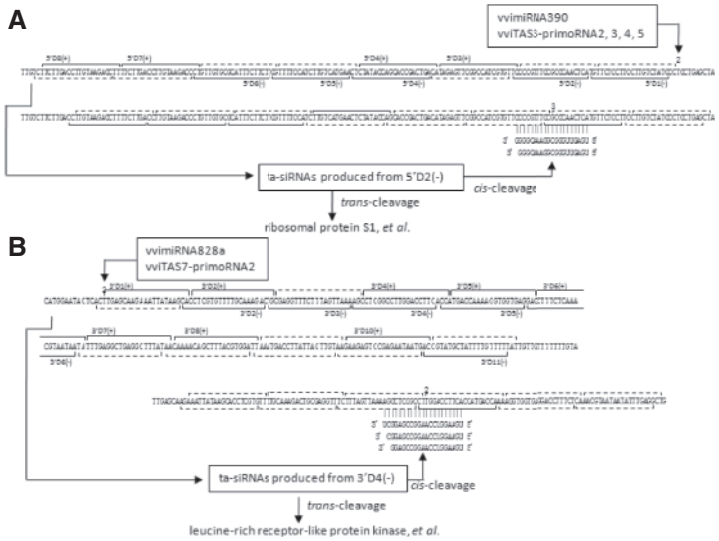

Fig. 2. *Cis*-activity of ta-siRNAs derived from vviTAS3 (**A**) and vviTAS7 (**B**). The numbers above sequences are the fragment number of RNA degradation fragments with the 5′-end mapping at that position. Horizontal square brackets indicate the phased region of TAS, and solid and dotted lines indicate regions with mapped and unmapped phased sRNAs, respectively

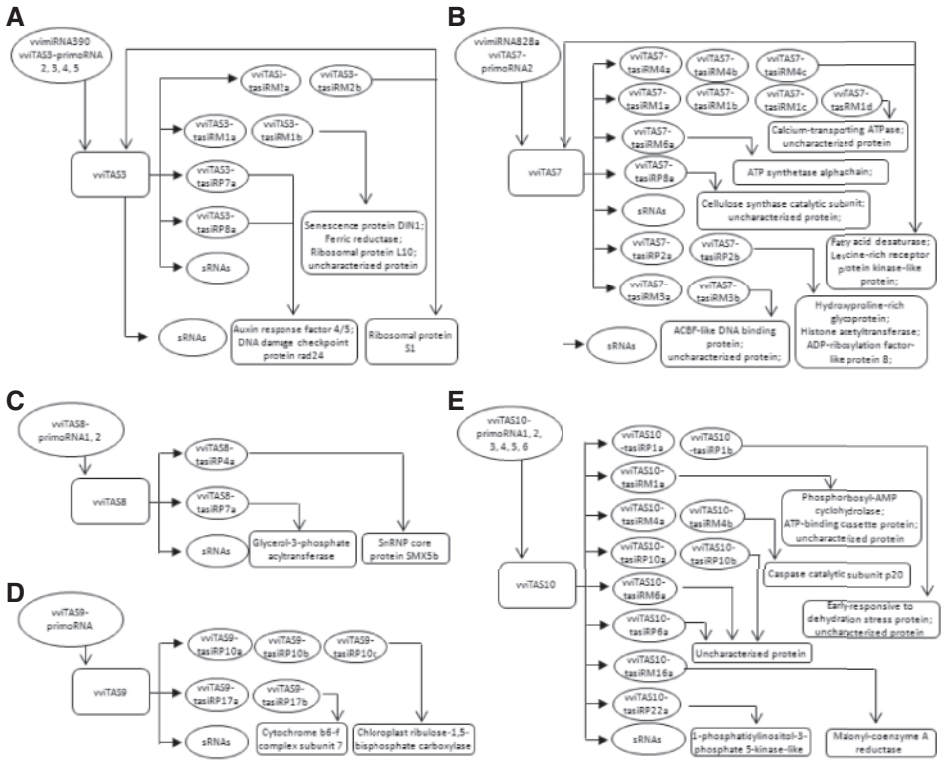

Fig. 3. Five ta-siRNA regulatory cascades related to vviTAS3 (**A**), vviTAS7 (**B**), vviTAS8 (**C**), vviTAS9 (**D**) or vviTAS10 (**E**)

