Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*

Ho-Ming Chen*†‡, Yi-Hang Li*, and Shu-Hsing Wu*†‡§

*Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan; †Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung-Hsing University and Academia Sinica, Taipei 11529, Taiwan; and ‡Department of Life Sciences, National Chung-Hsing University, Taichung 402, Taiwan

Communicated by Wen-Hsiung Li, University of Chicago, Chicago, IL, December 14, 2006 (received for review December 5, 2006)

Small RNAs play pivotal roles in regulating gene expression in higher eukaryotes. Among them, trans-acting siRNAs (ta-siRNAs) are a class of small RNAs that regulate plant development. The biogenesis of ta-siRNA depends on microRNA-targeted cleavage followed by the DCL4-mediated production of small RNAs phased in 21-nt increments relative to the cleavage site on both strands. To find *TAS* genes, we have used these characteristics to develop the first computational algorithm that allows for a comprehensive search and statistical evaluation of putative *TAS* genes from any given small RNA database. A search in *Arabidopsis* small RNA massively parallel signature sequencing (MPSS) databases with this algorithm revealed both known and previously unknown ta-siRNA-producing loci. We experimentally validated the biogenesis of ta-siRNAs from two *PPR* genes and the trans-acting activity of one of the ta-siRNAs. The production of ta-siRNAs from the identified *PPR* genes was directed by the cleavage of a *TAS2*-derived ta-siRNA instead of by microRNAs as was reported previously for *TAS1a*, *-b*, *-c*, *TAS2*, and *TAS3* genes. Our results indicate the existence of a small RNA regulatory cascade initiated by miR173-directed cleavage and followed by the consecutive production of ta-siRNAs from two *TAS* genes.

massively parallel signature sequencing | TAS

Small regulatory RNAs modulate transcriptional gene silencing (TGS), mRNA degradation (post-TGS; PTGS) and translational repression in a wide spectrum of organisms. These small regulatory RNAs include microRNAs (miRNAs), heterochromatic siRNAs (hc-siRNAs), repeat-associated siRNAs (ra-siRNAs), natural sense–antisense transcript siRNAs (1), trans-acting siRNAs (ta-siRNAs) (2, 3), and the recently identified Piwi-interacting RNAs (4).

In *Arabidopsis*, miRNAs are processed from hairpin precursors to play important roles in development and stress responses by either targeted cleavage of mRNA or translational repression (for review see ref. 5). The biogenesis of miRNAs requires a specific RNase III enzyme, DICER-LIKE protein 1 (DCL1) (5). *Arabidopsis* hc-siRNAs or ra-siRNAs trigger DNA methylation and histone modification and are thus involved in the assembly of heterochromatin and the control of transposon movement. hc-siRNAs or ra-siRNAs are usually derived from genomic repeats or transposons, a process requiring DICER-LIKE 3 (DCL3) and a specific RNA-dependent RNA polymerase, RDR2 (for reviews, see refs. 6 and 7).

The identification of ta-siRNAs in *Arabidopsis* bridged the miRNA and siRNA pathways previously considered independent (2, 3, 8–10). ta-siRNAs clustered in 21-nt increments in both sense and antisense strands of several noncoding *TAS* transcripts were first identified from the study of two genes involved in PTGS, *RDR6* and suppressor of gene silencing 3 (*SGS3*) (3, 8). Interestingly, the production of phased ta-siRNA is initially triggered by the targeted cleavage of primary *TAS* transcripts by miRNAs (2). After cleavage, the 5′ or 3′ cleavage products are converted into dsRNA with the assistance of RDR6 and SGS3 and subsequently processed by DCL4 to produce ta-siRNAs phased in 21-nt increments relative to the miRNA cleavage sites (10, 11).

Five *TAS* loci belonging to three families have been identified and experimentally validated in *Arabidopsis*. *TAS1a*, *TAS1b*, and *TAS1c* are targeted by miR173, and their 3′ cleavage products produce either identical or very closely related ta-siRNAs, which target a group of genes with unknown function (2) and, potentially, a group of pentatricopeptide repeat (PPR) genes (10). In addition to *TAS1* loci, *TAS2* is targeted by miR173, and the 3′ cleavage product generates ta-siRNAs targeting a group of *PPR* genes (2). Different from *TAS1a*, *-b*, and *-c* and *TAS2*, *TAS3* is targeted by miR390, and its 5′ cleavage product produces two ta-siRNAs that target three auxin response factors: *ARF2*, *ARF3*, and *ARF4* (2, 9). Recent studies showed that ta-siRNAs from *TAS3* regulate the transition from juvenile to adult phase and leaf morphology (12–15).

It has been reported that ta-siRNAs derived from *TAS1s*, *TAS2* and *TAS3* were highly enriched in small RNA populations from *Arabidopsis rdr2* mutant. With this as a reference, the *rdr2* massively parallel signature sequencing (MPSS) database was searched for potential new *TAS* loci by inspecting regions rich in RNA clusters. The analysis identified 28 potential *TAS* loci with such characteristics (16). Two loci, containing small49 and small58, were experimentally validated to produce small RNAs with typical ta-siRNA expression pattern. A locus containing small49 also yielded 21-nt phased RNAs. However, it remains unclear whether the other potential *TAS* loci also produce phased small RNAs.

A more recent report elegantly provided one possible mechanistic view for the production of phased RNA in moss (17). In essence, phased siRNAs were found to be derived from regions flanked by dual target sites of miRNA or siRNAs. A "two-hit trigger" was thus hypothesized as an evolutionarily conserved mechanism for the biogenesis of phased siRNA. However, whether the "two-hit trigger" will account for the sole mechanism for generating phased siRNA remains to be further investigated.

