

Washington University School of Medicine Digital Commons@Becker

Open Access Publications

2008

Conifers have a unique small RNA silencing signature

Elena V. Dolgosheina
Simon Fraser University

Ryan D. Morin
BC Cancer Agency

Gozde Aksay
University of Washington - Seattle Campus

S. Cenk Sahinalp
Simon Fraser University

Vincent J. Magrini
Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation

Dolgosheina, Elena V.; Morin, Ryan D.; Aksay, Gozde; Sahinalp, S. Cenk; Magrini, Vincent J.; Mardis, Elaine; Mattsson, Jim; and Unrau, Peter J., "Conifers have a unique small RNA silencing signature." *RNA*.14,8. 1508-1515. (2008).
http://digitalcommons.wustl.edu/open_access_pubs/4169

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.

Authors

Elena V. Dolgosheina, Ryan D. Morin, Gozde Aksay, S. Cenk Sahinalp, Vincent J. Magrini, Elaine Mardis, Jim Mattsson, and Peter J. Unrau

Conifers have a unique small RNA silencing signature

ELENA V. DOLGOSHEINA,¹ RYAN D. MORIN,² GOZDE AKSAY,³ S. CENK SAHINALP,⁴ VINCENT MAGRINI,⁵ ELAINE R. MARDIS,⁵ JIM MATTSSON,⁶ and PETER J. UNRAU¹

¹Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

²Genome Sciences Centre, BC Cancer Agency, Vancouver BC, V5Z 4S6, Canada

³Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA

⁴School of Computing Sciences, Simon Fraser University, Burnaby BC, V5A 1S6, Canada

⁵Washington University School of Medicine, Genome Center at Washington University, St. Louis, Missouri USA, 63108

⁶Biological Sciences, Simon Fraser University, Burnaby BC, V5A 1S6, Canada

ABSTRACT

Plants produce small RNAs to negatively regulate genes, viral nucleic acids, and repetitive elements at either the transcriptional or post-transcriptional level in a process that is referred to as RNA silencing. While RNA silencing has been extensively studied across the different phyla of the animal kingdom (e.g., mouse, fly, worm), similar studies in the plant kingdom have focused primarily on angiosperms, thus limiting evolutionary studies of RNA silencing in plants. Here we report on an unexpected phylogenetic difference in the size distribution of small RNAs among the vascular plants. By extracting total RNA from freshly growing shoot tissue, we conducted a survey of small RNAs in 24 vascular plant species. We find that conifers, which radiated from the other seed-bearing plants ~260 million years ago, fail to produce significant amounts of 24-nucleotide (nt) RNAs that are known to guide DNA methylation and heterochromatin formation in angiosperms. Instead, they synthesize a diverse population of small RNAs that are exactly 21-nt long. This finding was confirmed by high-throughput sequencing of the small RNA sequences from a conifer, *Pinus contorta*. A conifer EST search revealed the presence of a novel Dicer-like (DCL) family, which may be responsible for the observed change in small RNA expression. No evidence for DCL3, an enzyme that matures 24-nt RNAs in angiosperms, was found. We hypothesize that the diverse class of 21-nt RNAs found in conifers may help to maintain organization of their unusually large genomes.

Keywords: RNA silencing; Dicer (DCL); small RNA processing; plant evolution; gymnosperms

INTRODUCTION

One of the unifying features of small RNA biogenesis in different eukaryotic organisms is the presence of Dicers—endonuclease enzymes with RNase III domains, which process small RNAs prior to their incorporation into downstream multicomponent RNA-induced silencing complexes (RISCs) (Hammond 2005; MacRae et al. 2006). Unlike other eukaryotes, plants have significantly diversified their Dicer-mediated pathways. There are at least four functionally distinct Dicer-like (DCL) enzymes in flowering plants or angiosperms, and these produce three distinct size-classes of small RNAs (Xie et al. 2004; Brodersen and Voinnet 2006; Margis et al. 2006). DCL1 produces highly conserved microRNAs (miRNAs) from stem-loop regions of long primary transcripts, and these RNAs are typically 21-

