nature structural & molecular biology

# Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*

Josh T Cuperus[1,2,5], Alberto Carbonell[2,5], Noah Fahlgren[1,2], Hernan Garcia-Ruiz[2], Russell T Burke[2], Atsushi Takeda[3,4], Christopher M Sullivan[2,3], Sunny D Gilbert[2,3], Taiowa A Montgomery[1,4] & James C Carrington[2,3]



**RNA interference pathways can involve amplification of secondary siRNAs by RNA-dependent RNA polymerases. In plants, RDR6-dependent secondary siRNAs arise from transcripts targeted by some microRNAs (miRNAs). Here, *Arabidopsis thaliana* secondary siRNAs from mRNA as well as *trans*-acting siRNAs are shown to be triggered through initial targeting by a 22-nucleotide (nt) miRNA that associates with AGO1. In contrast to canonical 21-nt miRNAs, 22-nt miRNAs primarily arise from foldback precursors containing asymmetric bulges. Using artificial miRNA constructs, conversion of asymmetric foldbacks to symmetric foldbacks resulted in the production of 21-nt forms of miR173, miR472 and miR828. Both 21- and 22-nt forms associated with AGO1 and guided accurate slicer activity, but only 22-nt forms were competent to trigger RDR6-dependent siRNA production from target RNA. These data suggest that AGO1 functions differentially with 21- and 22-nt miRNAs to engage the RDR6-associated amplification apparatus.**

During RNA interference (RNAi), double-stranded RNA (dsRNA) is processed by Dicer, a dsRNA-specific RNase III–class RNase, into small ~20- to 30-nt RNA duplexes. Typically, one strand of the duplex preferentially associates with an Argonaute protein to form an effector complex (RNA-induced silencing complex (RISC)) that targets and silences transcripts based on sequence complementarity[1]. In plants, fungi, nematodes and some other organisms, this process can be amplified through the production of secondary short interfering RNAs (siRNAs) after transcription by RNA-dependent RNA polymerase (RdRp) on the primary target RNA[2]. As exemplified in *Caenorhabditis elegans*, primary siRNA, secondary siRNA and other small RNAs may serve as guides in association with distinct Argonaute proteins[3–5]. Additionally, dsRNA precursors or siRNAs have the potential to act non–cell autonomously in plants and some animals, leading to the spread of silencing signals, subsequent amplification and transgenerational silencing in some cases[2,6].

In amplification-competent organisms, the selective recruitment and activity of RdRp on primary targets is a key step that governs whether secondary siRNA production and amplification occurs. The partially redundant RdRps RRF-1 and EGO-1 from *C. elegans,* QDE-1 from *Neurospora crassa,* Rdp1 from *Schizosaccharomyces pombe* and RNA-dependent RNA polymerase 6 (RDR6) from *Arabidopsis thaliana* transcribe RNA in a primer-independent manner[7–12]. *A. thaliana* may encode six RdRps, at least three of which generate long dsRNAs that serve as Dicer-like (DCL) substrates in several distinct pathways[2,13]. RDR6, which frequently functions in combination with DCL4, participates in post-transcriptional silencing of exogenous targets (transgene transcripts, some viral RNA) as well as several types of endogenous transcripts[13,14]. Indiscriminate entry of transcripts into the RDR6-dependent siRNA amplification pathway does not occur in plants. However, amplification of secondary siRNA occurs from a minority of small RNA–targeted transcripts, and the genetic requirements for biogenesis of these siRNAs resemble those of *trans*-acting siRNA (tasiRNA)[15–20]. Tightly controlled entry of transcripts into amplification pathways makes intuitive sense, as feed-forward amplification of silencing signals could conceivably result in runaway suppression of off-target transcripts from related gene family members. In *Arabidopsis*, the XRN family of exonucleases functions to degrade sliced transcript fragments[21], and sequencing analysis using an *xrn4* mutant suggests that XRN4 is antagonistic to secondary siRNA biogenesis[22]. How secondary siRNA–producing transcripts are routed differentially from most other target transcripts is poorly understood.

Plant tasiRNAs represent a case in which an RdRp-dependent, secondary siRNA–producing pathway has evolved to produce specialized small RNAs that function to repress coevolved targets. The four known families of tasiRNA (*TAS1*–*TAS4*) in *Arabidopsis* are particularly informative as experimental systems to understand the formation of secondary siRNA. tasiRNAs form from primary *TAS* transcripts that are initially targeted and sliced by AGO1–miR173 (*TAS1* and *TAS2*), AGO1–miR828 (*TAS4*) or AGO7–miR390 (*TAS3*) complexes[18,23–26].

[1]Molecular and Cellular Biology Program, Oregon State University, Corvallis, Oregon, USA. [2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, USA. [3]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon, USA. [4]Present addresses: Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan (A.T.) and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA (T.A.M.). [5]These authors contributed equally to this work. Correspondence should be addressed to J.C.C. (carrington@cgrb.oregonstate.edu).

Received 25 March; accepted 10 June; published online 18 June 2010; doi:10.1038/nsmb.1866

