RNA
A PUBLICATION OF THE RNA SOCIETY

Small RNAs derived from snoRNAs

Ryan J. Taft, Evgeny A. Glazov, Timo Lassmann, et al.

RNA 2009 15: 1233-1240 originally published online May 27, 2009
Access the most recent version at doi:10.1261/rna.1528909

Small RNAs derived from snoRNAs

RYAN J. TAFT,[1] EVGENY A. GLAZOV,[2] TIMO LASSMANN,[3,4] YOSHIHIDE HAYASHIZAKI,[3,4] PIERO CARNINCI,[3,4] and JOHN S. MATTICK[1]

[1]Australian Research Council Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia
[2]Diamantina Institute for Cancer, Immunology and Metabolic Medicine, The University of Queensland, Princess Alexandra Hospital, Woolloongabba, QLD 4102, Australia
[3]Genome Exploration Research Group, RIKEN Omics Science Center and the Functional RNA Research Program, Frontier Research System, RIKEN Yokohama Institute, Yokohama, Kanagawa, 230-0045 Japan
[4]Genome Science Laboratory, Discovery and Research Institute, RIKEN Wako Institute, 2-1 Hirosawa, Wako-shi, Saitama, 351-0198 Japan

ABSTRACT

Small nucleolar RNAs (snoRNAs) guide RNA modification and are localized in nucleoli and Cajal bodies in eukaryotic cells. Components of the RNA silencing pathway associate with these structures, and two recent reports have revealed that a human and a protozoan snoRNA can be processed into miRNA-like RNAs. Here we show that small RNAs with evolutionary conservation of size and position are derived from the vast majority of snoRNA loci in animals (human, mouse, chicken, fruit fly), *Arabidopsis*, and fission yeast. In animals, sno-derived RNAs (sdRNAs) from H/ACA snoRNAs are predominantly 20–24 nucleotides (nt) in length and originate from the 3′ end. Those derived from C/D snoRNAs show a bimodal size distribution at ~17–19 nt and >27 nt and predominantly originate from the 5′ end. SdRNAs are associated with AGO7 in *Arabidopsis* and Ago1 in fission yeast with characteristic 5′ nucleotide biases and show altered expression patterns in fly *loquacious* and *Dicer-2* and mouse *Dicer1* and *Dgcr8* mutants. These findings indicate that there is interplay between the RNA silencing and snoRNA-mediated RNA processing systems, and that sdRNAs comprise a novel and ancient class of small RNAs in eukaryotes.

Keywords: small nucleolar RNA; microRNA; deep sequencing; Argonaute; Dicer

INTRODUCTION

Small nucleolar RNAs are a highly evolutionarily conserved class of RNAs, which are present throughout the Eukaryotes and whose origins lie in the Archaea (Gaspin et al. 2000; Omer et al. 2000; Matera et al. 2007). There are two classes of snoRNAs (C/D and H/ACA box) that function as ribonucleoprotein (RNP) complexes to guide the enzymatic modification of target RNAs at sites determined by RNA:RNA antisense interactions (Matera et al. 2007). Generally, C/D box snoRNAs are ~70–120 nucleotides (nt) and guide the methylation of target RNAs, while H/ACA box snoRNAs are ~100–200 nt and guide pseudouridulation (Matera et al. 2007). These RNAs were initially discovered in the nucleolus and thought to exclusively target ribosomal RNAs, but are now recognized to be a much broader class of RNAs with different functions, targets, and subcellular locations (Matera et al. 2007).

Argonaute proteins are key players in the RNA silencing pathways, including the microRNA (miRNA), small interfering RNA (siRNA), and PIWI-associated RNA-(piRNA) mediated regulation of gene expression (Peters and Meister 2007; Hutvagner and Simard 2008). Recent proteomic and immunofluorescence analyses of Argonaute-containing RNA–protein complexes in animals have indicated that there may be a link between the RNA silencing and snoRNA pathways (Hock et al. 2007; Hutvagner and Simard 2008). Mass spectrometry of human HEK-293 cell Argonaute-associated proteins revealed a direct interaction between AGO1 and AGO2, and Nop56 and Fibrillarin, respectively, both of which are components of the snoRNA RNP complex (Kiss 2002; Hock et al. 2007; Matera et al. 2007). In plants, at least four key components of the RNA-induced silencing complex (RNA DEPENDENT RNA POLYMERASE 2, DICER-LIKE 3, ARGONAUTE4, and the largest subunit of Pol IVb, NRPD1b) localize to the nucleolus (Li et al. 2006, 2008; Pontes et al. 2006), the primary site of snoRNA biogenesis and function, and concentrate in

Reprint requests to: John S. Mattick, Australian Research Council Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia; e-mail: j.mattick@imb.uq.edu.au; fax: 61 7 3346 2111.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1528909.

Cajal bodies to assemble Ago-siRNA complexes (Pontes et al. 2006; Pontes and Pikaard 2008). Moreover, recent reports have indicated that a human snoRNA and a protozoan snoRNA are associated with Argonaute, processed into small RNAs, and can function as miRNAs (Ender et al. 2008; Saraiya and Wang 2008).