With the recent advances in large-scale sequencing of small RNAs via MPSS or 454 pyrosequencing technologies (18), we were interested in developing a computational algorithm that could be broadly applied to predict *TAS* genes by data mining any given large-scale small RNA sequencing data with no restriction in prior knowledge of miRNA/siRNA target sites and of nonrestricted

Author contributions: H.-M.C. and S.-H.W. designed research; H.-M.C., Y.-H.L., and S.-H.W. performed research; H.-M.C. contributed new reagents/analytic tools; H.-M.C. and S.-H.W. analyzed data; and H.-M.C. and S.-H.W. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: TGS, transcriptional gene silencing; PTGS, post-TGS; miRNA, microRNA; DCL, DICER-LIKE; RDR, RNA-dependent RNA polymerase; ta-siRNA, trans-acting siRNA; PPR, pentatricopeptide repeat; MPSS, massively parallel signature sequencing.

§To whom correspondence should be addressed. E-mail: shuwu@gate.sinica.edu.tw.

This article contains supporting information online at www.pnas.org/cgi/content/full/0611119104/DC1.

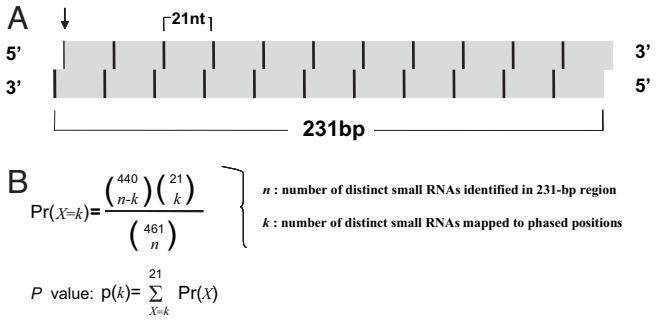

Fig. 1. Theoretical basis and derivation of the *TAS* prediction algorithm. (*A*) The vertical arrow indicates the start site for the small RNA used to determine the phased and nonphased positions. 21 phased sites relative to the start site are indicated as black vertical bars. Four hundred forty nonphased sites relative to the start site are indicated as gray. (*B*) Equation based on hypergeometric distribution for statistically evaluating the presence of phased siRNA in genomic fragment defined in *A*.

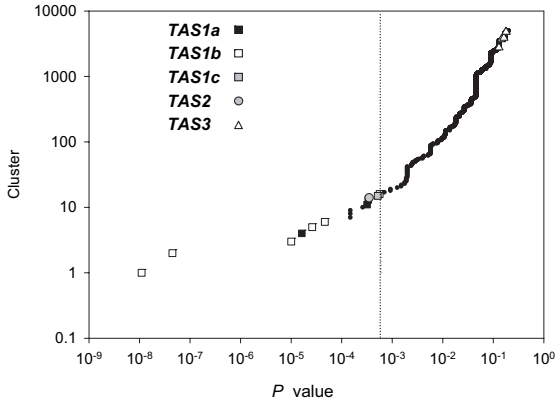

Fig. 2. *P* value plot of ranked small RNA clusters. Small RNA clusters with $P < 0.2$ are ranked according to their *P* values. The vertical dashed line indicates the threshold of $P = 0.0006$. The small RNA clusters mapped to *TAS1a*,- *b*, and -*c* and *TAS2* with *P* values <0.0006 and *TAS3* with $P < 0.2$ are shown as symbols indicated. Black dots represent other small RNA clusters.

organism origin. By integrating statistical evaluation, we developed an algorithm for detecting 21-nt phased small RNA clusters. The practical application of this algorithm was assessed by examining the *Arabidopsis* small RNA MPSS data (16, 19) to reveal putative *TAS* genes. The experimental validation of some of these new *TAS* genes exposed an miRNA → ta-siRNA → ta-siRNA → target gene cleavage cascade in *Arabidopsis*. We discuss the potential application of our computational algorithm and the possible significance of the ta-siRNA cascade.

Results

A Widely Applicable *TAS* Prediction Algorithm. Previous computational approaches have successfully identified miRNA-encoding *MIR* genes in many organisms on the basis of the secondary structure of precursors and conservation among species (20–22). *TAS* genes lack such sequence features, and thus require a different computational prediction strategy. miRNA-directed cleavage and phase setting are crucial for the production of ta-siRNA. However, small RNA clusters have not been commonly observed for most target genes of known miRNAs (19). This finding suggests that the cleavage of miRNA target genes does not ensure the production of ta-siRNAs and thus may not be a definitive criterion for *TAS* gene prediction. On the other hand, all known *TAS* genes produce phased small RNAs in 21-nt increments (2). Therefore, we developed an algorithm based on Perl [supporting information (SI) *Script*] to identify small RNAs phased in 21-nt increments that could represent potential *TAS* genes. The script inspects small RNA configuration in a 231-bp fragment downstream from the 5′ start site of each small RNA in the small RNA data analyzed (Fig. 1*A*). The 231-nt sequence from the antisense strand has a 2-nt shift to mimic the DCL4 cleavage result. This region is expected to contain 21 "phased" positions in 21-nt increments and 440 "nonphased" positions relative to the start site of each chosen small RNA. Both the number of distinct small RNAs identified in this region (n) and the number of distinct small RNAs mapped to the phased positions (k) are counted. To differentiate significant occurrences from random events, the *P* value of obtaining equal or more than k phased small RNAs is calculated with the equations on the basis of a hypergeometric distribution (Fig. 1*B*). The execution of this algorithm does not require prior knowledge of miRNA or siRNA target sites; thus, it can potentially be used to identify ta-siRNAs with a phase set by unidentified miRNAs/siRNAs. We also applied statistical evaluation to differentiate significant occurrences from random events.