nucleotides (nt) long (Jones-Rhoades et al. 2006). The combined action of DCL2 and DCL4 produce short interfering RNAs (siRNAs) that are 21–22-nt long and are derived from either viral or endogenous loci (Brodersen and Voinnet 2006; Deleris et al. 2006). Endogenous siRNAs are further subdivided into *trans*-acting siRNAs and natural antisense siRNAs, which control aspects of developmental timing and stress adaptation, respectively, in angiosperms (Borsani et al. 2005; Gascioli et al. 2005). Finally, DCL3 generates 24-nt RNAs (Hamilton et al. 2002; Xie et al. 2004). These RNAs mediate DNA and histone modifications of repetitive and transposable elements and are involved in the formation of heterochromatin (Chan et al. 2004; Matzke and Birchler 2005; Wassenegger 2005; Henderson et al. 2006; Henderson and Jacobsen 2007). This RNA size class is particularly fascinating in that it exhibits a much higher sequence diversity than any other small RNA population in flowering plants (Lu et al. 2005; Rajagopalan et al. 2006). In the angiosperm *Arabidopsis thaliana*, at least 35,000 distinct 24-nt sequences are known to be produced, constituting ~60% of its small RNA transcriptome (Lu et al. 2005).

Reprint requests to: Peter J. Unrau, Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby BC Canada, V5A 1S6; e-mail: punrau@sfu.ca; fax: (778) 782-5583.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1052008>.

To date, most of the research on RNA silencing in vascular plants has focused on model angiosperms, that is, the eudicot *A. thaliana* and the monocot *Oryza sativa* (rice). Yet, angiosperms only came to dominate the terrestrial floral diversity ~150 million years ago (mya) (Schneider et al. 2004), relatively recently considering the 480-million-year history of land plants (Kenrick and Crane 1997). Two other major divisions of vascular plants—gymnosperms and pteridophytes—are still extant and have much longer evolutionary histories. Gymnosperms, which are ancestral to angiosperms (Kenrick and Crane 1997), are the most ancient spermatophytes (seed-bearing plants) and, in contrast to angiosperms, produce seeds that are not protected by an ovary wall. Conifers, which include pines, spruces, and firs, radiated ~260 mya from the other seed plants (Schneider et al. 2004), and form the largest and most diverse group of the cone-bearing gymnosperms. The still more ancient pteridophytes are seedless plants, such as the clubmosses, ferns, and horsetails, which first appear in the fossil record ~430 mya (Kenrick and Crane 1997; Schneider et al. 2004).

To investigate the conservation of small RNA-generating machinery across vascular plants, we conducted a survey of small RNA expression profiles in the seed-bearing and seedless plants. We studied the length distribution and the presence of terminal modifications on small RNAs extracted from representatives of pteridophytes, gymnosperms, and flowering plants, including an ancient angiosperm, *Nymphaea* sp. or water lily (Schneider et al. 2004). Biochemical experiments and sequencing results that we describe here revealed that conifers produce an extremely diverse population of 21-nt RNAs and show no detectable levels of 24-nt RNAs, which are known to direct chromatin modifications in angiosperms (Hamilton et al. 2002; Xie et al. 2004; Matzke and Birchler 2005; Wassenegger 2005). This unique small RNA distribution correlates with the presence of a new DCL family that we found by performing extensive conifer EST searches, and the apparent absence of DCL3, which plays an important role in transcriptional gene silencing in angiosperms (Chan et al. 2004; Henderson et al. 2006).

RESULTS AND DISCUSSION

We analyzed small RNA size profiles from 24 vascular plant species (Fig. 1; Supplemental Table 1). Total RNA was extracted from meristematic (actively

dividing) shoot tissue, labeled by exchanging the 5'-phosphate (assumed to be present on all small RNAs) with [γ - 32 P]ATP and polynucleotide kinase (PNK), and resolved on a denaturing polyacrylamide gel. As expected, within the angiosperms, the monocots (Fig. 1, lanes 17–19) and the eudicots (Fig. 1, lanes 20–24) strongly expressed 24-nt populations of small RNAs and relatively modest levels of 21–22-nt small RNAs. In contrast, the majority of sampled gymnosperms (Fig. 1, lanes 6–15) expressed predominantly 21-nt small RNAs. In fact, all seven conifers (Fig. 1, lanes 6–12) lacked detectable 24-nt RNA bands. Intriguingly, vascular plants that precede spermatophytes in evolution, such as ferns, appear to be very similar in their small RNA size profiles to primitive angiosperms such as *Nymphaea caerulea* in that they produce both 24-nt and 21–22-nt size classes of small RNA (Fig. 1, cf. ferns in lanes 4,5 and *N. caerulea* in lane 16). This suggests that small RNAs undergo atypical processing in gymnosperms, specifically in the conifers. Given that the more anciently divergent pteridophytes produce 24-nt small RNAs, it appears that conifers may have lost their ability to produce this size class of RNA.