RESULTS AND DISCUSSION

We performed a systematic analysis of small RNA libraries we generated from a human monocytic leukemia cell line (THP-1) (Taft et al. 2009) and embryonic chicken (Glazov et al. 2008) and found extensive evidence for small RNAs derived from both classes of snoRNA. We then extended our analyses to other small-RNA deep sequencing libraries, including those from mouse embryonic stem (ES) cells (Babiarz et al. 2008), a variety of *Drosophila* tissues and life-cycle stages (Ruby et al. 2007; Czech et al. 2008), *Arabidopsis* ARGONAUTE2 and ARGONAUTE7 coimmunoprecipitations (IPs) (Montgomery et al. 2008), and Argonaute 1 IPs from wild-type (WT) and *cid14*Δ mutant fission yeast *Schizosaccharomyces pombe* (Supplemental Tables 1, 2; Buhler et al. 2008). More than 60% of human and mouse, 80% of chicken and *Drosophila*, and nearly all *Arabidopsis* and fission yeast snoRNAs show evidence of highly conserved sdRNAs (Supplemental Fig. 1), which occupy ~0.1 to >5% of total deep sequencing reads (Supplemental Table 1). In human THP-1 cells, we found that many highly abundant sdRNAs are derived from weakly expressed snoRNAs, suggesting that sdRNAs are the result of regulated processing rather than RNA turnover (Supplemental Fig. 2).

Indeed, metazoan sdRNAs exhibit other characteristics suggestive of regulated biogenesis, including distinct size distributions. Those derived from C/D box snoRNAs (C/D sdRNAs) are predominantly ~17–19 nt and ~30 nt (Fig. 1), the latter of which is typical of piRNAs (Peters and Meister 2007). These piRNA-sized C/D sdRNAs are detectable by Northern (Supplemental Fig. 3), but are only directly observable in the human data set, which contains deep sequencing reads extending up to 90 nt (see Materials and Methods). H/ACA box snoRNA-derived RNAs (H/ACA sdRNAs) show a peak at ~22 nt, similar to miRNAs, although in mouse this may be affected by context-dependent regulation of sdRNA biogensis (see below) (Fig. 1), and in flies, there are significant numbers of 23–24-nt small RNAs that may be related to small RNAs associated

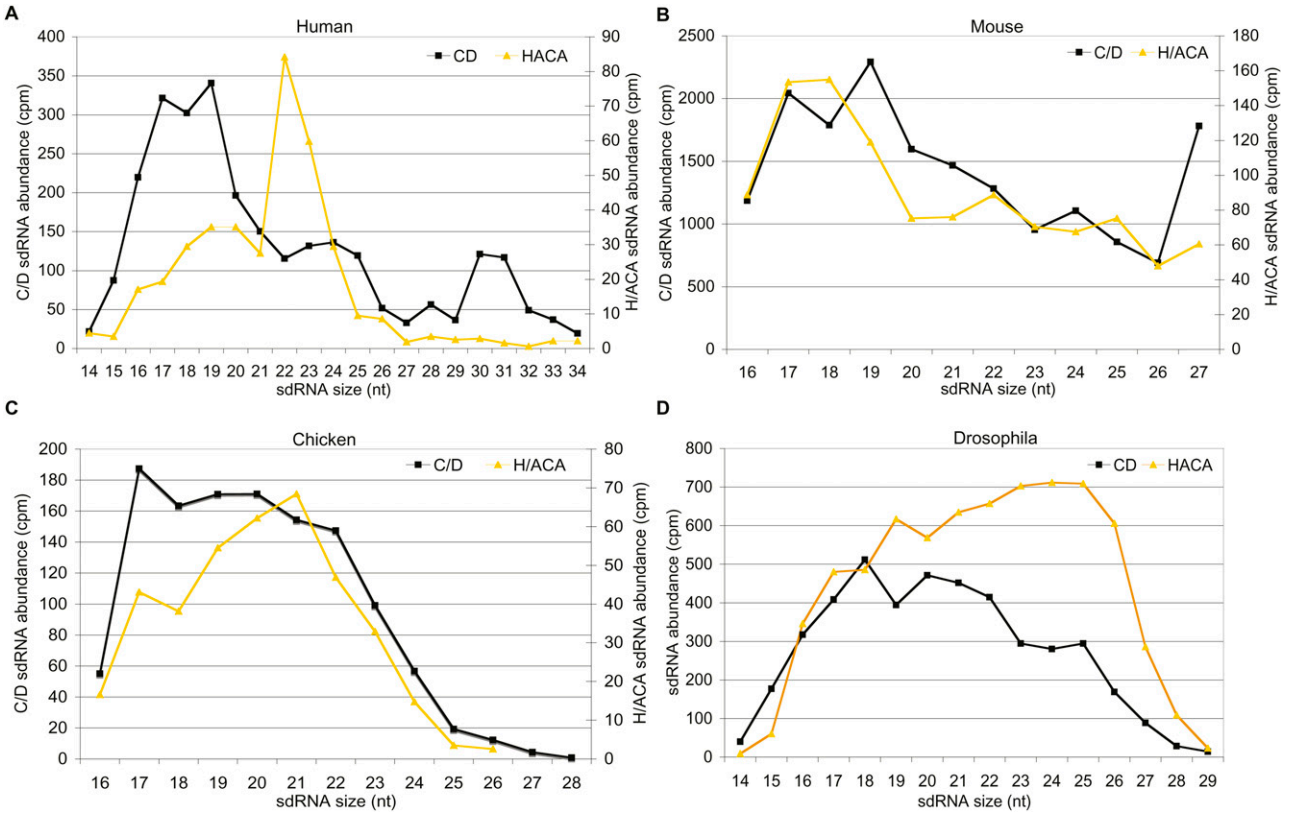

FIGURE 1. sdRNAs size distributions in metazoa. (Black) C/D and (yellow) H/ACA sdRNA sizes are plotted for (*A*) human, (*B*) mouse, (*C*) chicken, and (*D*) *Drosophila*. In all species, sdRNAs segregate by size dependent on the snoRNA type they originate from. In animals, C/D sdRNAs are more highly expressed and smaller than H/ACA sdRNAs. Vertical axes are in units of counts per million (cpm)—a ratio of the abundance of sdRNAs per million mapped tags in the total library. In *A–C*, C/D sdRNA abundance is denoted on the *left* of each panel, H/ACA on the *right*.

with AGO4 and Cajal bodies, one of the sites of H/ACA snoRNA function (Fig. 1; Li et al. 2008).