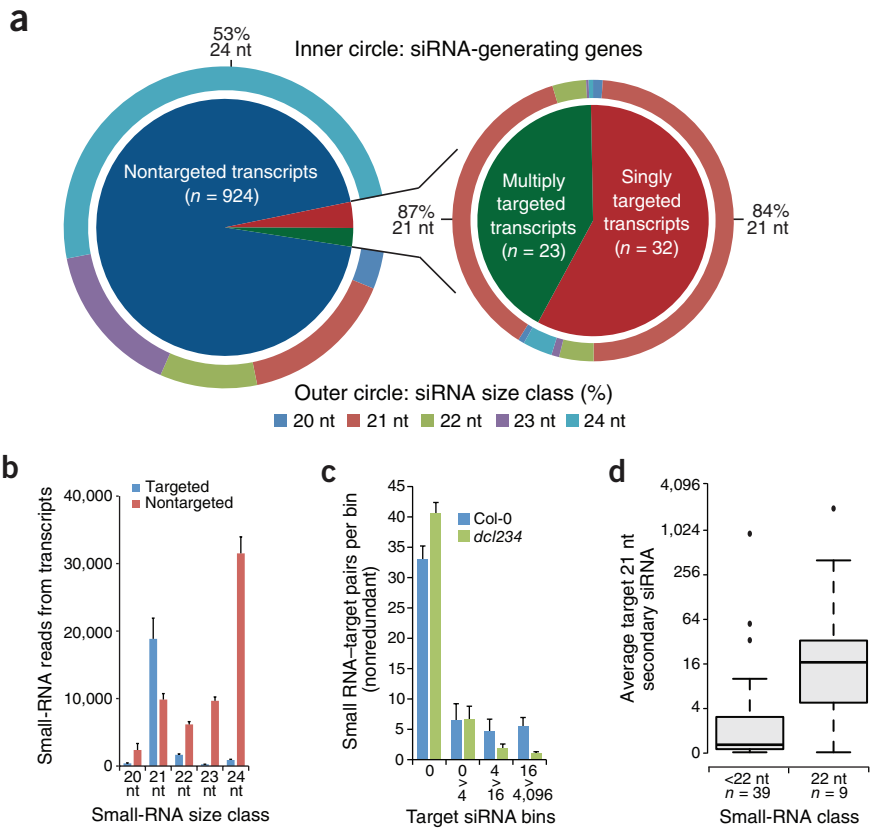


**Figure 1** Small RNAs from *Arabidopsis* annotated transcripts. (**a**) Proportion of small RNA–generating transcripts that are targeted (at single or multiple sites) or not targeted by miRNA or tasiRNA. Outer rings, proportion of small-RNA size. (**b**) Mean 20- to 24-nt siRNA levels from targeted or nontargeted transcripts. (**c**) Nonredundant small RNA–target transcript pairs yielding four levels (bins) of 21-nt siRNA in Col-0 and *dcl234* mutant plants. Data, averages of six (Col-0) and five (*dcl234*) replicates. (**d**) Box plots showing the mean numbers of 21-nt siRNAs originating from nonredundant transcripts targeted by small RNAs that are 22 nt or less in length.

*TAS1*, *TAS2* and *TAS4* tasiRNA-generating transcripts are cleaved at a 5′-proximal site. The 3′ RNA product is then transcribed by RDR6, leading to dsRNA that is sequentially processed by DCL4 to yield phased tasiRNAs in register with the cleavage site[18,23,26–30].

Here, we investigated the basis for the selective entry of miRNA-targeted transcripts through the RDR6-dependent siRNA-generating pathway. Combining genome-wide analyses of 21-nt siRNA and miRNA-targeting patterns with directed experiments, we found miRNA length to be a key determinant in triggering amplification in the context of AGO1–miRNA complexes. In addition, we identified determinants within *MIRNA* foldbacks that govern miRNA size.

## RESULTS

### Secondary siRNAs originate from some targeted mRNAs

We generated six *A. thaliana* small RNA libraries from aerial plant tissue and sequenced them in a multiplexed format, yielding an average of 658,999 reads per library that mapped to at least one genomic position. Among 20- to 24-nt reads, 53.9% were 21 nt in length, the vast majority of which corresponded to miRNAs (**Supplementary Fig. 1**). A significant proportion of 20- to 24-nt reads (74.4%) mapped to 979 annotated transcripts, 55 of which are known to be targeted by one or more miRNA or tasiRNA (**Fig. 1a** and **Supplementary Table 1**). Although targeted transcripts represented only 5.6% of the 979 small RNA–generating transcripts, reads from the targeted set represented 27.0% of reads from the entire transcript set. Among reads mapping to the targeted transcripts, 85.4% were 21 nt in length, whereas the majority (52.9%) of reads derived from nontargeted transcripts were 24 nt in length (**Fig. 1a**,**b**). Previously, 21-nt siRNAs originating from the targeted transcripts were shown to be largely RDR6 and DCL4 dependent[15]. Thus, among *Arabidopsis* siRNAs originating from annotated transcripts, those from targeted transcripts are more likely to yield siRNA of 21 nt in length.

We collated a total of 280 miRNA-target or tasiRNA-target pairs from published data (**Supplementary Table 2**). Given that multigene families with many members targeted by a specific miRNA or tasiRNA family tend to bias representation on this list, we generated 'nonredundant' sets of miRNA-target and tasiRNA-target pairs. One hundred nonredundant set iterations (50 pairs in each) contained only one paralog/target family/small RNA, with the targeted family member chosen randomly for each family in each iteration. Most targeted transcripts from the nonredundant list yielded no 21-nt siRNA (**Fig. 1c** and **Supplementary Table 2**). However, on average, 33.8% of targets yielded at least a few siRNAs, with several targets yielding highly abundant 21-nt siRNA (**Fig. 1c**). Levels of these abundant siRNAs were decreased in the *dcl2-1 dcl3-1 dcl4-2* (termed *dcl234*) triple mutant (**Fig. 1c**).

In exploring the differences between small RNA–target pairs that led to secondary siRNA production and those that did not, it was noticed that secondary siRNA–generating transcripts were frequently targeted by a miRNA or tasiRNA of 22 nt in length (**Fig. 1d**). Targets that interacted with miRNA or tasiRNA containing less than 22 nt yielded significantly ($P = 0.034$, permutation test) less siRNA on average. Additionally, 47% of singly targeted transcripts and 43% of multiply targeted transcripts presented in **Figure 1a** were targeted by a small RNA of predominantly 22 nt in length. We therefore subjected the biogenesis and activity of 22-nt miRNAs to further analysis.

### Asymmetric *MIRNA* foldbacks yield 22-nt miRNA

The vast majority of *Arabidopsis* miRNA reads were either 21 or 22 nt in length, with 22-nt families representing 22.5% of all annotated *MIRNA* families (**Fig. 2a**). However, the overall abundance of 22-nt miRNA reads was relatively low (2.3%). Both the percentage of 22 nt–generating *MIRNA* families and the numbers of 22-nt miRNA reads were comparably low in rice (**Fig. 2a** and **Supplementary Tables 3** and **4**). Size was also a distinct characteristic for most *MIRNA* families in both *Arabidopsis* and rice, with relatively few families expressing a mixed size distribution. Comparing the proportions of 21- or 22-nt miRNA from nonredundant loci in *Arabidopsis* and rice revealed a distinct bimodal distribution: over 90% of all *MIRNA* loci generated miRNA of which >80% were either 21 or 22 nt in length (**Fig. 2b** and **Supplementary Tables 3** and **4**). To understand the basis for 21- versus 22-nt miRNA formation, we examined foldback base pairing patterns of *Arabidopsis MIRNA*. Strikingly, those *Arabidopsis* foldbacks that contained asymmetry in the form of a nonpaired nucleotide within the miRNA sequence were significantly enriched for 22-nt miRNA reads ($P < 2.6 \times 10^{-9}$, Wilcoxon