To confirm the size distribution of small RNAs inferred from our biochemical data, we performed an extensive sampling of small RNA sequences from the conifer *Pinus contorta* (lodgepole pine) and a lighter sampling from the

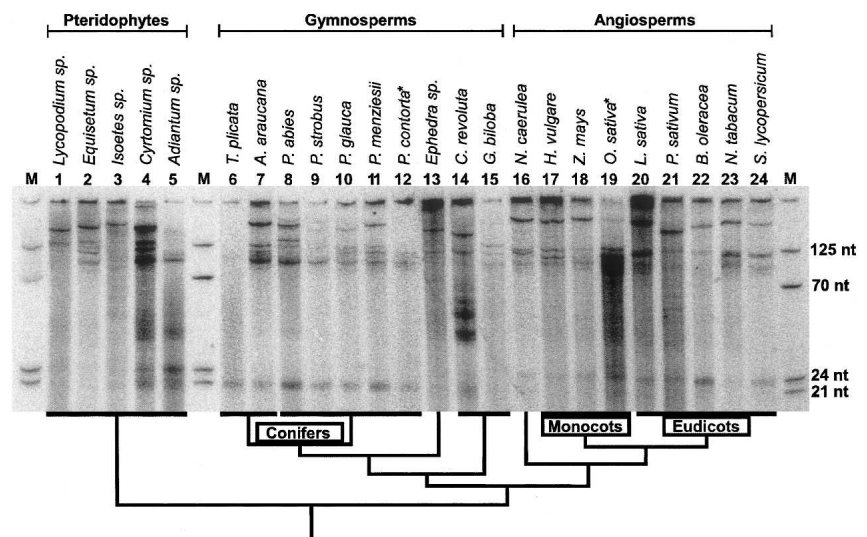


FIGURE 1. Conifers are notable in that they lack a detectable 24-nt population of small RNAs, otherwise found in a diverse range of vascular plants, while producing a dominant 21-nt class of small RNAs. Polyacrylamide gel analysis was performed to investigate total RNA size profiles from 24 vascular plant species. Pteridophytes: (lanes 1–3) clubmoss, horsetail, quillwort, and (lanes 4,5) ferns. (Lanes 6–12) Conifer gymnosperms: Western red cedar, monkey-puzzle tree, Norway spruce, Eastern white pine, white spruce, Douglas-fir, and lodgepole pine. (Lanes 13–15) Other gymnosperms: ephedra, cycad, and ginkgo. (Lane 16) Ancient angiosperms: water lily. (Lanes 17–19) Monocot angiosperms: barley, maize, and rice. (Lanes 20–24) Eudicot angiosperms: lettuce, pea, cauliflower, tobacco, and tomato. (M) Marker lane. Small RNAs from plant species marked with an asterisk were subjected to high throughput 454 sequencing. The phylogeny of major vascular plant divisions is shown below the gel image (Schneider et al. 2004).

previously studied monocot *O. sativa*. Small RNAs 15-nt to 30-nt long were extracted from growing pine needle tips and from young rice leaves and subjected to high-throughput 454 pyrosequencing (Morin et al. 2008). Altogether we obtained 130,998 small RNA sequences from *P. contorta* and 11,329 from *O. sativa*, of which 58,466 and 8615 were unique, respectively. To identify the potential genomic sources of *O. sativa* small RNAs, all sequences were aligned to the rice genome using Megablast, forcing a full-length alignment with no mismatches. Since no gymnosperm genome has ever been sequenced, *P. contorta* small RNAs were aligned to *Pinus taeda* ESTs and cDNAs currently available in GenBank. This data set is accessible at ftp://ftp03.bcgs.c.ca/public/pine_rice_smRNA/, and the bioinformatics analysis is discussed in greater detail in Morin et al. (2008).