Further evidence of the specificity of sdRNAs is provided by the observation that C/D sdRNAs are derived predominantly from the 5′ end of full-length snoRNAs, whereas H/ACA sdRNAs are derived predominantly from the 3′ end (Fig. 2), both of which form double-stranded RNA structures (Matera et al. 2007). These data indicate a strong tendency for sdRNAs to be derived from the 3′ arm of one of the two miRNA-like hairpins that compose H/ACA snoRNAs (Fig. 2E). We also observed processing specificity for ~30-nt human C/D sdRNAs that originate from the middle of the host snoRNA (Supplemental Fig. 4). Sno-derived RNA size and position distributions are conserved in human, mouse, chicken, and *Drosophila* (Supplemental Fig. 5). Indeed, orthologous sdRNAs are identifiable in all three vertebrate species at 31 snoRNA loci (26 C/D and 5 H/ACA) and in two or more vertebrate species at 141 loci (103 C/D and 38 H/ACA), and exhibit conservation of size and position within the host snoRNA, albeit with some terminal sequence heterogeneity, which is typical of other types of small RNAs, including miRNAs (Supplemental Fig. 5; Wu et al. 2007; Morin et al. 2008).

To determine if RNAi pathway proteins affect the biogenesis of sdRNAs, we examined small RNA libraries from *dicer1*Δ/Δ and *dgcr8*Δ/Δ mouse ES cells (Babiarz et al. 2008), and *loquacious* (*loqs*) and *Dicer-2* (*Dcr-2*) mutant flies (Czech et al. 2008). Dicer1 and DGCR8 are principal