result of this cleavage. They target fatty acid desaturase, leucine-rich receptor protein kinase-like protein, calcium-transporting ATPase, ATP synthetase alpha chain, cellulose synthase catalytic subunit, glycoprotein, histone acetyltransferase, adenosine diphosphate-ribosylation factor-like protein 8, ACBF-like DNA-binding protein, some uncharacterized proteins and vviTAS7 transcripts to produce new sRNAs. In the third cascade, related to vviTAS8 and illustrated in Figure 3C, vviTAS8-primoRNAs target vviTAS8 and direct the production of two ta-siRNAs. They can target glycerol-3-phosphate

acyltransferase and snRNP core protein SMX5b to suppress their expressions. In the fourth cascade, related to vviTAS9 and illustrated in Figure 3D, vviTAS9-primoRNA targets vviTAS9 to produce two groups of ta-siRNAs: vviTAS9-tasiRP10a, b, c and vviTAS9-tasiRP17a, b. These two groups can, respectively, target chloroplast ribulose-1,5-bisphosphate carboxylase and cytochrome b6-f complex subunit 7. The final cascade, related to vviTAS10 and illustrated in Figure 3E, begins with cleavage mediated by vviTAS10-primoRNA1, 2, 3, 4, 5 and 6. Eleven ta-siRNAs are produced as a result, which variously target dehydration stress protein, phosphoribosyl-AMP cyclohydrolase, ATP-binding cassette protein, caspase catalytic subunit p20, 1-phosphatidylinositol-3-phosphate 5-kinase-likeprotein, malonyl-coenzyme A reductase and various uncharacterized proteins.

To evaluate the conservation of these five ta-siRNA regulatory cascades, we examined the expressions of their TASs, phase initiators and ta-siRNAs in different species (Supplementary Table 3).

The top 100 BLAST hits from NCBI showed that ESTs of vviTAS3 exist for 28 species (*V. vinifera*, *Acacia mangium*, *Theobroma cacao*, various citrus species and members of Populus, Quercus, Papilionoideae and Euphorbiaceae); vviTAS8 ESTs exist for four species (*V. vinifera*, *Vitis riparia*, *Vitis amurensis* and *Vitis bryoniifolia*), but vviTAS7, vviTAS9 and vviTAS10 were only found from grapevine. A further examination of the number of sequenced phase initiators in 34 sRNA libraries belonging to grapevine, orange, Arabidopsis and tomato revealed that vvimiRNA390 is highly conserved. Among tested sRNA libraries, vvimiRNA390 was found in 31, whereas the other phase initiators only occurred in grapevine sRNA libraries. On examining the number of ta-siRNA sequences in the 34 sRNA libraries, we found some ta-siRNAs generated from vviTAS3 in at least two species of grapevine, orange, Arabidopsis and tomato, but ta-siRNAs from other vviTASs had only been sequenced from grapevine. This indicates that the ta-siRNA regulatory cascade related to vviTAS3 is universal and that the regulatory cascades related to other vviTASs are grape-specific.

A comparison of the 13 sequenced ta-siRNAs from vviTAS7 revealed that the number of sequences was not very uniform in each grapevine library. vviTAS7-tasiRP1d had been sequenced more than tens of thousands times, but the other ta-siRNAs less so, even though four ta-siRNAs—vviTAS7-tasiRP1a, vviTAS7-tasiRP1b, vviTAS7-tasiRP1c and vviTAS7-tasiRP1d—all originated from vviTAS7 3'D1(+). After further comparisons involving the number of sequenced ta-siRNAs from vviTAS9, the same phenomenon was also observed: vviTAS9-tasiRP10b had been sequenced several thousand times (at least), but the two other ta-siRNAs arising from the same gene as vviTAS9-tasiRP10b were also less sequenced from each sRNA library. Surprisingly, the two most highly sequenced ta-siRNAs were both 21 nt long, showing that 21-nt ta-siRNA is the most prevalent ta-siRNA among its analogs.