Figure 2 shows the size distribution of small RNAs from *P. contorta* and *O. sativa*. Consistent with our polyacrylamide gel analysis, *O. sativa* had nearly twice as many 24-nt RNAs as 21-nt RNAs (Fig. 2A), exemplifying the presence of a DCL3-mediated pathway in angiosperms. The 21-nt population contained many known miRNAs, whereas the 24-nt population was dominated by sequences from genomic repeats and intergenic regions (Morin et al. 2008). In contrast to *O. sativa*, the majority of small RNAs from

P. contorta were 21-nt long (55%) (Fig. 2B). A marked bias toward 21-nt sequences starting with a U was observed, which may be biologically significant (see Supplemental Table 2). The fraction of 24-nt RNAs was very small (2.6%) and had a lower diversity than the same size class in *O. sativa*. The dominant 21-nt class exhibited very high sequence diversity—out of 72,045 *P. contorta* sequences that were 21-nt long, 29,920 unique sequences were found (42% of the total 21-nt population). Furthermore, this size class was much more diverse in *P. contorta* than in *O. sativa*, suggesting that a diversification of the 21-nt-RNA-producing pathways has occurred in conifers.

A small fraction (0.3%) of the *P. contorta* sequence library was identified as highly conserved miRNAs previously identified in angiosperms (Fig. 2C; Jones-Rhoades et al. 2006), and their frequency by size is shown in Figure 2D. The majority of miRNAs from *P. contorta* fall in the 21-nt size range, which corresponds well to the size distribution of miRNAs observed in *O. sativa*. Furthermore, we discovered many novel putative miRNAs in *P. contorta* using EST folding and machine learning approaches (Morin et al. 2008), which indicates that the rapid evolution of miRNAs that is well documented in angiosperms (Rajagopalan et al. 2006) also occurs in conifers. The presence of many previously identified miRNAs in conifers is