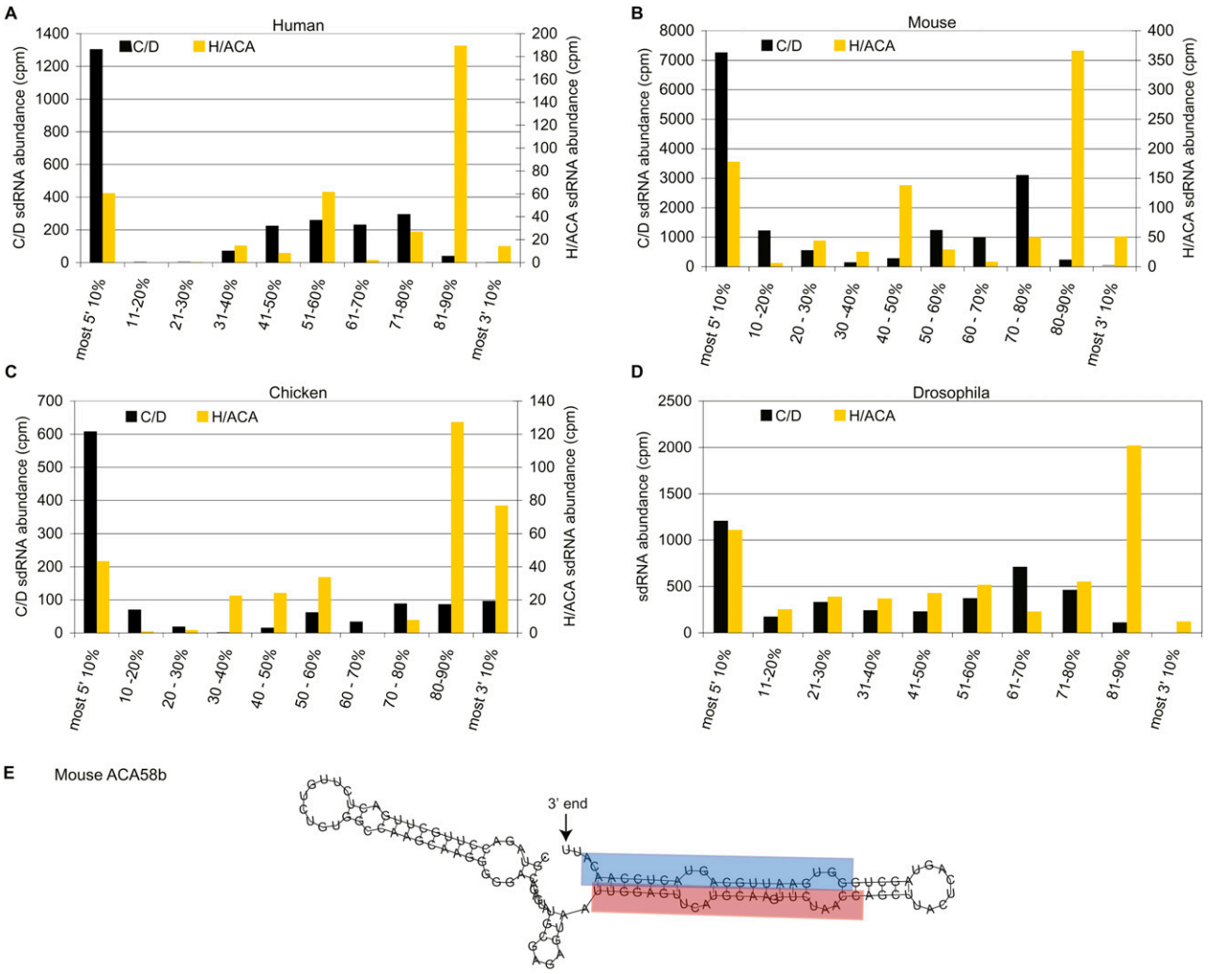

FIGURE 2. sdRNA position-of-origin within parent snoRNAs. The position-of-origin of (black) C/D and (yellow) H/ACA is shown for (*A*) human, (*B*) mouse, (*C*) chicken, and (*D*) *Drosophila*. To normalize for varying host snoRNA lengths, we divided snoRNAs into 10% blocks (deciles) from the 5′ to the 3′ end, and plotted the positions of the 5′ end of each sdRNA. C/D sdRNAs are primarily derived from the 5′ end of snoRNAs. H/ACA sdRNAs are predominantly derived from the 3′ end, a position that correlates to one of the two H/ACA snoRNA hairpins. C/D sdRNA abundance is denoted on the *left* of each panel, H/ACA on the *right*. (*E*) The structure of mouse ACA58b. SdRNAs are predominantly localized to the 3′ end of the RNA secondary structure. Sequences shaded in red and blue indicate miRNA-like mature and star sequences, respectively.

components of the miRNA biogenesis pathway in mammals. Pre-miRNA hairpins are cleaved from primary miRNA transcripts by Drosha-DGCR8 complexes and then processed into ~22-nt miRNAs by Dicer1. Consistent with the original report describing these data (Babiarz et al. 2008), we normalized against the abundance of tRNA, snRNA, and srpRNAs in each library (see Materials and Methods; Supplemental Table 3), which implicitly corrects for varying library depth, the loss of specific small RNA (e.g., miRNA or siRNA) species, and varying amounts of RNA degradation. The proportions of miRNAs in these libraries are ~100-fold and 20-fold less abundant in *dicer1*Δ/Δ and *dgcr8*Δ/Δ ES cells, respectively. In contrast, C/D sdRNAs are only mildly down-regulated and show size distributions identical to wild type (Fig. 3A).

H/ACA sdRNAs, however, show pronounced responses to the loss of *Dicer1* and *Dgcr8*. SnoRNA-derived small RNAs from two H/ACA loci are more than fourfold down-regulated in the absence of *Dicer1*, and a further nine are at least twofold down-regulated. Intriguingly, three, including those that are the most affected by the loss of Dicer, are twofold up-regulated in the absence of *Dgcr8* (Fig. 3B), suggesting that in some cases, DGCR8 may serve as an inhibitor of small RNA biogenesis. Indeed, *dgcr8*Δ/Δ H/ACA sdRNAs exhibit a dominant ~22-nt peak, similar to known miRNAs, with >2.5-fold increases in the number of 22 and 23 mers compared with wild type, which is primarily driven by three snoRNAs (Fig. 3B, ACA36b, ACA58b, ACA41). These data indicate that sdRNAs derived from at least two H/ACA snoRNAs are *Dicer1* sensitive and that DGCR8 may negatively regulate a subset of H/ACA sdRNAs. Indeed, H/ACA snoRNAs are principally composed of two short miRNA-like hairpins, which would appear to be ideal Droha-DGCR8 or Dicer substrates. Consistent with this hypothesis, the snoRNA ACA36b was recently annotated as a Dicer-sensitive endogenous shRNA (Babiarz et al. 2008) capable of generating miRNA-like small RNAs (Ender et al. 2008).