Taking the advantage of abundant small RNA data and extensive studies of small RNA in *Arabidopsis*, we first applied our *TAS* prediction algorithm to *Arabidopsis* MPSS small RNA data of Col-0 to validate the concept for developing this algorithm. A plot depicting the ranked small RNA clusters in Col-0 against their corresponding *P* values is shown in Fig. 2. Our algorithm calculates *P* values for all RNA clusters initiated by every single small RNA species; thus, genomic regions rich in small RNAs are usually represented by multiple small RNA clusters. Most of the small RNA clusters with the low *P* values ($P < 0.001$) are derived from known *TAS* genes, which suggests that the *P* value calculated based on the hypergeometric distribution is a useful indicator to capture the phasing phenomenon and identify potential *TAS* genes (Fig. 2). In addition, a close examination of these small RNA clusters corresponding to known *TAS* loci reveals that the phase identified by our algorithm is perfectly consistent with that set by the miRNA-directed cleavage reported previously (data not shown).

To investigate whether phasing also occurs in small RNAs with phase sizes different from 21 nt, we compared the number of phased small RNA clusters in 21-nt increments to those in different phasing intervals ranging from 19 to 24 nt for different *P* values. The analyses of the Col-0 MPSS data revealed the small RNA clusters phased in 21-nt intervals with the most significant *P* values ($P < 0.0005$) (Table 1), which suggests that the phased siRNAs are predominantly produced by DCLs yielding 21-nt siRNAs in wild-type *Arabidopsis*.

The 19- and 20-nt small RNAs are less abundant than 21- to 24-nt small RNAs (16, 23). However, the number of small RNA clusters with 19- or 20-nt phasing at *P* value <0.01 is similar to that with 22- or 23-nt phasing (Table 1). Thus, the phased small RNA clusters identified from the analyses of 19- and 20-nt phasing likely represent random occurrence events and could be used to estimate the background noise in computational analyses. We found no small

Table 1. Number of phased small RNA clusters of 19- to 24-nt phasing intervals in Col-0 (wild-type) and *rdr2* mutant

	Number of phased small RNA clusters at *P* values indicated (Col-0/*rdr2*)				
Phasing intervals, nt	<0.0001	0.0001–0.0005	0.0005–0.001	0.001–0.005	0.005–0.01
19	0/0	0/1	4/1	38/7	44/13
20	0/0	0/0	4/6	29/14	36/28
21	6/65	8/17	5/11	42/58	54/54
22	0/0	0/0	4/1	24/3	56/15
23	0/0	0/0	6/0	39/6	56/26
24	0/0	1/0	7/0	68/3	79/11

GENETICS

Table 2. Output loci from the *TAS* gene prediction algorithm ($P < 0.0006$) using Col-0 MPSS small RNA data

Locus	Name	Coordinate*	Strand*	$n^{\dagger}$	$k^{\dagger}$	P value†	Ref.‡
At1g50055	TAS1b	18,553,151	W	14	8	1.10×10^{-8}	2, 8
At2g27400	TAS1a	11,729,045	W	15	6	1.62×10^{-5}	2, 3, 8
At1g63070§	PPR	23,390,084	W	21	6	1.48×10^{-4}	16
At1g63080§	PPR	23,393,675	W	21	6	1.48×10^{-4}	16, 17
At1g63130	PPR	23,417,284	C	21	6	1.48×10^{-4}	16, 17
At2g39681	TAS2	16,546,913	W	9	4	3.49×10^{-4}	2
At2g39675	TAS1c	16,544,875	W	17	5	5.18×10^{-4}	2, 8

*The strand and the coordinate for the start site of the small RNA used to determine the phased positions in calculating the corresponding *P* value. W, Watson; C, Crick.
†*n*, *k*, and *P* value abbreviations are described in *Results*.
‡Literature with *TAS* or potential *TAS* loci initially described.
§These two genes share the same small RNA cluster.

RNA clusters phased in 19- or 20-nt at *P* value 0.0005 (Table 1). Thus, *P* values between 0.0005 and 0.001 would provide stringent cutoffs for the identification of phased 21-nt small RNA clusters for potential *TAS* loci in the analysis of Col-0 MPSS data with the most satisfactory false discovery rate.

Identification of Phased Small RNAs Clustered in 21-nt Increments From *Arabidopsis* MPSS Small RNA Data. In Col-0, we obtained 16 small RNA clusters mapped to seven loci with $P < 0.0006$; those with the lowest *P* values for each locus are listed in Table 2. Among them, the small RNA cluster with the lowest *P* value is derived from *TAS1b*. A total of 14 distinct small RNAs exist for this small RNA cluster, and eight are in 21-nt increment phase. Four of the five known *TAS* genes are included in the list, which indicates that our approach was effective in identifying *TAS* genes. In addition to revealing known *TAS* genes, our analyses of Col-0 MPSS data revealed three potential *TAS* genes with protein-coding nature, encoding PPR proteins (At1g63130, At1g63070, and At1g63080). SI Table 4 shows an extended list of 35 loci with $P < 0.005$ (4 known and 31 putative *TAS* loci) and their predicted target genes.