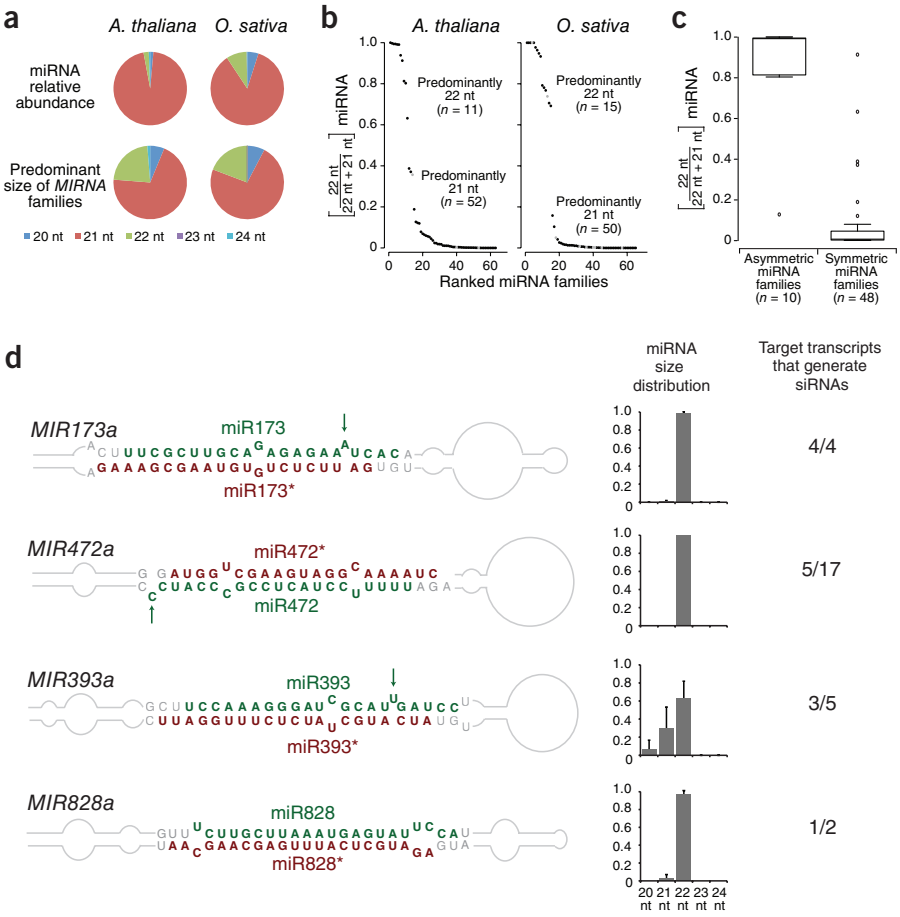


**Figure 2** *MIRNA* foldback asymmetry leads to the formation of 22-nt miRNAs. (**a**) Mean proportions of distinct miRNA size classes in read datasets (top) and of predominant size class for MIRNA families (bottom) from *Arabidopsis* and rice. The rice miRNAs were from a filtered subset that passed basic criteria for bona fide miRNA[57] (see **Supplementary Table 4**). (**b**) Rank order showing proportion of 22-nt size class, from averages of sequencing datasets, corresponding to nonredundant *MIRNA* loci in *Arabidopsis* and rice. Gray, multigene *MIRNA* families with loci encoding the identical miRNA but with both symmetric and asymmetric foldbacks. (**c**) Proportion of 22-nt miRNA from nonredundant *MIRNA* loci with base pair asymmetry or symmetry within the miRNA-miRNA* segment of the foldback. (**d**) Examples of 22-nt miRNA–generating *MIRNA* foldbacks, average miRNA size distribution and proportion of target transcripts that yield 21-nt siRNA (at least four reads from 4–6 replicate libraries). Green arrows, predicted asymmetric position within the foldbacks.

rank sum test) (**Fig. 2c**). In nearly all cases of predominantly 22-nt miRNA, including miR173, miR393 and miR472, we detected an asymmetric nonpaired base. There were a few exceptions, such as miR828, which arose from a foldback containing only symmetric mispairs within the miRNA-miRNA* segment (**Fig. 2d** and **Supplementary Table 3**).

To determine if the asymmetric positions within miRNA sequences of foldbacks led to the formation of the 22-nt miRNA, we expressed artificial miRNAs (amiRNAs) using foldbacks containing either asymmetric (wild-type configuration) or symmetric miRNA-miRNA* segments. We produced these using the *MIR390a* foldback, in which we replaced the miR390-miR390* sequences with miR173-miR173* sequences[25]. We engineered the symmetric foldback by the addition of a uridine residue to base-pair with the normally mispaired adenosine residue within the *MIR173* foldback (**Fig. 3a**). We coexpressed the authentic *MIR173* foldback and the asymmetric and symmetric amiRNA foldbacks with the tasiRNA-generating construct *35S:TAS1c* in transient-expression assays using *Nicotiana benthamiana*[23], and we measured both miRNA and tasiRNA accumulation (see below) in blot assays and by sequence analysis. Neither miR173 nor *TAS1c* tasiRNA are conserved in *Nicotiana* species, so the accumulation of these small RNAs is entirely dependent on the transient-expression assay. The asymmetric foldback from either the authentic (*35S:MIR173*) or artificial (*35S:amiR173*) constructs yielded predominantly 22-nt miR173 in blot assays (**Fig. 3b**). In contrast, the symmetric foldback (*35S:amiR173-21*) yielded a 21-nt miR173 (**Fig. 3b**). Sequence analysis of small RNAs from these assays confirmed that the asymmetric foldbacks yielded predominantly 22-nt miR173, although we sequenced a substantial amount of off-size products of 19 and 20 nt (**Fig. 3c**, **Supplementary Fig. 2** and **Supplementary Table 5**). The symmetric foldback, however, yielded miR173 of which 91% was 21 nt in length (**Fig. 3c**, **Supplementary Fig. 2** and **Supplementary Table 5**). The 21-nt form lacked the 3′-C residue present in authentic miR173.

Several previous reports collectively showed that miR173 formation requires DCL1 but not DCL2 (refs. 30–32). However, DCL2 was shown to produce 22-nt siRNA from endogenous *Arabidopsis* dsRNA[27,29]. To determine if DCL2, or siRNA-generating DCL3 and DCL4, are necessary for 22-nt miRNA, we compared miRNA reads sequenced from Col-0 and the *dcl234* triple mutant. Accumulation and family representation levels of most miRNAs in the *dcl234* triple mutant, regardless of size, were similar to those in wild-type Col-0 plants (**Fig. 2a** and **Supplementary Fig. 3a**,**b**). miRNA reads from Col-0 and the *dcl234* mutant plants for most families, except for two (miR822 and miR839, which are known to be DCL4 dependent[18]), were highly correlated (Pearson product-moment correlation coefficient $r = 0.95$) (**Supplementary Fig. 3b**). A high correlation ($r = 0.94$) was shown when we compared only 22-nt miRNA reads (**Supplementary Fig. 3b**). Therefore, all known 22-nt miRNA in *Arabidopsis* are likely DCL1 dependent.

## 22-nt forms of miR173, miR472 and miR828 trigger siRNAs

As shown previously[23,30] and in **Figure 3b**, coexpression of *35S:MIR173* with *35S:TAS1c* triggered tasiRNA biogenesis. Authentic-size miR173 generated from *35S:amiR173* also triggered *TAS1c* tasiRNA formation (**Fig. 3b**, lane 5). The tasiRNA pools triggered by both 22-nt miR173 sources were highly phased, with the register set by the miR173 cleavage site (**Fig. 3c**, **Supplementary Fig. 2** and **Supplementary Table 6**). However, the 21-nt form produced from the amiR173-21 foldback possessed very low tasiRNA trigger activity, and the few *TAS1c*-derived siRNA that were produced showed no phasing pattern (**Fig. 3b**,**c**, **Supplementary Fig. 2** and **Supplementary Table 6**).