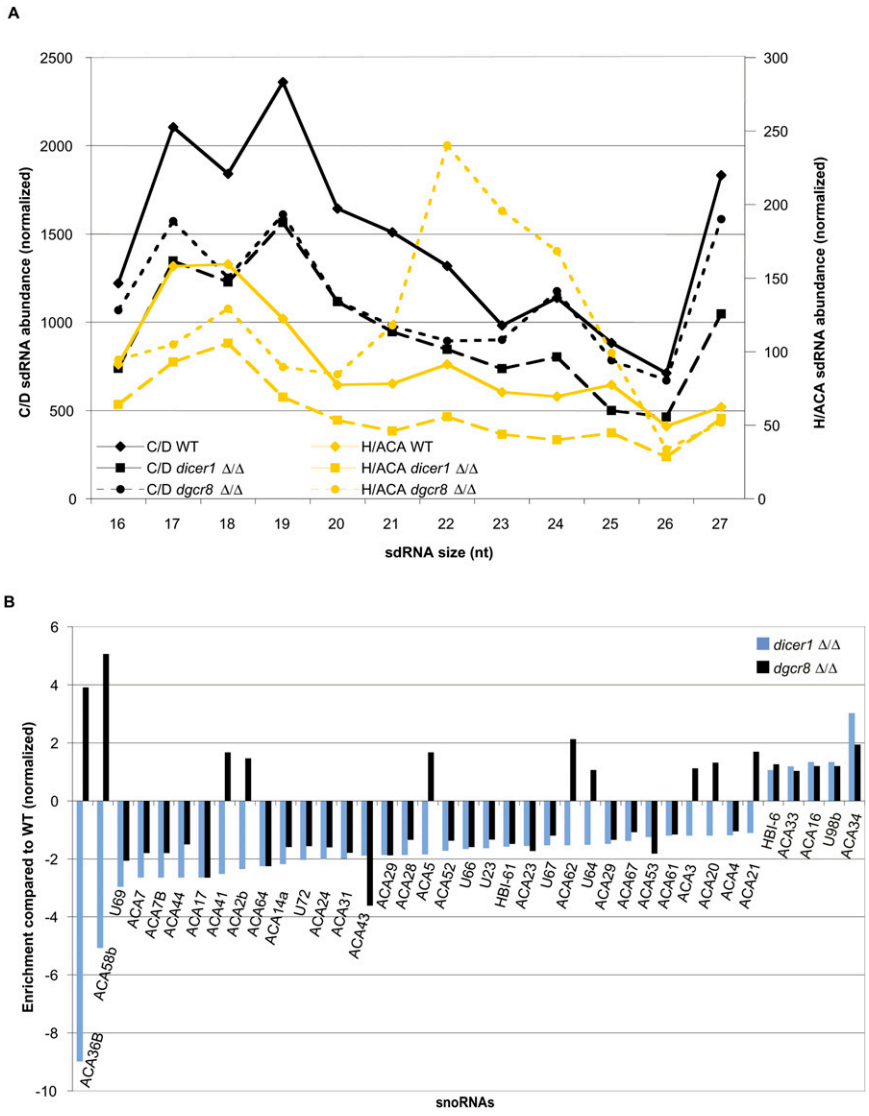

FIGURE 3. Mouse sdRNAs are affected by the loss of *Dicer1* and *Dgcr8*. Wild-type and mutant ES cell libraries were normalized to t/sn/srpRNA abundance (Supplemental Table 4; see the text). (*A*, black) C/D and (yellow) H/ACA sdRNA sizes in (solid lines) wild-type (WT), (large dashed lines) *dicer1*Δ/Δ, and (small dashed lines) *dgcr8*Δ/Δ embryonic stem cells. H/ACA sdRNAs also exhibit an increase of 21–24-nt species in *dgcr8*Δ/Δ ES cells. (*B*) The relative enrichment of sdRNA abundance compared to WT of 39 H/ACA snoRNAs in (blue) *dicer1*Δ/Δ and (black) *dgcr8*Δ/Δ embryonic stem cells. SnoRNAs were only included in the analysis if their sdRNAs had an abundance greater than 5 in each library and greater than 10 in at least one library.

Endogenous 21-nt siRNAs have been described in *Drosophila* and mammals (Babiarz et al. 2008; Czech et al. 2008; Ghildiyal et al. 2008; Watanabe et al. 2008). In flies, siRNAs are *Dcr-2* dependent, although a subset also relies on the double-stranded RNA-binding protein and Dicer partner, *loquacious*. We normalized small RNA libraries from mutant flies to t/sn/srpRNA fragments (Supplemental Table 4; see Materials and Methods) and found that C/D and H/ACA sdRNAs are more than fivefold, up-regulated in *loqs*$^{-/-}$ ovaries (Supplemental Figs. 6, 7) and show characteristics consistent with other *Drosophila* libraries (Supplemental Fig. 6), suggesting that the absence of *loquacious* leads to specific sdRNA enrichment. Indeed, sdRNAs from some snoRNA loci are more than 35-fold up-regulated (Supplemental Fig. 7). Additionally, sdRNAs from nine C/D and three H/ACA

snoRNAs are more than twofold down-regulated in the absence of *Dicer-2* (Supplemental Fig. 7). These data provide further evidence that sdRNA biogenesis is regulated by components of the RNAi pathway including proteins conventionally associated with siRNAs.

MiRNAs and siRNAs associate with the Argonaute family of proteins (Peters and Meister 2007; Hutvagner and Simard 2008). To test if sdRNAs are Argonaute associated, we queried AGO2 and AGO7 libraries from *Arabidopsis* (Montgomery et al. 2008) and Ago1 libraries from *S. pombe* (Buhler et al. 2008). We found that sdRNAs in *Arabidopsis* are strongly associated with AGO7 but not AGO2 (Fig. 4A; Supplemental Fig. 8). Eight H/ACA box and 25 C/D box snoRNAs are more than twofold enriched in the AGO7 library, corresponding to ~4% of all AGO7 associated small RNAs (Fig. 4A). Like metazoan sdRNAs, AGO7 H/ACA and C/D sdRNAs differ in size. C/D sdRNAs are smaller and dominantly 21 nt, while H/ACA sdRNAs are larger, ~27–29 nt (Supplemental Fig. 9). *Arabidopsis* Argonaute proteins preferentially load small RNAs with specific 5′ nucleotides and AGO7 is selective for 5′ uracil (U) and 5′ adenine (A). Intriguingly, H/ACA sdRNAs are dominantly 5′A, while C/D box are dominantly 5′U (Fig. 4A,B).

We observed similar trends in fission yeast. *S. pombe* Ago1-associated snoRNA fragments were recently reported in WT and *cid14*Δ [a non-canonical poly(A) polymerase and member of the exosome-mediated RNA degradation pathway] cells (Buhler et al. 2008). Consistent with previous findings, we find that sdRNAs comprise ~1.6% of the WT library and are up-regulated almost twofold to ~3.3% of the *cid14*Δ library (Supplemental Fig. 10; Supplemental Table 1). H/ACA sdRNAs dominate both Ago1 libraries (greater than fourfold over C/D sdRNAs), and show an up-regulation of 20-nt small RNAs in *cid14*Δ cells (Supplemental Fig. 11) that are derived from the central portion of the host snoRNA and fall on the 5′ arm of the 3′ hairpin (Supplemental Fig. 11). C/D sdRNA sizes are dominantly 22 and 23 nt in both data sets (similar to the library as a whole) (Supplemental Fig. 11), and unlike metazoan C/D sdRNAs, are principally derived from the 3′ end of the parent snoRNA (Supplemental Fig. 12). Like Argonautes in higher eukaryotes, *S. pombe* Ago1 preferentially loads small RNAs with a 5′U (Montgomery et al. 2008). Congruent with the 5′ nucleotide bias observed in *Arabidopsis*, we found that the vast majority of C/D and H/ACA sdRNAs of all sizes are 5′U enriched (Supplemental Fig. 13). These data strongly suggest that *Arabidopsis* and fission yeast sdRNAs are preferentially loaded into Argonaute complexes, which has also recently been reported for a small set of human sdRNAs (Ender et al. 2008), and are likely to have roles in the regulation of gene expression and transcriptional silencing. In particular, the fact that both miRNAs and sdRNAs are 5′U biased strengthens the link between them and suggests that some snoRNAs, including "orphan" snoRNAs whose targets are unknown, may function solely as intermediates in the sdRNA pathway.