Although 28 potential *TAS* loci were hypothesized on the basis of their enrichment in the *rdr2* mutant (16), whether these loci will produce phased small RNAs remain unaddressed. Application of our algorithm for the analysis of *rdr2* MPSS data identified five known and 14 potential *TAS* loci with $P < 0.005$ (Table 3). The prediction of *TAS* genes with the *rdr2* MPSS data resulted in an extended list in addition to that in Table 2, likely due to the enrichment of ta-siRNAs in the *rdr2* mutant. A more relaxed *P* value ($P < 0.005$) was applied for the *rdr2* data, since the number of small RNA clusters phased in 21-nt increments remains significantly higher than those phased in 19- to 20- or 22- to 24-nt increments (Table 1). Our data revealed that, among the 28 potential *TAS* loci reported in ref. 16, only 11 significantly produce phased small RNAs. This finding further strengthens the necessity for introducing statistical analyses in *TAS* gene prediction.

Among the 11 common loci described above, At5g39370 (S-locus glycoprotein) has not been annotated to be a target of known miRNAs (24). It is, however, interesting to observe an miR447 target site localized at the potential 5′ UTR sequence of At5g39370. This miR447 target site has been unannotated previously because no ESTs are available to extend the gene model beyond the predicted ORF (www.arabidopsis.org). As shown in Fig. 3*A*, the targeted cleavage of At5g39370 by miR447 is expected to set the phase for the production of a cluster of small RNAs, by a significant *P* value of 0.00062. The appearance of *TAS3* only in the analyses of *rdr2* data further justified the advantage of obtaining small RNA data sets from *Arabidopsis* mutants defective in small RNA biosynthetic pathways.

Similar to the known *TAS* genes, these potential *TAS* genes produce small RNAs from both strands with 2-nt overhangs and some with perfect complementary duplexes. As examples, we provide the locations of both phased and nonphased small RNAs for S-locus glycoprotein and three PPR genes described above (Fig. 3 *A* and *B*). Of note, in contrast to the noncoding nature of *TAS1*, *TAS2*, and *TAS3*, many potential *TAS* genes listed in Tables 2 and 3 are annotated as protein-coding genes, especially the PPR-encoding genes. Among them, At1g63070, At1g63080, and At1g63130 were identified in both Col-0 (Table 2) and *rdr2* (Table 3) MPSS data and thus especially caught our interest.

***At1g63130* Is a *TAS* Gene Producing ta-siRNAs.** Recent 454 sequencing data indicated that the locus At1g63130 is rich in small RNAs, and a significant portion of these small RNAs exist in 21-nt increments (17). Whether the expression of these small RNAs, like previously identified ta-siRNAs, depends on functional RDR6, SGS3, and DCL1 and, in part, DCL4, is unclear. We thus examined the presence of two At1g63130-derived small RNAs, a sense siR5s and an antisense siR9as (for location, see Fig. 3*B*) in various *Arabidopsis* genetic backgrounds (siR9as is represented in the *rdr2* MPSS data but not in the Col-0 MPSS data). At least six mismatches exist between these two small RNAs and *TAS1abc*/*TAS2*/*TAS3* transcripts (data not shown; also true for siR3as from At1g63080; see below). This sequence divergence is sufficient to distinguish these small RNAs from ta-siRNAs derived from known *TAS* genes. Northern blot analyses indicate that the expression of siR5s and siR9as is normal in the wild type, *dcl2-1* and *dcl3-1*; reduced in *dcl4-2*; and absent in *dcl1-9*, *rdr6-11* and *sgs3-11* (Fig. 4*A*). These results indicate that the expression pattern of siR5s and siR9as is identical to that of the reported ta-siRNAs (2, 3, 8, 10, 11, 25). Taken together, these results indicate that the small RNAs derived from At1g63130 are likely ta-siRNAs.

The Production of Secondary ta-siRNAs Directed by the Cleavage of a *TAS2*-Derived ta-siRNA. As predicted, At1g63130 could be targeted by miR400, miR161, and *TAS2*-derived ta-siR2140. Interestingly, the small RNA cluster in At1g63130 identified by our algorithm is immediately next to the ta-siR2140 target site (Fig. 3*B*; ref. 10). The production of phased siRNAs from At1g63130 through a dual small RNA-targeted mechanism was recently proposed (17). However, whether the production of phased siRNA from At1g63130 really depends on *TAS2* remains to be experimentally addressed. Our data showed that, indeed, two newly identified At1g63130-derived ta-siRNAs, siR5s and siR9as, failed to accumulate in the *tas2* mutant (Fig. 4*A*). These data confirm that ta-siRNAs, like miRNAs, can direct the biogenesis of ta-siRNAs. The production of ta-siRNAs thus relies more on the nature of miRNA or siRNA target transcripts than that of the miRNAs or siRNAs themselves. The discovery of dual small RNA targeting on one single transcript represents one elegant elucidation of this nature (17).

We also examined whether another *TAS* gene, At1g63080, is also

Table 3. Output loci from the *TAS* gene prediction algorithm ($P < 0.005$) using *rdr2* MPSS small RNA data

Locus*	Name*	Coordinate†	Strand†	n‡	k‡	P value‡	Ref.§
At1g50055	TAS1b	18,553,151	W	46	13	2.04×10^{-9}	2, 8
At2g39681	TAS2	16,547,078	C	73	15	6.58×10^{-9}	2
At1g63150	PPR	23,423,856	C	23	9	7.12×10^{-8}	16, 17
At1g63080	PPR	23,393,654	W	27	9	3.70×10^{-7}	16, 17
At2g27400	TAS1a	11,729,168	C	56	12	4.18×10^{-7}	2, 3, 8
At1g63130	PPR	23,417,098	W	36	10	4.56×10^{-7}	16, 17
At1g62910	PPR	23,303,391	C	10	6	7.87×10^{-7}	16, 17
At1g63070	PPR	23,390,043	W	24	7	3.00×10^{-5}	16
At1g12820	AFB3	4,368,842	W	6	4	4.55×10^{-5}	16, 17
At2g39675	TAS1c	16,544,896	W	37	8	7.66×10^{-5}	2, 8
At1g15400/At1g15410	IGR	5,297,945	W	51	9	1.30×10^{-4}	16
At1g48410	AGO1	17,895,163	C	10	4	5.64×10^{-4}	17
At5g39370	SLG	15,774,916	W	37	7	6.17×10^{-4}	16
At1g62930	PPR	23,311,213	C	40	7	1.03×10^{-3}	16, 17
At3g17185	TAS3	5,862,213	W	42	7	1.40×10^{-3}	2
At1g63400	PPR	23,511,931	C	31	6	1.50×10^{-3}	16, 17
At3g06433/At3g06435	IGR	1,966,491	W	175	15	1.54×10^{-3}	
At2g16586	Unknown	7,198,532	W	14	4	2.38×10^{-3}	16
At3g23690	MYM9.3	8,530,582	C	7	3	2.55×10^{-3}	17