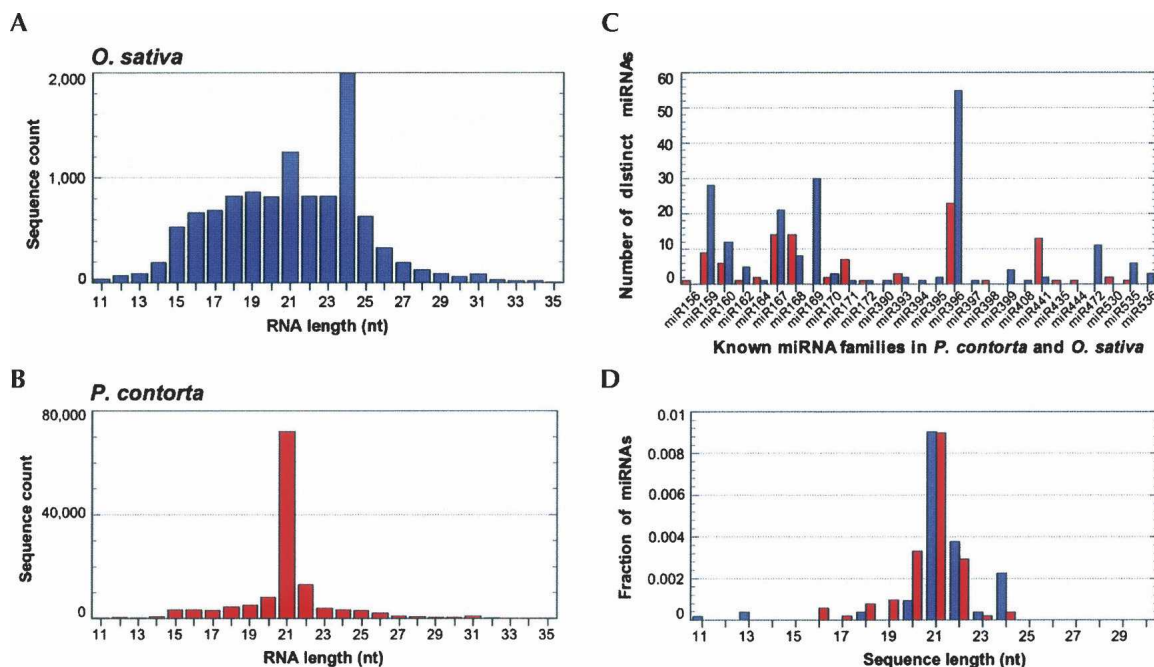


FIGURE 2. Length distribution of total small RNA and conserved miRNA sequences obtained from *O. sativa* and *P. contorta*. (A) Small RNAs from *O. sativa* exhibit the profile typical of an angiosperm sample having a dominant 24-nt RNA population together with a smaller fraction of 21–22-nt RNA. (B) In contrast, small RNAs from *P. contorta* were dominated by a 21-nt class of small RNAs. This class was very diverse in sequence composition. (C) Detection of known miRNAs. Conserved miRNAs from *P. contorta* (red bars) and *O. sativa* (blue bars) were identified by sequence similarity comparisons with known miRNA sequences deposited in miRBase release 10.0 (<http://microrna.sanger.ac.uk>). *O. sativa*: 99 distinct miRNA sequences were found corresponding to 18 known miRNA families (i.e., having a unique random sequence tag). *P. contorta*: 173 distinct miRNA sequences were found corresponding to 22 known miRNA families. (D) Fractions of known miRNAs are plotted by size; (red bars) *P. contorta*; (blue bars) *O. sativa*. Data adapted from Morin et al. (2008).

consistent with the high conservation of miRNA biogenesis and processing pathways across the land plants (Axtell et al. 2007). As the identified miRNAs comprised only a small fraction of the 21-nt RNA population in *P. contorta*, it is tempting to propose that the remaining 21-nt RNAs in conifers may substitute for the missing 24-nt class that mediates chromatin modification in angiosperms.

Since small RNAs are known to be terminally methylated by HEN1 in angiosperms (Ebhardt et al. 2005; Yu et al. 2005) and are terminally modified in the nonvascular moss *Physcomitrella patens* (Axtell et al. 2007), we tested selected angiosperm, gymnosperm, and fern samples for the presence of 3' modifications using periodate-mediated β -elimination of the terminal nucleoside. When both 2'- and 3'-hydroxyls are unmodified, these groups are oxidized by periodate to form an unstable dialdehyde, resulting in β -elimination of the terminal nucleoside upon heating and an \sim 2-nt downward shift in electrophoretic mobility (Yu et al. 2005). Small RNAs from a pteridophyte, conifer, and angiosperm were significantly resistant to periodate cleavage (Fig. 3), indicating that small RNAs from each sample were

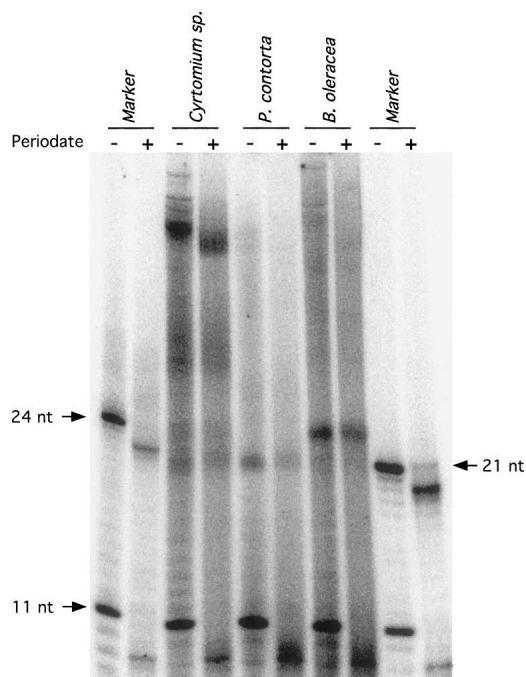


FIGURE 3. Sodium periodate treatment of small RNAs from *Cyrtomium sp.* (fern), *P. contorta* (lodgepole pine), and *Brassica oleracea* (cauliflower). Only if both the 2'- and 3'-hydroxyl groups are unmodified will β -elimination of the terminal nucleoside take place, resulting in an \sim 2-nt increase in electrophoretic mobility. Small RNAs in the 21–24-nt range for all three plant samples were highly resistant to the sodium periodate treatment, as exemplified by the comigration of small RNA bands in the – and + periodate lanes. A synthetic 11-nt RNA 5'-end-labeled with ^{32}P was added as an internal positive control (lower band) to all RNA extracts before chemical treatment to confirm the efficient periodate-mediated elimination of the terminal nucleoside from unmodified RNA. The size marker on the left is 24-nt long, the marker on the right is 21-nt long.

modified on either the 2'- or the 3'-hydroxyl. Given that small RNAs from *P. contorta* were also substrates for T4 RNA ligase—an enzyme that requires a free terminal 3'-hydroxyl group—during preparation for sequencing (see Materials and Methods), our results strongly suggest that a post-transcriptional process modifies small RNAs on the terminal 2'-hydroxyl group of this gymnosperm. Together with the periodate resistance observed for the pteridophyte sample and the nonvascular *P. patens* (Axtell et al. 2007), this evidence indicates that the terminal modification of small RNAs is likely to be a conserved aspect of RNA silencing in land plants that is not sensitive to size-dependent changes in RNA processing.