The findings presented here indicate that sno-derived small RNAs are a general feature of both animal and plant biology, and that snoRNAs may have an ancient link with RNA silencing (Saraiya and Wang 2008), given that (1) sdRNAs are found in vertebrates, invertebrates, plants, and unicellular eukaryotes; (2) show a characteristic size distribution and origin that is distinctive with respect

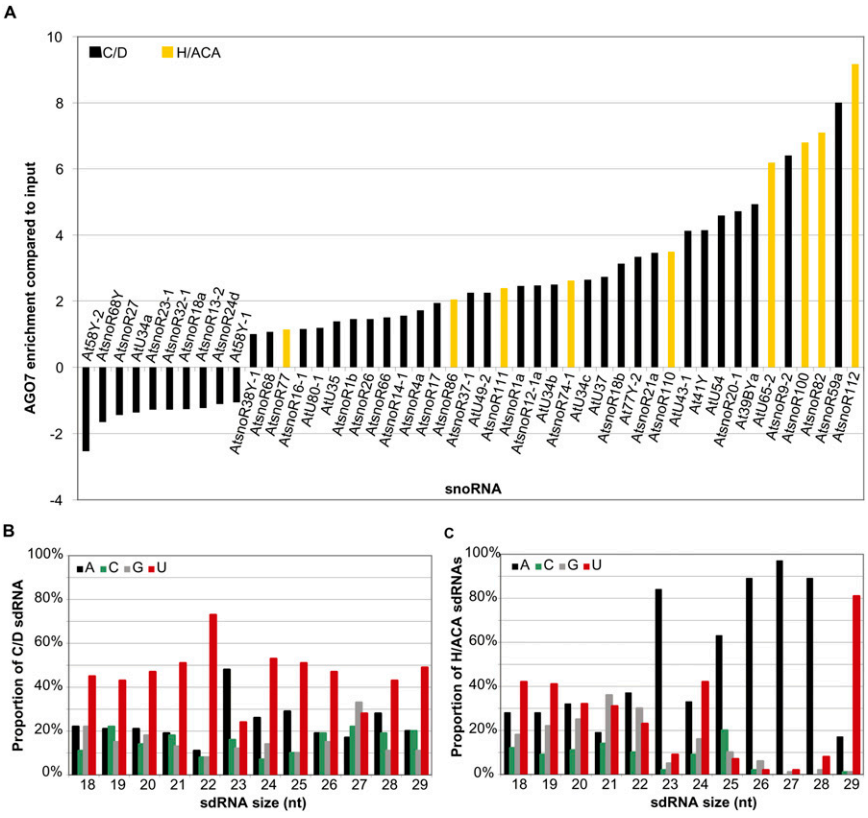

FIGURE 4. sdRNAs are associated with *Arabidopsis* AGO7. (*A*) The relative enrichment of (black) C/D and (yellow) H/ACA sdRNA abundance from individual snoRNAs in *Arabidopsis* AGO7 immunoprecipitations. Enrichments are calculated in comparison to IP input deep sequencing libraries. SnoRNAs were only included in the analysis if their sdRNAs had an abundance greater than 5 in each library and greater than 10 in at least one library. H/ACA and C/D sdRNAs are strongly enriched in the AGO7 library. (*B*,*C*) The proportions of C/D and H/ACA sdRNAs 5′ end nucleotides associated with *Arabidopsis* AGO7. C/D sdRNAs are 5′U biased and H/ACA sdRNAs are 5′A biased.

to the two major types of snoRNAs; (3) show specific responses to the loss of components of the RNAi pathway; and (4) are associated with specific members of the Argonaute family of proteins. It also reinforces the likelihood of an interplay between the RNA silencing and snoRNA-mediated RNA processing and RNA-directed regulatory systems. Indeed, given that snoRNAs are active in the nucleolus, the strong association of C/D and H/ACA sdRNAs with Argonautes is consistent with the emerging picture of the nucleolus as a site of small RNA biogenesis, stabilization, and function in animals, plants, and fungi (Pontes and Pikaard 2008). Additionally, sdRNA up-regulation in the absence of DGCR8, *loquacious*, and Cid14 suggests that sdRNA biogenesis is tightly controlled by known dsRNA binding proteins and members of the exosome-mediated small RNA processing pathway. These findings also suggest that approaches combining comparative genomics and deep sequencing are capable of revealing new classes of small RNAs and that snoRNAs, like other small RNAs, may be more multifaceted than previously supposed.