The lack of tasiRNA-trigger activity of miR173-21 could be for one or more reasons, including (i) failure to associate with AGO1, (ii) failure to interact with and guide cleavage of the *TAS1c* primary transcript or (iii) loss of post-cleavage functions necessary to recruit RDR6 to

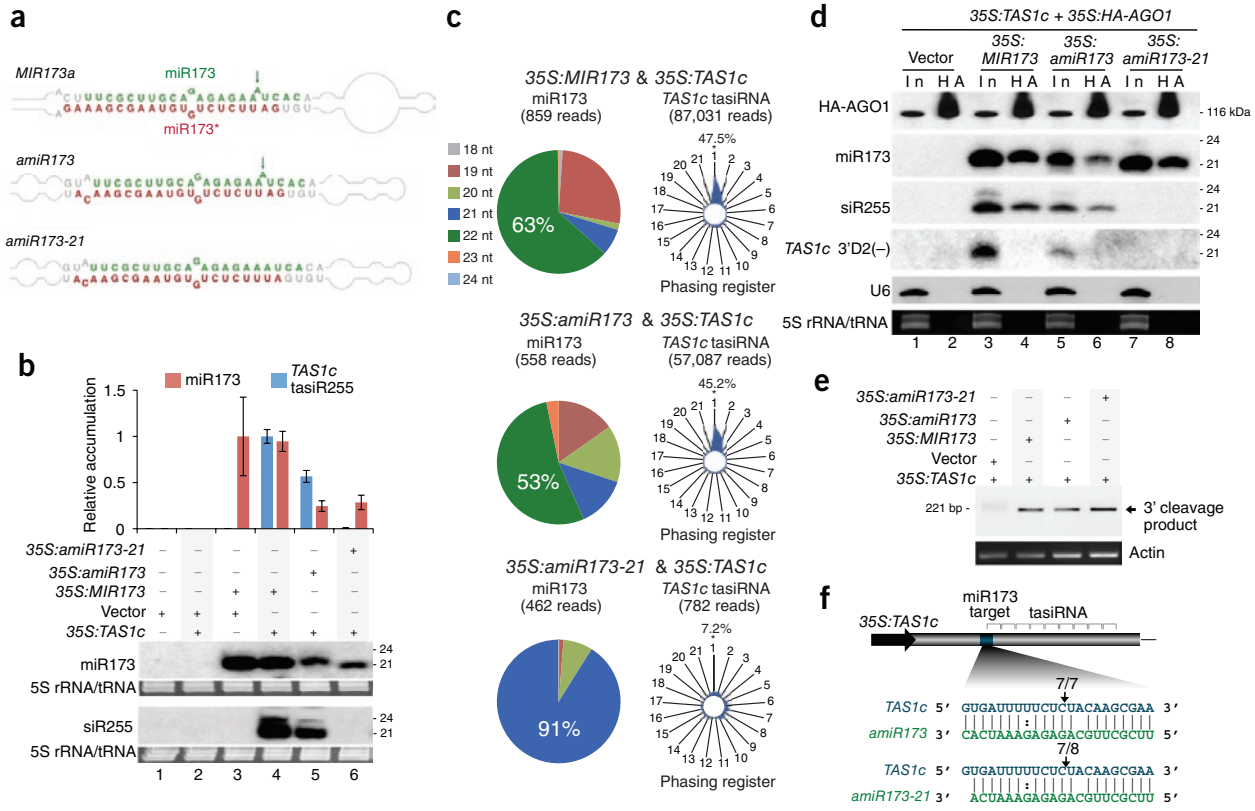


**Figure 3** Production and activities of 21- and 22-nt miR173 forms. (**a**) Foldbacks of wild-type *MIR173, amiR173* and *amiR173-21*. Artificial miRNAs were engineered within the *MIR390a* foldback. Green and red, miRNA guide and miRNA* strands, respectively. Arrows, predicted asymmetric position in *MIR173* and *amiR173* foldbacks. (**b**) Accumulation of miR173 and *TAS1c* tasiRNA (siR255) in *N. benthamiana* transient assays. Constructs were coexpressed as indicated above the blot panels. Top, mean ($n = 3$) relative miR173 (red) and siR255 (blue) levels ± s.d. (lane 2 and lane 3 = 1.0 for miR173 and tasiRNA255, respectively). Bottom, one of three biological replicates of the blot data as well as ethidium bromide (EtBr)-stained rRNA as loading controls. (**c**) Analysis of miR173 (from *35S:MIR173*, *35S:amiR173* and *35S:amiR173-21*) and *TAS1c*-derived siRNA sequences by high-throughput sequencing after transient assays in *N. benthamiana*. Pie charts, percentage of 18- to 24-nt reads; radar plots, percentages of 21-nt reads corresponding to each of the 21 registers from *TAS1c* transcripts, with position 1 designated as immediately after the miR173-guided cleavage site. (**d**) Analysis of co-immunoprecipitation of 21- and 22-nt miR173 variants with HA-AGO1. Protein and RNA blot assays using input (in) and immunoprecipitated (HA) fractions from *N. benthamiana* following coexpression of *35S:HA-AGO1* and *35S:TAS1c* with *35S:MIR173*, *35S:amiR173* and *35S:amiR173-21*. The *TAS1c* 3′D2(–) panel shows an HA-AGO1–nonassociated tasiRNA generated from the *TAS1c* transcript as an immunoprecipitation control. U6 RNA and EtBr-stained rRNA were included as input loading and HA-AGO1–nonassociated controls. (**e**) EtBr-stained 5′ RACE products corresponding to the 3′ cleavage product from miR173-guided cleavage. *N. benthamiana* actin RT-PCR products are shown as a control. (**f**) Proportion of cloned 5′ RACE products corresponding to cleavage within *TAS1c* transcripts at the canonical miR173-guided site in assays with *amiR173* and *amiR173-21*.