As DCL enzymes are involved in the production of small RNAs and are important determinants of small RNA size (MacRae et al. 2006; MacRae and Doudna 2007), we proceeded to identify candidate DCL genes in conifers. Figure 4A shows the linear organization of an angiosperm Dicer protein that contains DexD-helicase, helicase-C, Duf283, PAZ, RNase III, and double-stranded RNA-binding (dsRBD) domains (Margis et al. 2006). Helicase domains are found toward the N-terminal region of the protein. Duf283 is a domain of unknown function, which is strongly conserved among plant DCLs and is thought to contain a double-stranded RNA-binding fold (Dlatic 2006). There are always two RNase III domains, which together with the PAZ domain are crucial for specific recognition and cleavage of dsRNA precursors. The dsRBD is generally thought to mediate nonspecific interactions with dsRNA (Margis et al. 2006; MacRae and Doudna 2007). Given that no gymnosperm genome has ever been sequenced, we searched *Pinus taeda* and *Picea spp.* EST libraries (\sim 367,000 reads; see Materials and Methods) for putative conifer DCLs using four known Dicer nucleotide sequences from *A. thaliana* (Xie et al. 2005; Margis et al. 2006). ESTs obtained through similarity searches were translated into amino acid sequences using the ExpASY translate tool (see Materials and Methods), and the sequences with the longest open reading frames were used for further analysis. We identified four *P. taeda* and three *Picea sp.* ESTs with high-quality alignments ($>$ 35% similarity on the amino acid level) that were orthologous to the C-terminal region of *A. thaliana* DCLs and that spanned the C-terminal region of the second RNase III domain and the dsRBD found in the DCLs of flowering plants (Fig. 4A). A multiple sequence alignment was made with these predicted protein sequences and the corresponding region from each of the four *A. thaliana*, six *O. sativa* DCLs (Xie et al. 2005; Margis et al. 2006), and two predicted DCL protein sequences from the angiosperm *Medicago truncatula*. Figure 4B shows the most conserved region (90 amino acids long) of the alignment. Eleven residues are absolutely conserved in this region (Fig. 4B, marked with asterisks). Four of the seven conifer sequences (Fig. 4B, bottom of alignment) shared an additional distinctive set of conserved residues (i.e., residues 96, 120,