*Loci mapped to intergenic regions (IGR) were indicated between two AGI loci. An rRNA locus with P value of 3.29×10^{-3} was excluded from the list.

†The strand and the coordinate for the start site of the small RNA were used to determine the phased positions in calculating the corresponding P value. W, Watson; C, Crick.

‡n, k, and P value abbreviations are described in *Results*.

§Literature with *TAS* or potential *TAS* loci initially described.

targeted by ta-siR2140 for cleavage. We used modified 5′ RACE, commonly used to identify cleavage site(s) on RNA molecules complementary to miRNA or ta-siRNAs. ta-siR2140 could indeed direct At1g63080 for cleavage (Fig. 4*B*). The cleavage site is identical to that for At1g63130 targeted by ta-siR2140 (10). Of note, most of the small RNAs clustered in At1g63080 have a 1-nt shift as compared with the phase set by ta-siR2140. On inspection of other *TAS* genes, we found small RNAs with a 1-nt shift from the phase set by cleavage also observed previously (17). Whether this finding implies a general or alternative cleavage pattern remains to be studied when more *TAS* genes or ta-siRNAs become available.

The accumulation pattern of At1g63080-derived siR3as mimics that of ta-siRNAs (Fig. 4*C*), which indicates that At1g63080 also represents a *TAS* locus. Similar to At1g63130-derived siR5s and siR9as, the production of siR3as depends on *TAS2*, a locus producing primary ta-siRNA (Fig. 4*C*). Therefore, the ta-siRNAs produced from these two loci are logically secondary ta-siRNAs.

Targets of Secondary ta-siRNAs from PPR Genes. The targets of newly identified ta-siRNAs were predicted by use of miRU (http://bioinfo3.noble.org/miRNA/miRU.htm) (26). Although ta-siRNAs from the sense strand had few good targets, most of the ta-siRNAs derived from the antisense strand of the three *PPR* genes were predicted to target a few *PPR* genes (SI Table 4). Among the candidate target genes, At1g62930 (also a *PPR* gene) possesses an identical complementary sequence to At1g63130-derived siR9as and thus was selected for cleavage-site validation. Results from modified 5′ RACE indicate that At1g63160-derived siR9as could target At1g62930 for cleavage, and we also determined the cleavage site (Fig. 4*D*).

miRNA-Directed Production of a Tandem ta-siRNA Cascade in *Arabidopsis*. Our studies indicate the existence of an endogenous small RNA regulatory cascade, as illustrated in Fig. 4*E*. The process starts with the miR173-mediated cleavage of *TAS2* transcript and the production of phased primary ta-siRNAs. One of the *TAS2*-derived ta-siRNAs, ta-siR2140, targets two *PPR* genes, At1g63130 and At1g63080. The 3′ cleavage products of At1g63130 and At1g63080 then produce secondary ta-siRNAs, with the phase set by ta-siRNA rather than by miRNAs as described for *TAS1a*, *-b*, *-c*, *TAS2*, and *TAS3*. As an example, one of the secondary ta-siRNAs, siR9as, targets another *PPR* gene, At1g62930, for cleavage.

Discussion

Tandem ta-siRNA cascade in *Arabidopsis*. The production of ta-siRNAs from the 3′ cleavage product of At1g63130 is similar to that for the *TAS1* family and *TAS2*. Because ARGONAUTE 1 (AGO1) is responsible for the targeted cleavage by *TAS1*-derived ta-siRNAs (27, 28), AGO1 is likely the Slicer for siR9as-At1g62930 pairing. Because RDR6, SGS3, and DCL4 are involved in the biogenesis of primary ta-siRNAs, accessing their direct contribution in generating secondary ta-siRNAs is genetically challenging at the present time.

Our analyses suggest the existence of an miRNA → ta-siRNA → ta-siRNA → target gene cleavage cascade in *Arabidopsis*. The results also imply an amplification effect of PTGS initially triggered by miRNA and immediately followed by a tandem cascade of ta-siRNAs. That is, by producing cascades of ta-siRNAs, one or a few miRNAs could transmit the gene-silencing signal to genes far beyond their original targets, a process inversely analogous to the gene-induction cascade triggered by transcription factors. This finding also expands the current knowledge of small regulatory RNAs in *Arabidopsis* (29–31).

***PPR*, Tandem Repeats, and *TAS* Genes.** Tandem repeats frequently produce 24-nt siRNA in a DCL3- and an RDR2-dependent manner, and these siRNAs act by guiding methylation in TGS. The small RNA clusters identified in the three *PPR* genes locate in the regions containing 105-nt tandem repeats. Even though tandem repeats are present in 103 of the 448 *PPR* genes, only 11 of the 103 *PPR*s yield small RNA clusters ($n \geq 10$ in the Col-0 MPSS database), which suggests that tandem repeats within *PPR* genes alone do not guarantee the production of ta-siRNAs unless the transcripts are targeted by miRNAs or ta-siRNAs. In addition, that the production of small RNAs from these putative *TAS* genes depends on DCL4 and RDR6 but not DCL3 (Fig. 4) argues against the possibility of these small RNAs being ra-siRNAs.