the precursor transcript. Previous work[33] showed that, among AGO1, AGO2, AGO4 and AGO5, miR173 associated most commonly with AGO1. To confirm that both the 21- and 22-nt forms of miR173 associated with AGO1, we performed *TAS1c* biogenesis assays in *N. benthamiana* with coexpression of HA-tagged AGO1, followed by co-immunoprecipitation analysis. Both the 22- and 21-nt forms of miR173 co-immunoprecipitated with HA-AGO1 (**Fig. 3d**, lanes 4, 6 and 8). *TAS1c* tasiR255 produced in the presence of 22-nt miR173 also co-immunoprecipitated with HA-AGO1 (**Fig. 3d**, lanes 4 and 6). As HA-AGO1–nonassociated controls, we analyzed the accumulation of *TAS1c* 3′D2(–) tasiRNA, which contains an AGO1-nonpreferred 5′-A and U6 RNA. We detected no or very low levels of *TAS1c* 3′D2(–) tasiRNA and U6 RNA in the immunoprecipitated samples (**Fig. 3d**), indicating that HA-AGO1 was selective. To assess more broadly whether AGO1 associates preferentially with 21- or 22-nt miRNA, we identified small RNAs from total RNA (input fraction) and immunoprecipitated HA-AGO1 complexes by high-throughput sequencing (two replicates each). Of the miRNAs that were enriched at least two-fold in the immunoprecipitated fraction, miRNA families that were predominantly 20–24 nt were recovered, suggesting that AGO1 does not include or exclude miRNA based on length (**Supplementary Fig. 4**). All predominantly 22-nt miRNA families that met read thresholds for inclusion in the enrichment calculation were enriched in the immunoprecipitated fraction. Additionally, for two predominantly 21-nt miRNAs (miR397a and miR167a,b,d) that had a >30% subpopulation of 22-nt variants (**Fig. 2b**), both the 21- and 22-nt variants were enriched more than two-fold in the immunoprecipitated fraction (**Supplementary Fig. 4**).

To assess the guide function of the 22- and 21-nt miR173 forms, we performed rapid amplification of 5′ complementary DNA ends (5′ RACE) assays for each coexpression assay using primers to detect the 3′ product of miR173-guided cleavage. We detected products in all three coexpression assays, and sequencing confirmed that cleavage guided by both 22- and 21-nt miR173 forms occurred at the canonical target site (**Fig. 3e**,**f**). Furthermore, both the 22- and 21-nt miR173 forms destabilized *TAS1c*

full-length transcripts to very low or undetectable levels (**Supplementary Fig. 5**).

Because the 21–nt form derived from the *35S:amiR173-21* foldback contained a 3′-A rather than a 3′-C residue, as in 22-nt authentic miR173, we examined the effect of 3′-nucleotide identity on both the 21- and 22-nt forms of miR173. First, we tested a construct (*35S:amiR173-21–3′-C*) that generated a symmetric foldback yielding a 21-nt form containing a 3′-C (**Supplementary Fig. 5a**). This 21-nt miR173 failed to trigger tasiRNA accumulation despite guiding accurate cleavage of the *TAS1c* transcript (**Supplementary Fig. 5b**,**c**, lane 2). Second, to assess the possibility of a preference for a specific 3′-nucleotide in a 22-nt context, we tested 22-nt miR173 variants containing 3′-G, 3′-A or 3′-U. Each of the variants accumulated to within 36% of the level of 22-nt miR173 containing a natural 3′-C and guided accurate cleavage of the *TAS1c* transcript (**Supplementary Fig. 5b**,**c**). Notably, although each of the 22-nt 3′-G, 3′-A and 3′-U variants triggered tasiRNA accumulation, the levels of tasiRNA generated were relatively low (**Supplementary Fig. 5b**).

To determine if the 22-nt requirement to trigger tasiRNA is specific to miR173, we generated 22- and 21-nt artificial miRNAs for both miR472 and miR828 using the *MIR390a*-based foldback and tested these using the *TAS1c* tasiRNA transient-expression assay system. We engineered both asymmetric and symmetric foldbacks using the same strategy as that for the miR173 variants (**Fig. 4a**) and coexpressed them with modified *TAS1c* constructs (*35S:TAS1c-472* and *35S:TAS1c-828*) in which we substituted the normal miR173 target site with sites recognized by miR472 or miR828. Based on the mobility of miR472 and miR828 forms in blot assays, 22- and 21-nt variants were generated as predicted from asymmetric and symmetric foldbacks, respectively, although the asymmetric miR828 foldback yielded a mixture of both size classes (**Fig. 4b**). Notably, we detected tasiRNA only in those samples containing 22-nt variants of miR472 and miR828 (**Fig. 4b**). Like the miR173 size variants, both the 22- and 21-nt forms of miR472 and miR828 guided accurate cleavage of the engineered tasiRNA transcripts (**Fig. 4b**,**c**) and co-immunoprecipitated with HA-AGO1 (**Fig. 4d**). These data clearly show that 22-nt forms of three *Arabidopsis* miRNAs possess unique functionality to direct tasiRNA or secondary siRNA biogenesis.