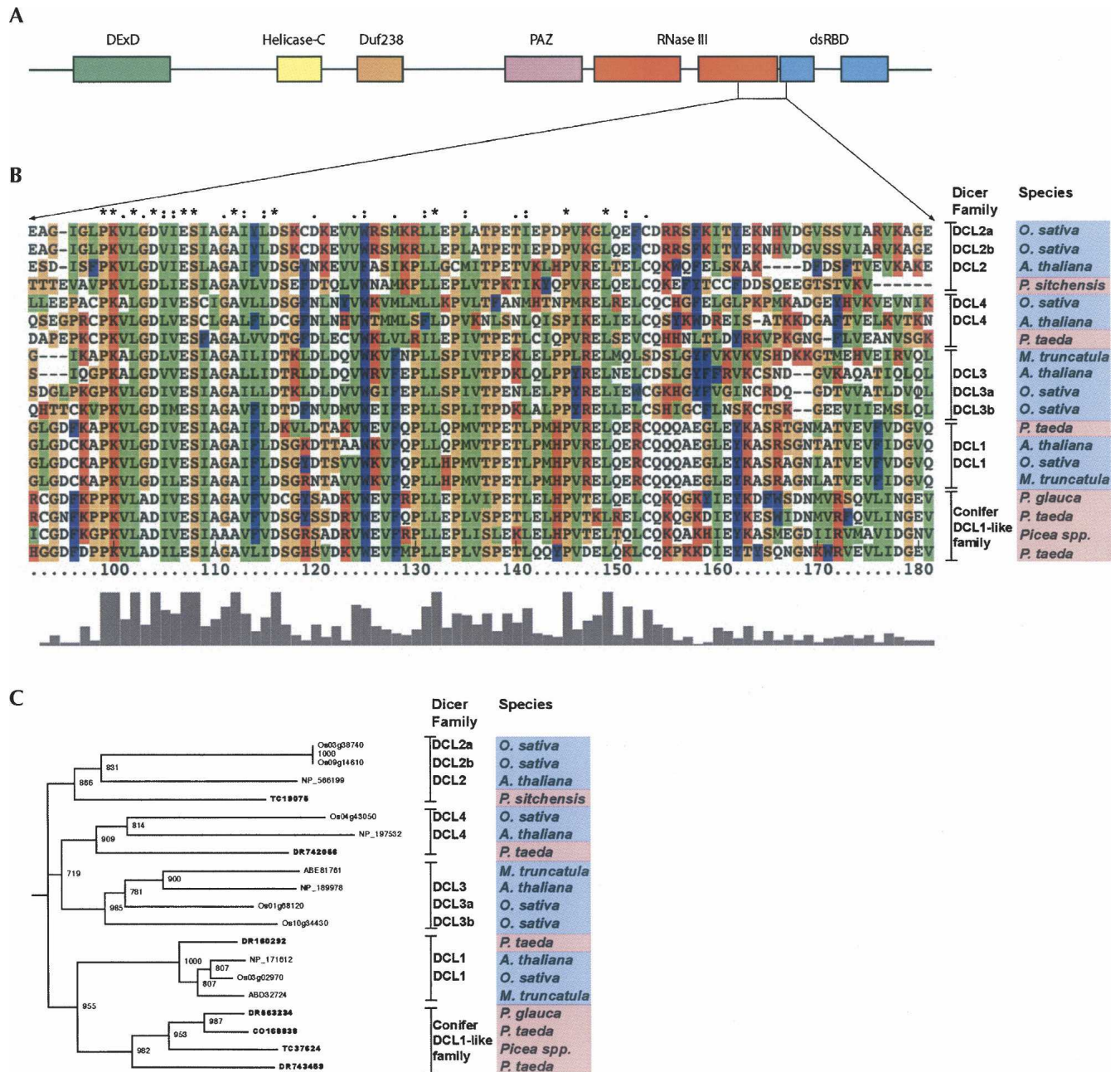


FIGURE 4. Identification of a new Dicer-like family in conifers. (A) Domain organization of a flowering plant Dicer. Dicers in *A. thaliana* and *O. sativa* have a total of eight domains with the following exceptions: DCL2 lacks second dsRBD; DCL3 lacks Duf283 (Margis et al. 2006). (B) Amino acid sequences corresponding to seven DCL-like ESTs from *P. taeda* and *Picea spp.*, and the C-terminal regions from four *A. thaliana* Dicers, six *O. sativa* Dicers, and two predicted Dicers from *M. truncatula* were aligned using ClustalX and default parameters. Only the alignment of the region exhibiting the highest sequence conservation is shown (90 amino acids). This region is predicted to correspond to the end of the second RNase III domain and to part of the dsRBD. Color code for amino acid residues: (orange) GPST; (red) HKR; (blue) FWY; (green) ILMV. Bars below the alignment indicate prevalence of the most common amino acid residue at each position, with the highest bar corresponding to 100% conservation. (C) Phylogenetic tree of Dicers in the seed-bearing plants generated from the alignment in B. Positions with gaps were excluded during tree building.

133, 155, 158, 172, and 181) that were not commonly found in the other DCL sequences used to construct the alignment. Considering that these residues lie in a region of high overall sequence conservation, their distinctiveness

suggests that the corresponding proteins are divergent from the established DCL families found in angiosperms. Unfortunately, since ESTs span only the C-terminal protein regions, we were not able to find any ESTs homologous

to other Dicer domains toward the N terminus of the protein (e.g., PAZ, DexD-helicase).

The alignment shown in Figure 4B was used to generate the most likely phylogenetic tree of Dicers in the seed-bearing plants (Fig. 4C, based on 1000 rounds of bootstrapping). One conifer EST (DR160292, *P. taeda*) aligned very well with the angiosperm DCL1, which together with the presence of highly conserved miRNAs in *P. contorta* provides compelling evidence for the conservation of miRNA-generating pathways between conifers and angiosperms (Fig. 2C). Orthologs for angiosperm DCL2 (TC19075, *Picea sitchensis*) and DCL4 (DR742056, *Pinus taeda*) were also found. While we have no direct experimental evidence for the function of DCL2 and DCL4 in conifers, their sequence conservation with the corresponding Dicers in angiosperms suggests that they play a similar role. The remaining four conifer ESTs formed a new DCL family that is distinct from any known angiosperm DCL family, but is most similar to DCL1. Although the role of this new family is unknown, we hypothesize that this family is responsible for production of the diverse set of 21-nt small RNAs that are so highly expressed in conifers. Interestingly, the most similar sequences in this new family are from different species (DR563234, *Picea glauca* and CO168838, *P. taeda*) suggesting a conserved functionality unique to these two proteins. Finally, no evidence for ESTs containing a DCL3-like coding region was found in any of the conifer libraries screened.