4 DISCUSSION

Because ta-siRNAs are produced from TAS, TAS is rich in sRNAs. We can thus detect TASs by initially searching for regions containing large numbers of sRNAs. Using this approach, a total of 49 055 regions were found. When we counted the number of clusters overlapped to annotated gene loci and exons, we found 21-nt clusters were much more likely than 24-nt clusters to overlap with exon regions. In previous studies, 24-nt clusters have been implicated in transcriptional silencing of transposon regions (Johnson *et al.*, 2009). The grapevine 24-nt clusters are seldom associated with gene regions, consistent with their roles in directing silencing of transposable elements.

To further identify TAS loci from regions highly enriched in sRNA, we used an improved computational method that allows for a variable TAS length to predict 49 TASs. Compared with current strategies, the greatest advantage of this improved method is that it is not constrained by a 231-bp length requirement. This leads to a more accurate TAS prediction, especially for the TASs with lengths longer or shorter than 231 bp, and has great potential for future use. However, because the cleavage position cannot be offset, it still requires further refinement. To further validate the cleavage sites of TASs targeted by phase initiators, an RNA degradome was used. Because the RNA degradome was obtained from high-throughput sequencing of products generated from 5′-RACE, the degradome can be used to analyze RNA degradation patterns, which is helpful for identifying siRNA cleavage sites on mRNA or other cDNAs. Here, we used it to determine phase initiator and ta-siRNA cleavage sites.

We identified 22 ta-siRNAs, produced from vviTASs and 37 *trans*-cleaved targeted transcripts. The 37 targeted transcripts are involved in a wide range of biological processes, including basic biochemical responses, such as those related to ribosomal protein S1, DNA damage checkpoint protein rad24, calcium-transporting ATPase, fatty acid desaturase, cellulose synthase catalytic subunit, and chloroplast ribulose-1,5-bisphosphate carboxylase, and development-related processes, such as those related to senescence-associated protein DIN1, auxin response factor and pathogenic defense-related protein LRPK. This wide range of biological processes suggests that ta-siRNAs play a basic regulatory role in plant growth and development.

ta-siRNAs have been suggested as key factors conferring regulatory diversity in members of multi-gene families involved in auxin signaling (Allen *et al.*, 2005). Because leucine-rich receptor-like kinases make up one of the largest protein families, with at least 223 members in Arabidopsis thaliana, and are targeted by ta-siRNAs, they may also be diversely regulated. Any regulatory control by miRNAs and other endogenous sRNAs is still speculative, however.

The function of ta-siRNA *trans*-activity has been the focus of many studies, but its *cis*-activity has rarely been investigated. In a previous study (Allen *et al.*, 2005), a function was suggested but the phase initiator was not found. The *cis*-cleavage position was validated and a secondary siRNA product was detected, but that was not enough information to establish the *cis*-acting role of ta-siRNA. In this study, we have confirmed the *cis*-activity of ta-siRNAs by identifying the phase initiator, as well as detecting degradation products and determining the site of *cis*-cleavage.

The biological role of ta-siRNA regulatory cascades is also interesting. In this article, we have reported five cascades and evaluated their expressions in various species and tissues. The results indicate that the vviTAS3-related cascade is conserved

across species; in contrast, the cascades related to vviTAS7, vviTAS8, vviTAS9 and vviTAS10 are found only in grape. There is a previously known cascade in which ath-miR828 targets ath-TAS4 and triggers production of ta-siRNAs, which in turn regulate a group of MYB genes. The cascade related to vviTAS7 is similar to the ath-miR28-associated cascade in that the phase initiators for both cascades belong to the miR828 family; however, the TASs, ta-siRNAs produced and ta-siRNAs targets, are all different, indicating the discovery of a new miR828 regulation pathway. Previous studies have posited that viral siRNAs are abundant in plants, which implies that ta-siRNA cascades may be involved in antiviral defense. Our findings further suggest that different ta-siRNA cascades might be specific to different species or tissues, opening the possibility that some might have specialized developmental regulatory roles.