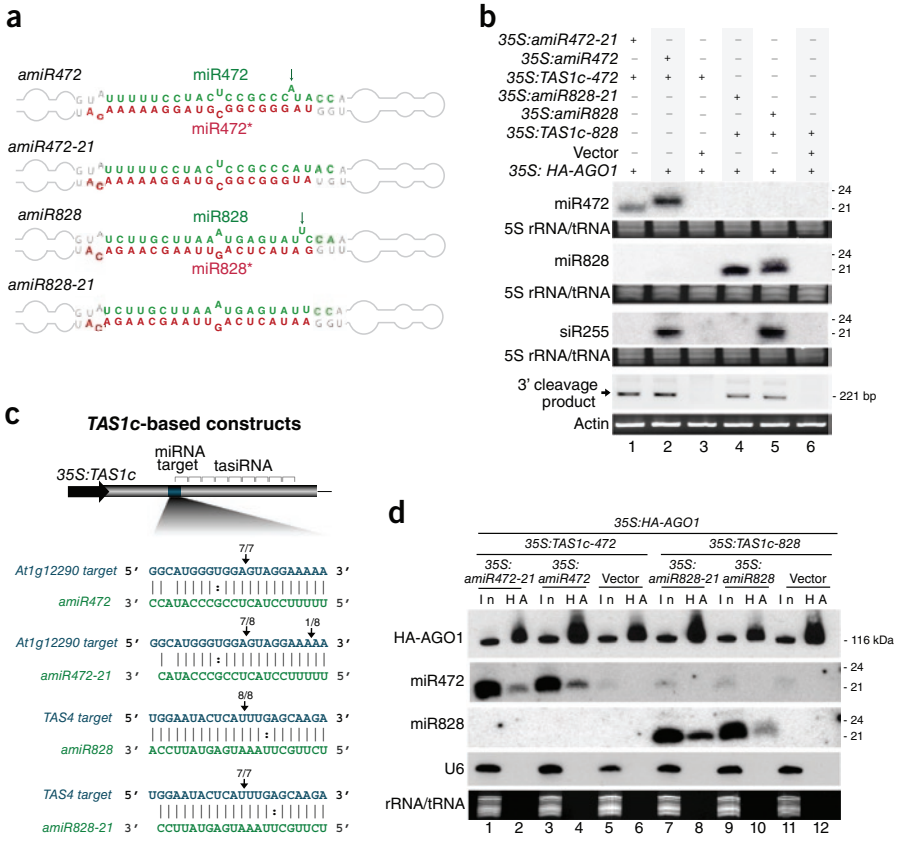


**Figure 4** Production and activities of 21- and 22-nt miR472 and miR828 forms. (**a**) Foldbacks of *amiR828*, *amiR828-21*, *amiR472* and *amiR472-21*. (**b**) Accumulation of miR472, miR828 and modified *TAS1c* tasiRNA (siR255) as well as 5′ RACE to detect miRNA-guided cleavage products of the modified *TAS1c* transcripts in *N. benthamiana* transient assays. (**c**) Proportion of cloned 5′ RACE products corresponding to cleavage within modified *TAS1c* transcripts at the canonical miR472- or miR828-guided sites in assays with the designated artificial miRNAs. The target-site sequences are actual sites from *At1g12290* and *TAS4* transcripts, which are recognized by miR472 and miR828, respectively. (**d**) Analysis of co-immunoprecipitation of 21- and 22-nt amiR472 and amiR828 variants with HA-AGO1. Protein and RNA assays for input (in) and immunoprecipitated (HA) fractions from *N. benthamiana* expressing the amiR472 and amiR828 variants were performed using blots containing samples from both sets of experiments. U6 RNA and EtBr-stained rRNA were included as input loading and HA-AGO1–nonassociated controls.

## DISCUSSION

This study associated a unique class of *Arabidopsis* miRNA—those that are 22 nt in length—with triggering formation of secondary, RDR6- or DCL4-dependent 21-nt siRNAs from primary miRNA targets. Although both the 22- and 21-nt forms of miR173, miR472 and miR828 associated with AGO1 and guided accurate cleavage of target transcripts, advancement of target RNA fragments through the 21-nt siRNA biogenesis pathway was a unique property of the 22-nt miRNA–target RNA pairs. Previous studies provided hints that 22-nt miRNAs might be associated with secondary siRNA. *Arabidopsis MIR168b* was shown to generate both 21- and 22-nt forms of miR168 (ref. 34), which targets the transcript encoding AGO1 and leads to secondary siRNAs in a homeostatic regulatory loop[34,35]. In rice, 22-nt miRNA target sites were identified in several abundant siRNA-generating transcripts[36], suggesting that the 22-nt property may be conserved in angiosperms.

The fact that most plant miRNAs, particularly those from the most highly expressed families, are 21 nt in length may explain why secondary siRNAs from most targeted transcripts are absent or in low abundance. Although the tasiRNAs represent highly refined examples of functional, discrete secondary siRNAs from a noncoding RNA, it is interesting to consider potential functions for the cases in which abundant siRNAs originate from protein-coding transcripts. In at least some examples, secondary siRNAs might specifically target transcripts from related family members. For gene families that are evolving or expanding rapidly, secondary targeting may suppress dosage effects. Evidence for the functionality of secondary siRNA from large multigene family transcripts has been presented[15,37,38].

It is proposed that 22-nt miRNA–AGO1 complexes but not 21-nt miRNA–AGO1 complexes mark transcripts for RDR6-dependent

siRNA formation. This could involve AGO1 adopting one of two states, an amplification-trigger state or a nontrigger state. Both states would be competent to interact with and slice targets containing a suitable target site, but only the trigger state would recruit RDR6 to the decapped (sliced) target. Previous results from experiments with a noncleavable target of miR173 suggested that secondary siRNA formation depends on a functional cleavage site[30]. Several studies have also shown that all known tasiRNA as well as many other RDR6-dependent siRNA in *Arabidopsis*[15,18,19,23] preferentially originate adjacent to target cleavage sites. Therefore, recruitment of RDR6 may depend on target cleavage.

Conformational changes of well-characterized Argonaute proteins are known to occur upon interaction with target RNA and release of the guide 3′ end from the PAZ domain[39,40]. How a 22-nt RNA would effect an AGO1 conformation distinct from that effected by a 21-nt RNA is not immediately obvious. However, the proposed trigger state could involve a direct interaction with RDR6 or associated factors, such as SGS3 (refs. 16,26). To date, there are no data suggesting direct interaction between AGO1 and RDR6 or known RDR6-associated factors. Alternatively, conformationally distinct AGO1–small RNA complexes may interact differentially with Gly-Trp (GW) domain proteins, which interact directly with AGO proteins as scaffolds to mediate interactions with other factors to suppress translation and promote deadenylation[41]. *Arabidopsis* AGO1-specific GW proteins are not yet characterized, although AGO4 likely interacts with GW domains in an RNA polymerase V subunit (NRPDE1) and the transcription elongation factor SPT5L[42–44].

Although it is clear that AGO1 shows a preference for small RNAs with a 5′-U, which docks within a MID domain binding pocket, there are no clear structural or functional data supporting a 3′-nucleotide preference for interaction with the PAZ domain. Among 22-nt miRNAs in *Arabidopsis*, a 3′-C is most common, although this implies neither a binding preference nor a particular functionality. Within a 22-nt context, a 3′-C promoted siRNA trigger function of miR173 most effectively. However, there was no strict requirement for any specific 3′-nucleotide, and the 22-nt siRNA triggers miR828 and miR393 both contain 3′-A. The studies presented here do not allow quantitative comparison of the affinity of AGO1 for 21- versus 22-nt miRNAs or for 22-nt miRNAs with different 3′-nucleotides.

The finding that 22-nt miRNAs trigger RDR6-dependent secondary siRNAs leads to the question of whether other siRNA amplification pathways have guide RNA–specific activities to recruit RDR1 or RDR2. Signals for amplification mediated by RDR1 and RDR6 during antiviral RNAi are not known, although the results presented here suggest that 22-nt siRNAs, working through AGO1 or other AGO proteins, might be plausible triggers. As DCL2 is known to both promote a full silencing response against several viruses[32,45–48] and to generate 22-nt siRNAs[27,29], this idea may have support. DCL2 was also shown to promote DCL4-dependent post-transcriptional silencing of several transgenes[49]. Furthermore, overproduction of DCL2-dependent 22-nt siRNA may account for the severe phenotypes of *dcl134* and *dcl14* mutant *Arabidopsis* plants, in which DCL2 might generate 22-nt siRNA from dsRNA that would normally be processed to 21-nt siRNA by DCL4 (ref. 32).