Tandem repeats have been proposed to function in sustaining the

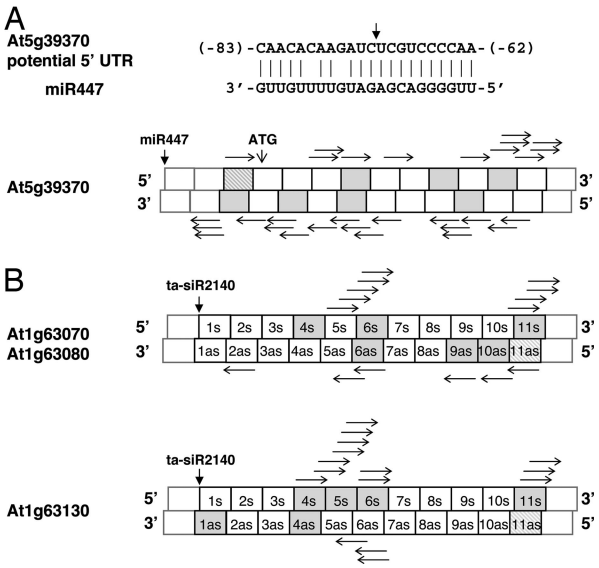

Fig. 3. Positions of phased and nonphased small RNAs on candidate *TAS* genes. Small RNAs represented in the MPSS database and clustered in 21-nt increments for At5g39370 in *rdr2* MPSS (*A*) and At1g63070/At1g63080/At1g63130 in Col-0 MPSS (*B*) data are indicated as gray boxes. Striped boxes correspond to the small RNAs whose start sites were used to determine the phased positions and to calculate the corresponding *P* value listed in Tables 2 and 3. The validated cleavage sites directed by ta-siR2140 from *TAS2* are marked by vertical arrows for At1g63070/At1g63080/At1g63130. The predicted cleavage site by miR447 on At5g39370 and the position of its initial codon (ATG) are also marked. Labels in each box indicate the predicted ta-siRNAs derived from the 3′ cleavage product of the candidate *TAS* genes. Horizontal arrows mark all small RNAs (less than or equal to six hits) not in phase but represented in the MPSS databases. s, sense strand; as, antisense strand.

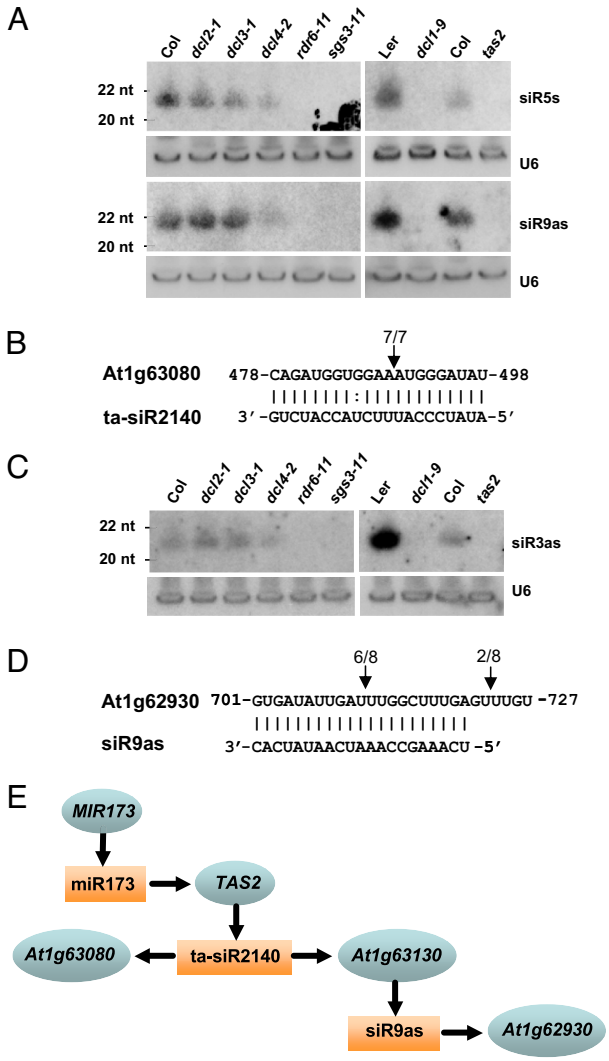

Fig. 4. At1g63130 and At1g63080 produce ta-siRNAs. (*A*) Expression pattern of the At1g63130-derived ta-siRNAs, siR5s and siR9as, in the wild type, mutants defective in various small RNA pathways, and *tas2*. The hybridization with U6 was used as a loading control. (*B*) At1g63080 is targeted for cleavage by *TAS2*-derived ta-siR2140. The cleavage site is mapped by modified 5′ RACE as described in *Materials and Methods* and marked with an arrowhead. The frequency of the mapped position is also indicated. (*C*) Expression pattern of the At1g63080-derived ta-siRNA, siR3as, in the wild type, mutants defective in various small RNA pathways, and *tas2*. The hybridization with U6 was used as a loading control. (*D*) Experimental validation of the siR9as-directed cleavage of At1g62930 via modified 5′ RACE. Positions of cleavage sites are marked with arrowheads, with sequence frequencies shown. (*E*) A cascade showing the tandem production of small RNAs in *Arabidopsis*. Vertical arrows indicate the production of small RNAs from corresponding genetic loci. Horizontal arrows indicate the sequence-specific targeting of small RNAs to target transcripts.