MATERIALS AND METHODS

Additional small RNA data sets

Small RNAs were obtained from 26 publicly available small RNA deep sequencing libraries (identifiers are listed in parentheses): human THP-1 small RNAs (DNA Database of Japan, AIAAA0000001–AIAAT0000001) (Taft et al. 2009); mouse WT, *dicer1*Δ/Δ, and *dgcr8*Δ/Δ ES stem cells (NCBI GEO, GSE12521) (Babiarz et al. 2008); chicken libraries from embryonic days 5, 7, and 9 (NCBI GEO, GSE10686) (Glazov et al. 2008); a variety of *Drosophila* tissues and life-cycle stages, and *Dcr-2*$^{-/-}$ and *loqs*$^{-/-}$ mutants (NCBI GEO, GSE7448 and GSE11086) (Ruby et al. 2007; Czech et al. 2008); *Arabidopsis* ARGONAUTE4 and ARGONAUTE7 IPs (NCBI GEO, GSE12037) (Montgomery et al. 2008); and Argonaute-1 IPs from WT and *cid14*Δ mutant *S. pombe* (NCBI GEO, GSE311595) (Buhler et al. 2008).

Small nucleolar RNA annotations

SnoRNA annotations were compiled from multiple sources. Human snoRNA annotations were obtained through the small RNA UCSC Genome Browser track (wgRNA) (Karolchik et al. 2008). Mouse snoRNAs were compiled from two sources: (1) we obtained a set of curated snoRNAs from the NCBI Entrez Gene database (~50 snoRNAs) (Wheeler et al. 2008); and (2) we identified orthologous evolutionary conserved mouse snoRNAs by mapping the human snoRNA annotations to the mouse genome using the UCSC syntenic alignment program, liftOver, requiring 95% sequence identity. Chicken snoRNAs annotations were obtained from the Ensemble chicken genome database (Flicek et al. 2008). The complete set of *Drosophila* snoRNAs was obtained from FlyBase (Drysdale 2008). *Arabidopsis* snoRNA annotations from Brown et al. (2001) were obtained from the Lowe Lab snoRNA database website (http://lowelab.ucsc.edu/snoRNAdb/). *S. pombe* snoRNA annotations were obtained from the Sanger Institute's GeneDB (Hertz-Fowler et al. 2004). Human, mouse, chicken, and *Drosophila* snoRNAs C/D and H/ACA box annotations were derived from the data files described above. *Arabidopsis* and *S. pombe* snoRNAs were manually curated for C/D and H/ACA annotations by querying the Entrez Gene and NCBI nucleotide databases with the sequence of interest using BLAST. A subset of *S. pombe* snoRNAs without snoRNA-type annotations were designated as C/D or H/ACA based on the presence of canonical sequence motifs and total sequence length, yielding a total of 17 and 27 high-confidence C/D and H/ACA box *S. pombe* snoRNAs, respectively.

sdRNA analysis

Bioinformatics analysis was done on a high-performance computing station that houses a local mirror of the UCSC Genome Browser (Karolchik et al. 2008) and a local installation of the Biopieces toolset (http://www.biopieces.org/, developed by Martin Hansen). All small RNA data sets were mapped onto reference genomes using Vmatch (http://www.vmatch.de/). We used the UCSC builds of the following genome assemblies: human (hg18 NCBI build 36.1), mouse (mm8 NCBI build 36), chicken (galGal3 v2.1 draft assembly Genome Sequencing Center at WUSTL), and *Drosophila* (dm3 BDGP Release 5). We created Vmatch indexes from *Arabidopsis* whole-chromosome sequence files from The *Arabidopsis* Information Resource website (Poole 2007). Likewise, we created Vmatch mapping indexes from *S. pombe* chromosome sequences (ftp://ftp.sanger.ac.uk/pub/yeast/pombe/Chromosome_contigs/). We required small RNAs to map uniquely to the genome of interest without any mismatches. Sno-derived RNAs were identified by intersecting snoRNA annotations with small RNA data sets using a modified version of the UCSC back end C++ tool, bedIntersect (Karolchik et al. 2008). Conservation profiles for sdRNA were computed using phastCons scores from the phastCons17way and phastCons15way MySQL tables for human and *Drosophila* snoRNA regions of interest, respectively (Siepel et al. 2005).

The human THP-1 small RNA data set was generated on the Roche FLX Genome Sequencer and contains sequence reads up to 90 nt long. Deep sequencing reads from all other species, however, were generated on the Illumina Genome Analyzer, which has an upper limit of ~30 nt and in many libraries only polled small RNAs <28 nt. A complete assessment of sdRNAs ≥29 nt was only possible in human THP-1 cells.

We took advantage of long THP-1 deep sequencing reads to assess the relationship between snoRNA and sdRNA expression. We identified snoRNAs with individual deep sequencing reads covering at least 90% of their length, pooled them for each snoRNA, and then compared total snoRNA abundance with the abundance of smaller (<40 nt) sno-derived RNAs.

SdRNA size distributions were assessed using custom AWK, Shell, and Perl scripts. The relative position of sdRNAs within the parent snoRNA was accomplished by parsing individual snoRNAs into 10% blocks (to normalize for differences in total snoRNA length) and then tabulating the abundance of sdRNA 5′ ends in each decile. SnoRNA enrichments and sdRNA size distributions in mouse wild-type, *dgcr8*Δ/Δ, and *dicer1*Δ/Δ ES cells and *Drosophila* wild-type, *Dcr-2*$^{-/-}$, and *loqs*$^{-/-}$ ovaries were performed on libraries normalized against the total abundance of tRNA, snRNA, and srpRNA-derived small RNA fragments (Supplemental Tables 3 and 4), as described previously (Babiarz et al. 2008). We found that

t/sn/srpRNA derived reads are ~2.6× more highly represented in *dgcr8*Δ/Δ and *dicer1*Δ/Δ ES cells compared to wild-type, and ~1.4× and ~2× more abundant in *Dcr-2*$^{-/-}$ and *loqs*$^{-/-}$ ovaries compared to wild-type, respectively. For all other analyses, we normalized the counts in each library to obtain counts per million, or "tags per million mapped sequences"—the total abundance of the sdRNAs associated with each snoRNA was divided by the total abundance of the library tags that mapped to the genome and multiplied by 10^6. SnoRNAs were only considered in enrichment analyses if their sdRNAs had counts >5 in all investigated libraries and had at least one library with sdRNA abundance >10. To calculate fold enrichment, the normalized counts of each mutant libraries were divided by the normalized counts of the appropriate WT library. Previous analyses have normalized libraries using the number of small RNA "degradation products," including sequences derived from snoRNAs. We reasoned that a direct comparison of the relative sequence depths of the libraries would be a more accurate measure of abundance, particularly in light of the fact that it appears that small RNAs can be derived from longer small RNA species including snoRNAs and tRNAs.