Finally, the basis for the production of either a 21- or a 22-nt miRNA depends on the nature of the foldback in plants. Most *Arabidopsis* miRNAs are formed by DCL1-mediated processing first at the loop-distal end of the foldback, then at the loop proximal end[50–52]. DCL1, like other Dicers, functions as a molecular ruler. For the loop-proximal cuts, DCL1 'measures' the length of helical RNA from the PAZ domain–bound end to the positions juxtaposed to the active centers[53–55]. For foldbacks that are symmetrically paired within the miRNA-miRNA* region, it is inferred that DCL1 measures a length equivalent to 21 base pairs in most cases. However, it is also inferred that the length of an A-form helix containing a single asymmetric bulge is frequently equivalent to that of a symmetrically paired helix. Three-dimensional modeling of 21 nt– and 22 nt–generating *MIR173* foldbacks using MC-Fold and MC-Sym[56] supports this idea. The most likely models show that an unpaired nucleotide does not substantially increase the length of the miRNA-miRNA* duplex region (**Supplementary Fig. 6**). This explains why, when the asymmetric nucleotide occurs within the miRNA sequence, the resulting miRNA contains a noncanonical 22-nt length. Given the unique functionality of 22-nt miRNA to trigger siRNA amplification, it seems reasonable to propose that foldback structures with asymmetric positions, and that would result in 22-nt miRNA, should be under evolutionary constraints.

## METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

**Accession codes**. High-throughput sequencing datasets for HA-AGO1-associated small RNA and input controls were deposited in Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under the series accession GSE22252. Datasets for Col-0 and *dcl2-1 dcl3-1 dcl4-2* small-RNA libraries were published previously[47] and are available under the series accession GSE20197. As described in **Supplementary Methods**, 28 *O. sativa* Nipponbare libraries were also used.

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

### ACKNOWLEDGMENTS

A.C. was supported by a postdoctoral fellowship from the Spanish Ministerio de Ciencia e Innovacion (BMC-2008-0188). This work was supported by grants from the US National Science Foundation (MCB-0618433 and MCB-0956526), the US National Institutes of Health (AI43288) and Monsanto Corporation.

### AUTHOR CONTRIBUTIONS

J.T.C., A.C., N.F. and J.C.C. designed the experiments; J.T.C., A.C., N.F., H.G.-R., R.T.B., A.T., C.M.S., S.D.G. and T.A.M. performed the experiments; J.T.C., A.C., N.F., S.D.G., T.A.M. and J.C.C. analyzed the data; J.T.C., A.C., N.F. and J.C.C. wrote the paper.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Kim, V.N., Han, J. & Siomi, M.C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
2. Ghildiyal, M. & Zamore, P.D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).
3. Guang, S. *et al.* An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321**, 537–541 (2008).
4. Gu, W. *et al.* Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol. Cell* **36**, 231–244 (2009).
5. Claycomb, J.M. *et al.* The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**, 123–134 (2009).
6. Carthew, R.W. & Sontheimer, E.J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
7. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
8. Makeyev, E.V. & Bamford, D.H. Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* **10**, 1417–1427 (2002).
9. Sijen, T., Steiner, F.A., Thijssen, K.L. & Plasterk, R.H. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315**, 244–247 (2007).
10. Pak, J. & Fire, A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**, 241–244 (2007).
11. Motamedi, M.R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
12. Correa, R.L., Steiner, F.A., Berezikov, E. & Ketting, R.F. MicroRNA-directed siRNA biogenesis in *Caenorhabditis elegans*. *PLoS Genet.* **6**, e1000903 (2010).
13. Voinnet, O. Use, tolerance and avoidance of amplified RNA silencing by plants. *Trends Plant Sci.* **13**, 317–328 (2008).
14. Mallory, A.C., Elmayan, T. & Vaucheret, H. MicroRNA maturation and action–the expanding roles of ARGONAUTEs. *Curr. Opin. Plant Biol.* **11**, 560–566 (2008).

15. Howell, M.D. *et al.* Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**, 926–942 (2007).
16. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L. & Poethig, R.S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**, 2368–2379 (2004).
17. Vazquez, F. *et al.* Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16**, 69–79 (2004).
18. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D.P. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis* thaliana. *Genes Dev.* **20**, 3407–3425 (2006).
19. Axtell, M.J., Jan, C., Rajagopalan, R. & Bartel, D.P.A. Two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565–577 (2006).
20. Ronemus, M., Vaughn, M.W. & Martienssen, R.A. MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by argonaute, dicer, and RNA-dependent RNA polymerase in *Arabidopsis*. *Plant Cell* **18**, 1559–1574 (2006).
21. Chiba, Y. & Green, P.J. mRNA degredation machinery in plants. *J. Plant Biol.* **52**, 114–124 (2009).
22. Gregory, B.D. *et al.* A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev. Cell* **14**, 854–866 (2008).
23. Allen, E., Xie, Z., Gustafson, A.M. & Carrington, J.C. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121**, 207–221 (2005).
24. Baumberger, N. & Baulcombe, D.C. *Arabidopsis* ARGONAUTE1 is an RNA slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. USA* **102**, 11928–11933 (2005).
25. Montgomery, T.A. *et al.* Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 *trans*-acting siRNA formation. *Cell* **133**, 128–141 (2008).
26. Yoshikawa, M., Peragine, A., Park, M.Y. & Poethig, R.S. A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* **19**, 2164–2175 (2005).
27. Xie, Z., Allen, E., Wilken, A. & Carrington, J.C. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **102**, 12984–12989 (2005).
28. Dunoyer, P., Himber, C. & Voinnet, O. DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat. Genet.* **37**, 1356–1360 (2005).
29. Gasciolli, V., Mallory, A.C., Bartel, D.P. & Vaucheret, H. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr. Biol.* **15**, 1494–1500 (2005).
30. Montgomery, T.A. *et al.* AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc. Natl. Acad. Sci. USA* **105**, 20055–20062 (2008).
31. Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
32. Bouche, N., Lauressergues, D., Gasciolli, V. & Vaucheret, H. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* **25**, 3347–3356 (2006).
33. Mi, S. *et al.* Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* **133**, 116–127 (2008).
34. Vaucheret, H. AGO1 homeostasis involves differential production of 21-nt and 22-nt miR168 species by MIR168a and MIR168b. *PLoS One* **4**, e6442 (2009).
35. Mallory, A.C. & Vaucheret, H. ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. *EMBO Rep.* **10**, 521–526 (2009).
36. Johnson, C. *et al.* Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res.* **19**, 1429–1440 (2009).
37. Chen, H.M., Li, Y.H. & Wu, S.H. Bioinformatic prediction and experimental validation of a microRNA-directed tandem *trans*-acting siRNA cascade in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **104**, 3318–3323 (2007).
38. Addo-Quaye, C., Eshoo, T.W., Bartel, D.P. & Axtell, M.J. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* **18**, 758–762 (2008).
39. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
40. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754–761 (2009).
41. Eulalio, A., Tritschler, F. & Izaurralde, E. The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. *RNA* **15**, 1433–1442 (2009).
42. El-Shami, M. *et al.* Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**, 2539–2544 (2007).
43. Bies-Etheve, N. *et al.* RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep.* **10**, 649–654 (2009).
44. He, X.J. *et al.* An effector of RNA-directed DNA methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**, 498–508 (2009).
45. Blevins, T. *et al.* Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.* **34**, 6233–6246 (2006).
46. Deleris, A. *et al.* Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* **313**, 68–71 (2006).
47. Diaz-Pendon, J.A., Li, F., Li, W.X. & Ding, S.W. Suppression of antiviral silencing by cucumber mosaic virus 2b protein in *Arabidopsis* is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**, 2053–2063 (2007).
48. Garcia-Ruiz, H. *et al. Arabidopsis* RNA-dependent RNA polymerases and Dicer-like proteins in antiviral defense and small interfering RNA biogenesis during turnip mosaic virus infection. *Plant Cell* **22**, 481–496 (2010).
49. Mlotshwa, S. *et al.* DICER-LIKE 2 plays a primary role in transitive silencing of transgenes in *Arabidopsis*. *PLoS One* **3**, e1755 (2008).
50. Mateos, J.L., Bologna, N.G., Chorostecki, U. & Palatnik, J.F. Identification of MicroRNA processing determinants by random mutagenesis of *Arabidopsis* MIR172a precursor. *Curr. Biol.* **20**, 49–54 (2010).
51. Song, L., Axtell, M.J. & Fedoroff, N.V. RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. *Curr. Biol.* **20**, 37–41 (2010).
52. Werner, S., Wollmann, H., Schneeberger, K. & Weigel, D. Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Curr. Biol.* **20**, 42–48 (2010).
53. Macrae, I.J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
54. MacRae, I.J., Zhou, K. & Doudna, J.A. Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.* **14**, 934–940 (2007).
55. Qin, H. *et al.* Structure of the *Arabidopsis* thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA* **16**, 474–481 (2010).
56. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55 (2008).
57. Meyers, B.C. *et al.* Criteria for annotation of plant microRNAs. *Plant Cell* **20**, 3186–3190 (2008).