RDR-dependent synthesis of dsRNA, from which hc-siRNAs involved in heterochromatin remodeling are derived (32). Intrinsic direct repeats of transgenes have been demonstrated to effectively trigger PTGS, even under the control of a weak promoter (33). Whether the tandem repeats in the *TAS* genes are also involved in the production of ta-siRNAs remains to be studied and will potentially help uncover the features for predicting new *TAS* genes.

The finding of *PPR* genes producing ta-siRNAs further supports the hypothesis that *TAS2*, and possibly three *TAS1* genes, evolved from members of the *PPR* family (10). These three *PPR* genes are predicted or validated targets of *TAS2*-derived ta-siRNAs and share high sequence similarity with other PPR genes targeted by *TAS2* ta-siRNAs. This observation suggests that the three *PPR* genes identified in this study might be young *TAS* genes that still maintain the coding nature. Alternatively, these genes could have dual-function transcripts, both for producing protein products and for yielding siRNAs with regulatory roles toward their target genes. Of note, these *PPR* transcripts are presumably targets of miRNAs and validated targets of ta-siRNAs. This observation implies that a multilayered gene-silencing mechanism is involved in regulating the expression of this group of *PPR* genes. Indeed, phased siRNAs could be generated from *PPR* loci with two or more target sites of miRNAs and/or ta-siRNAs (17). This study provides a possible mechanistic explanation for the high representation of *PPRs* that produce ta-siRNAs.

Computational Prediction of *TAS* Genes. Here we report a computational method for comprehensive and genome-wide prediction of *TAS* genes. The identification of all known and additional *TAS*-like genes by our approach validated the central concept of the computational design. The successful prediction of new *TAS*-like genes was also cross-validated experimentally.

A few unique features of our computational method make it a nonbiased and highly applicable means to predict loci yielding phased small RNAs. First, this computational method does not require prior knowledge of miRNA/siRNA cleavage sites. Although a few phased siRNA-producing loci were elegantly revealed by searching 454 sequencing data for transcripts with two or more small RNA targets (17), our current computational algorithm could identify *TAS* loci generating ta-siRNAs possibly via alternative biogenesis mechanisms. Second, this methodology does not rely on the comparison between two small-RNA

datasets such as the wild type and *rdr2* mutant; thus, it offers superior prediction power when a specialized small RNA database, such as *rdr2* MPSS data, is unavailable. It is especially useful when analyzing data from organisms with incomplete knowledge of small RNA biosynthetic pathways or with lack of proper molecular means in generating desirable mutants. Third, we applied statistical evaluation for adjustable assessment of false discovery rate. Fourth, this computational method can be broadly applied and easily implemented. The future availability of *Arabidopsis* small RNA data from various genetic backgrounds, tissues, and plants grown under various environmental stimuli will allow for the identification of more *TAS* genes with specific regulatory missions.

Although *TAS* genes were first found in higher plants, recent reports have indicated their existence in moss (17, 34). Whether *TAS* genes are present in an even wider spectrum of organisms is of great interest, especially RDRs also present in *C. elegans*, *D. discoideum*, and many fungi. The study of *TAS* gene evolution will become feasible when *TAS* genes are revealed in species other than higher plants. Because of the 17-bp read nature of MPSS data, the exact size for each small RNA signature is unspecified. Thus, the abundant representation of 24-nt small RNAs in wild-type plants will likely increase the noise of this current analysis performed in Col-0 (Table 1). Future size information of small RNAs derived from 454 sequencing will further reduce the false discovery rate and thus increase the sensitivity of this approach. We expect that our method will be easily adopted for the identification of new *TAS* genes when more large-scale small RNA data sets become available. Such research will further advance the study of more regulatory RNAs that play diverse and significant roles in regulating gene expression in eukaryotic cells (29–31).

Materials and Methods

Computational Prediction of *TAS* Genes. The *Arabidopsis* MPSS small RNA signatures was downloaded from http://mpss.udel.edu/at (16, 19) (as of September 2005 for Col-0 and November 2006 for *rdr2*). A total of 73,086 and 18,927 distinct signatures corresponding to the first 17-nt sequence of 21- to 24-nt small RNA sequences were extracted from the Col-0 and *rdr2* data sets, respectively. Small RNA signatures with greater than six hits in the *Arabidopsis* genome were further filtered and removed because of mapping ambiguity. The coordinates of MPSS small RNA signatures were then subjected to the analyses using the *TAS* prediction algorithm developed based on Perl. For Table 1, the fragment size is determined by multiplying the phasing interval size by 11 for calculating small RNAs phased in 19- to 20- or 22- to 24-nt intervals. The above equation is also adjusted accordingly to reflect the numbers of phased and nonphased positions. New potential *TAS* loci with $P < 0.0006$ for Col-0 and $P < 0.005$ for *rdr2*, respectively, were searched against the *Arabidopsis* Small RNA Project (ASRP) data set (24) (http://asrp.cgrb.oregonstate.edu/db/) for potential small RNA(s) that could target and set the phase for the production of small RNA clusters identified in our study.

Plant Materials and Growth Conditions. Seeds of *Arabidopsis* mutant lines *dcl1-9* (CS3828) (35), *dcl2-1* (SALK_064627), *dcl3-1* (SALK_005512), *dcl4-2* (GABI_160G05), *rdr6-11* (CS24285) (8), *sgs3-11* (CS24289) (8) and *tas2* (SALK_014168) were obtained from the *Arabidopsis* Biological Research Center or Nottingham *Arabidopsis* Stock Center. Seedlings of Col-0, Ler, and all mutants used in this study were grown in 1% agar plates containing 1× MS and 1% sucrose under a 16/8-h light/dark cycle at 22°C.