Orthologous sdRNAs were identified by mapping human and chicken sdRNAs to the mouse genome using the UCSC syntenic alignment tool, liftOver, and intersecting them with previously mapped mouse sdRNAs. For sdRNAs to be considered orthologous, we required that the 5′ ends of the sdRNAs from each species map to the same position in the mouse snoRNA, a >90% sequence identity, and similarity in size (±20%).

SdRNA 5′ nucleotide enrichments were assessed by extracting the sequences of interest, parsing the sequences by size, and tabulating the percentage of each base using the Biopieces program weight_matrix. RNA structures (see Fig. 3) were calculated using the Vienna RNA fold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (Hofacker 2003).

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank Martin Hansen for access to, and assistance with, the Biopieces toolset. R.J.T. is supported by a National Science Foundation Graduate Research Fellowship. J.S.M. is supported by an Australian Research Council Federation Fellowship (FF0561986), the University of Queensland, and the Queensland State Government. This work was also supported by grants for the Genome Network Project from the Ministry of Education, Culture, Sports, Science and Technology, Japan (Y.H.), and a Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government, the RIKEN Frontier Research System, Functional RNA Research Program (Y.H.).

Received December 22, 2008; accepted April 9, 2009.

REFERENCES

Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes & Dev* **22:** 2773–2785.

Brown JW, Clark GP, Leader DJ, Simpson CG, Lowe T. 2001. Multiple snoRNA gene clusters from *Arabidopsis*. *RNA* **7:** 1817–1832.

Buhler M, Spies N, Bartel DP, Moazed D. 2008. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. *Nat Struct Mol Biol* **15:** 1015–1023.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453:** 798–802.

Drysdale R. 2008. FlyBase: A database for the *Drosophila* research community. *Methods Mol Biol* **420:** 45–59.

Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. 2008. A human snoRNA with microRNA-like functions. *Mol Cell* **32:** 519–528.

Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, et al. 2008. Ensembl 2008. *Nucleic Acids Res* **36:** D707–D714.

Gaspin C, Cavaille J, Erauso G, Bachellerie JP. 2000. Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: Lessons from the *Pyrococcus* genomes. *J Mol Biol* **297:** 895–906.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320:** 1077–1081.

Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. 2008. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res* **18:** 957–964.

Hertz-Fowler C, Peacock CS, Wood V, Aslett M, Kerhornou A, Mooney P, Tivey A, Berriman M, Hall N, Rutherford K, et al. 2004. GeneDB: A resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res* **32:** D339–D343.

Hock J, Weinmann L, Ender C, Rudel S, Kremmer E, Raabe M, Urlaub H, Meister G. 2007. Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. *EMBO Rep* **8:** 1052–1060.

Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31:** 3429–3431.

Hutvagner G, Simard MJ. 2008. Argonaute proteins: Key players in RNA silencing. *Nat Rev Mol Cell Biol* **9:** 22–32.

Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, et al. 2008. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* **36:** D773–D779.

Kiss T. 2002. Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109:** 145–148.

Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, Chan SW, Lagrange T, Pikaard CS, Jacobsen SE. 2006. An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126:** 93–106.

Li CF, Henderson IR, Song L, Fedoroff N, Lagrange T, Jacobsen SE. 2008. Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in *Arabidopsis thaliana*. *PLoS Genet* **4:** e27. doi: 10.1371/journal.pgen.0040027.

Matera AG, Terns RM, Terns MP. 2007. Noncoding RNAs: Lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8:** 209–220.

Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC. 2008. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 *trans*-acting siRNA formation. *Cell* **133:** 128–141.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, et al. 2008. Application of massively parallel sequencing to microRNA

profiling and discovery in human embryonic stem cells. *Genome Res* **18:** 610–621.

Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP. 2000. Homologs of small nucleolar RNAs in Archaea. *Science* **288:** 517–522.

Peters L, Meister G. 2007. Argonaute proteins: Mediators of RNA silencing. *Mol Cell* **26:** 611–623.

Pontes O, Pikaard CS. 2008. siRNA and miRNA processing: new functions for Cajal bodies. *Curr Opin Genet Dev* **18:** 197–203.

Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS. 2006. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126:** 79–92.

Poole RL. 2007. The TAIR database. *Methods Mol Biol* **406:** 179–212.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* **17:** 1850–1864.

Saraiya AA, Wang CC. 2008. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog* **4:** e1000224. doi: 10.1371/journal.ppat.1000224.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15:** 1034–1050.

Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest ARR, Grimmond SM, Schroder K, et al. 2009. Tiny RNAs associated with transcription start sites in animals. *Nat Genet* **41:** 572–578.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453:** 539–543.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, et al. 2008. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **36:** D13–D21.

Wu H, Neilson JR, Kumar P, Manocha M, Shankar P, Sharp PA, Manjunath N. 2007. miRNA profiling of naive, effector and memory CD8 T cells. *PLoS One* **2:** e1020. doi: 10.1371/journal.pone.0001020.