Surprisingly, vvimiRNA390 has lower copy numbers in certain Arabidopsis mutant libraries than in wild-type libraries (Supplementary Table 3). We believe that this situation stems primarily from the different experimental conditions used to produce the various libraries. For example, sRNA libraries with lower numbers of vvimiRNA390 copies, including GSM154336, GSM154361,GSM154362,GSM154363,GSM154364,GSM154365,GSM154367, GSM154368, GSM154370, GSM154372, GSM154374, GSM154375 and GSM154376, were all published by the James Lab at Oregon State University, while the wild-type library, GSM149079, was published by the Hannon Lab at Cold Spring Harbor Laboratory. For those Arabidopsis libraries created under identical experimental conditions, the reason that few or no vvimiR390 reads were detected is simply due to sequencing depth. Libraries from the James Lab were generated using 454 sequencing in 2007; the total number of reads in several of those libraries is below 20 000, which may be too low to detect vvimiR390.

When we compared the number of sequenced ta-siRNAs from vviTAS3, vviTAS7, vviTAS8, vviTAS9 and vviTAS10, we found very different numbers in each library from grapevine. This is very surprising as they are all produced from the same TAS transcript simultaneously. This suggests a potentially different degradation rate. In addition, among ta-siRNAs from the same position, such as vviTAS7-tasiRP1a, vviTAS7-tasiRP1b, vviTAS7-tasiRP1c and vviTAS7-tasiRP1d that all originate from vviTAS7 3'D1(+), the 21-nt ta-siRNA is always the most abundant. This suggests that the degradation rate of 21-nt ta-siRNA is very slow, and that its function might be more long lasting than its other analogs. Further studies are required to obtain better insights into the mechanism of ta-siRNA degradation.

Funding: This work was supported by the National Science Foundation of China (Nos. 31171273 and 60901053) and the Qing Lan Project of Jiangsu Province.

Conflict of Interest: none declared.

REFERENCES

Allen,E. *et al.* (2005) microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell*, **121**, 207–221.

Chen,H.M. *et al.* (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem *trans*-acting siRNA cascade in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3318–3323.

Chen,H.M. *et al.* (2010) 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15269–15274.

Cuperus,J.T. *et al.* (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat. Struct. Mol. Biol.*, **17**, 997–1003.

Dai,X. and Zhao,P.X (2008) pssRNAMiner: a plant short small RNA regulatory cascade analysis server. *Nucleic Acids Res.*, **36**, W114–W118.

Dai,X. and Zhao,P.X (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, **39**, W155–W159.

Hsieh,L.C. *et al.* (2009) Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. *Plant Physiol.*, **151**, 2120–2132.

Hu,Q. *et al.* (2011) Specific impact of tobamovirus infection on the Arabidopsis small RNA profile. *PloS One*, **6**, e19549.

Jaillon,O. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Johnson,C. *et al.* (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res.*, **19**, 1429–1440.

Katiyar-Agarwal,S. and Jin,H (2010) Role of small RNAs in host-microbe interactions. *Annu. Rev Phytopathol.*, **48**, 225–246.

Laubinger,S. *et al.* (2010) Global effects of the small RNA biogenesis machinery on the Arabidopsis thaliana transcriptome. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 17466–17473.

Lu,C. *et al.* (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res.*, **16**, 1276–1288.

Luo,Q.J. *et al.* (2012) An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis. *Plant Mol. Biol.*, **80**, 117–129.

Meng,Y. *et al.* (2011) The regulatory activities of plant microRNAs: a more dynamic perspective. *Plant Physiol.*, **157**, 1583–1595.

Molnar,A. *et al.* (2007) miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. *Nature*, **447**, 1126–1129.

Pantaleo,V. *et al.* (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J.*, **62**, 960–976.

Rajagopalan,R. *et al.* (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.*, **20**, 3407–3425.

Talmor-Neiman,M. *et al.* (2006) Identification of *trans*-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. *Plant J.*, **48**, 511–521.

Urano,K. *et al.* (2010) Omics' analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.*, **13**, 132–138.

Wang,C. *et al.* (2011) Deep sequencing of grapevine flower and berry short RNA library for discovery of novel microRNAs and validation of precise sequences of grapevine microRNAs deposited in miRBase. *Physiol Plant.*, **143**, 64–81.

Zhang,C *et al.* (2011) A mutation degree model for the identification of transcriptional regulatory elements. *BMC Bioinformatics*, **12**, 262.

Zuker,M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.