Small RNA Northern Blot Analyses. Total RNA was extracted from 14-d-old seedlings with use of TRIZOL reagent (Invitrogen, Carlsbad, CA). Fifty micrograms of total RNA was separated by 15% denaturing polyacrylamide TBE-Urea gels (Invitrogen) and transferred to Hybond-N^+ membranes (GE Healthcare, Piscataway, NJ) by use of a transblot semidry transfer cell (Bio-Rad, Hercules, CA). Antisense DNA oligonucleotides of 21 nt complementary to predicted ta-siRNAs were used as probes (for At1g63130-siR5s, 5′-CACAACCATCCGGTCAACTAA-3′; At1g63130-siR9as, 5′-GTGATATTGATTTGGCTTTGA-3′; and At1g63080-siR3as, 5′-TCATGGGCTTTTTCAACACAA-3′). The probes were end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. The membrane was hybridized with ULTRAhyb-Oligo buffer (Ambion, Austin, TX) and exposed to Kodak BioMAX MS x-ray films for 3–5 days.

Validation of ta-siRNA Targets. Modified 5′ RACE with the GeneRacer kit (Invitrogen) was adapted to validate the cleavage site determined by ta-siRNA targeting. Nested primers were used in PCRs and the cleavage sites were revealed by sequence analyses of the PCR product. Two nested primers for At1g63080 were primary, 5′-ACATATCTGTGACCAAGCCATAAGTTG-3′, and secondary, 5′-CGTCTGGCAGACAATCTTTGCTAACCAT-3′. Three nested primers for At1g62930 were primary, 5′-CTGACTGCAGAGAACTGTACCAGTCATG-3′; secondary, 5′-GGAAGAGTCCCATCTTCTTTCATTTCTC-3′; and tertiary, 5′-CAAAGGCATCTTATGAGGGAGTTGTAGG-3′.

We thank Dr. Jeng-Min Chiou for valuable assessment of the statistical methods; Dr. Kuo-Chen Yeh and Ms. Nien-Chen Lee for technical assistance; and Drs. Kuo-Chen Yeh, Tzyy-Jen Chiou, and Erh-Min Lai for helpful discussions. This research was supported by Academia Sinica Grant AS91IB1PP (to S.-H.W.).

1. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) *Cell* 123:1279–1291.
2. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) *Cell* 121:207–221.
3. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gascioli V, Mallory AC, Hilbert JL, Bartel DP, Crete P (2004) *Mol Cell* 16:69–79.
4. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE (2006) *Science* 313:363–367.
5. Chen X (2005) *FEBS Lett* 579:5923–5931.
6. Wassenegger M (2005) *Cell* 122:13–16.
7. Zilberman D, Henikoff S (2005) *Curr Opin Genet Dev* 15:557–562.
8. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) *Genes Dev* 18:2368–2379.
9. Williams L, Carles CC, Osmont KS, Fletcher JC (2005) *Proc Natl Acad Sci USA* 102:9703–9708.
10. Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) *Genes Dev* 19:2164–2175.
11. Xie Z, Allen E, Wilken A, Carrington JC (2005) *Proc Natl Acad Sci USA* 102:12984–12989.
12. Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouche N, Gascioli V, Vaucheret H (2006) *Curr Biol* 16:927–932.
13. Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, Carrington JC (2006) *Curr Biol* 16:939–944.
14. Garcia D, Collier SA, Byrne ME, Martienssen RA (2006) *Curr Biol* 16:933–938.
15. Hunter C, Willmann MR, Wu G, Yoshikawa M, de la Luz Gutierrez-Nava M, Poethig SR (2006) *Development (Cambridge, UK)* 133:2973–2981.
16. Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Tej SS, Poethig RS, Henderson IR, Jacobsen SE, Wang W, Green PJ, Meyers BC (2006) *Genome Res* 16:1276–1288.
17. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) *Cell* 127:565–577.
18. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, *et al.* (2005) *Nature* 437:376–380.
19. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ (2005) *Science* 309:1567–1569.
20. Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) *Proc Natl Acad Sci USA* 101:11511–11516.
21. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) *Cell* 110:513–520.
22. Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) *Genome Biol* 5:R65.
23. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) *Nat Genet* 38:721–725.
24. Gustafson AM, Allen E, Givan S, Smith D, Carrington JC, Kasschau KD (2005) *Nucleic Acids Res* 33:D637–D640.
25. Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005) *Curr Biol* 15:1494–1500.
26. Zhang Y (2005) *Nucleic Acids Res* 33:W701–W704.
27. Baumberger N, Baulcombe DC (2005) *Proc Natl Acad Sci USA* 102:11928–11933.
28. Qi Y, Denli AM, Hannon GJ (2005) *Mol Cell* 19:421–428.
29. Brodersen P, Voinnet O (2006) *Trends Genet* 22:268–280.
30. Mallory AC, Vaucheret H (2006) *Nat Genet* 38(Suppl):S31–S36.
31. Vaucheret H (2006) *Genes Dev* 20:759–771.
32. Martienssen RA (2003) *Nat Genet* 35:213–214.
33. Marker C, Zemann A, Terhorst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie JP, Brosius J, Huttenhofer A (2002) *Curr Biol* 12:2002–2013.
34. Talmor-Neiman M, Stav R, Klipcan L, Buxdorf K, Baulcombe DC, Arazi T (2006) *Plant J* 48:511–521.
35. Jacobsen SE, Running MP, Meyerowitz EM (1999) *Development (Cambridge, UK)* 126:5231–5243.