## ONLINE METHODS

**Plant materials and growth conditions.** *N. benthamiana* and *A. thaliana* plants, including the previously described *dcl234* triple mutant[48], were grown in normal greenhouse conditions with supplemental light on a 16 h light/8 h dark cycle.

**Transgene constructs.** *35S:TAS1c*- and *35S:MIR173a*-derived constructs were generated in pMDC32 and described previously[23,25,30,58]. *MIR390a*-derived artificial miRNA constructs were designed by ligating overlapping oligonucleotides into a pMDC32-derived vector containing ~200 base pairs upstream and downstream of the *MIR390a* foldback.

**Transient expression assays.** Transient expression assays in *N. benthamiana* leaves were carried out as previously described[59] with *Agrobacterium tumefaciens* GV3101. Bacterial cultures were resuspended at $OD_{600}$ = 1.0. When multiple constructs were coexpressed, equal amounts of each culture were used. Plants were grown in the greenhouse after infiltration, and zones of infiltration were harvested 48 h after infiltration for RNA isolation.

**RNA blot assays.** Total RNA from *Arabidopsis* and *N. benthamiana* was isolated using TRIzol reagent (Invitrogen). Two chloroform extractions were performed, and RNA was precipitated in an equal volume of isopropanol for 20 min. Triplicate samples from pools of *N. benthamiana*–infiltrated leaves were analyzed. RNA blot assays were performed as described[30]. Briefly, 5 μg, 10 μg or 20 μg of total normalized RNA was resolved by denaturing polyacrylamide gel electrophoresis for small RNA analysis, or 5 μg of total RNA was resolved by denaturing 1.5% agarose gel electrophoresis. RNA was transferred to positively charged nitrocellulose membrane. DNA or LNA probes were end-labeled using [$^{32}$]ATP and Optikinase (USB). Probes were hybridized to RNA on membranes in Sigma Perfect-Hyb at 68 °C (HMW blots) or 38–42 °C (small RNA blots). An Instant Imager (Packard Bioscience) was used to measure blot hybridization signals.

**Small-RNA sequencing.** Small-RNA amplicons from transient assays in *N. benthamiana* were prepared as described[58,60]. Four synthetic oligoribonucleotides were added in different amounts (Std11, 0.0001 pmol; Std6, 0.001 pmol; Std3, 0.01 pmol; Std2, 0.1 pmol) to 90 μg total RNA per sample before amplicon preparation. Sequencing-by-synthesis was performed with 5 pmol of each amplicon using an Illumina Genome Analyzer II (GAII, www.illumina.com). Reads were parsed using the first six nucleotides of the 3′ adaptor (CTGTAG). Reads were normalized per million as described[60]. Read proportions were based on total reads (18–24 nt) that matched perfectly within a 29-base window surrounding the miR173 or miR173* sequences from samples containing *35S:MIR173a*, *35S:amiR173* and *35S:amiR173-21*. Radar plots to display phasing from *35S:TAS1c* were generated as previously described[19,25]. Small-RNA amplicons were generated and analyzed from HA-AGO1 immunoprecipitated samples as described[25].

**Small-RNA analysis.** Small RNA from *A. thaliana* were mapped to the genome and annotated transcriptome (TAIR9; http://www.arabidopsis.org), including annotated *MIRNA* precursors (miRBase v14; http://www.miRBase.org), using the CASHX pipeline[60]. Those transcripts to which four or more small RNA reads per million were mapped in at least four of the six libraries were defined as small RNA–generating transcripts. For analysis of small RNA–generating transcripts, annotated structural RNA, pseudogenes or transposable elements were removed.

**Statistical analyses.** All statistical analyses were performed using R v2.9.2. Permutation tests were done using the R 'twotPermutation' function, with 1 million simulations, from the DAAG package[61]. A Wilcoxon rank sum test was used to compare empirical logits for the ratio of 22- to 21-nt miRNA reads. The rank sum test was done using the R 'wilcox.test' function from the stats package.

**5′ RACE.** Modified RNA ligase–mediated 5′ RACE was performed as described[30] with the Generacer kit (Invitrogen) using the following gene-specific primers: TAS1c_707_5′_RACE (GATGATGCTTCTTCGCTACACCTCGGAG) and TAS1c_573_5′_RACE (AGCAACTGTTCTTTAGACGACTTGAAAATCTCAT). We gel-purified 5′ RACE products using Invitrogen PureLink gel extraction kit, cloned them in TOPO TA (Invitrogen), introduced them into *E. coli*, screened them for inserts and sequenced them.


58. Cuperus, J.T. *et al*. Identification of MIR390a precursor processing-defective mutants in *Arabidopsis* by direct genome sequencing. *Proc. Natl. Acad. Sci. USA* **107**, 466–471 (2010).
59. Llave, C., Xie, Z., Kasschau, K.D. & Carrington, J.C. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
60. Fahlgren, N. *et al*. Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA* **15**, 992–1002 (2009).
61. Maindonald, J.H. & Braun, J. *Data Analysis and Graphics Using R: An Example-Based Approach*. (Cambridge University Press, New York, USA, 2007).