It is believed that all DCL genes derive from a single parental gene and that the expansion of this family occurred in plants prior to the divergence of monocots from eudicots (~200 mya) (Cerutti and Casas-Mollano 2006; Margis et al. 2006). The survey that we have performed places the DCL gene family expansion prior to the angiosperm–gymnosperm radiation (~300 mya) (Schneider et al. 2004) and suggests that RNA silencing in conifers has been subjected to a unique set of selective pressures relative to other land plants. While we provide strong evidence for the conservation of the DCL1-mediated RNA processing and weaker support for that of DCL2 and DCL4, conifers appear to be distinct in that they seemingly lack DCL3 and express a new DCL family not found in angiosperms. These differences between gymnosperms and angiosperms are intriguing from an evolutionary point of view. Since we found neither the DCL3-produced 24-nt RNA class, nor a DCL3-like EST in conifers, DCL3 could have either been lost or its expression could have been severely reduced in these plants. The presence of the DCL3 pathway in more ancient lineages of land plants is consistent with the 24-nt RNA expression observed in ferns (Fig. 1, lanes 4,5), combined with recent data that suggest that the anciently divergent moss *P. patens* possesses a gene orthologous to the angiosperm DCL3 (Axtell et al. 2007). In this case, the DCL3 pathway appears to be truly ancient and presumably evolved prior to the radiation of the seed-bearing plants.

Both DCL3 and 24-nt RNAs are key players in siRNA-directed DNA methylation and heterochromatin formation in flowering plants (Chan et al. 2004; Matzke and Birchler 2005; Wassenegger 2005; Henderson and Jacobsen 2007). If this pathway was lost from the gymnosperm/conifer lineage, it is possible that these plants rely on RNA silencing-independent transcriptional gene silencing. The partial or complete loss of RNA silencing pathways, thought to be present in the last common ancestor of eukaryotes, has been well documented in many organisms (Cerutti and Casas-Mollano 2006), and RNA silencing has been found to differ substantially even between members of the same phylum. For example, in RNA-silencing-positive eukaryotes such as *Neurospora crassa*, *Schizosaccharomyces pombe*, and *Dictyostelium discoideum*, RNA silencing machinery appears to be dispensable for DNA and histone methylation of repetitive elements (Freitag et al. 2004; Jia et al. 2004; Kaller et al. 2006). This certainly provides support that a similar loss could have occurred in conifers. Significantly, however, the hypothesis that conifers rely on RNA silencing-independent transcriptional gene silencing explains neither the existence of a new DCL family, nor the abundance of 21-nt RNAs that we observe in these plants.

Alternatively, and we think more likely, the discovery of a new DCL member and a dominant and sequence-rich 21-nt class of RNA in conifers may indicate that small RNA-mediated regulation of heterochromatin has undergone a significant diversification that has resulted in the evolution of novel DCL genes—and by implication, RNA silencing pathways. Considering that the diverse population of DCL3-produced 24-nt RNAs in *A. thaliana* targets the methylation of repetitive and transposable DNA elements (Chan et al. 2004; Xie et al. 2004; Matzke and Birchler 2005), it is likely that, in the absence of the 24-nt RNA class, conifers may use their diverse 21-nt class, presumably produced by the new DCL family, to satisfy a similar role. Yet, due to limitations of the EST data set, the absence of the sequenced gymnosperm genome and the lower sequence and domain structure conservation of DCL proteins relative to other RNA silencing effectors in eukaryotes (Cerutti and Casas-Mollano 2006), it cannot be ruled out that we simply missed the coding sequence for DCL3 in conifers. In this case, the apparent lack of DCL3-generated 24-nt-long RNAs and the unusual diversity of the 21-nt RNA class in conifers remains to be explained. Sequencing of the DCL proteins and/or the genome of a gymnosperm will shed further light on this important issue.

Interestingly, additional support for our hypothesis may stem from the poorly explained fact that gymnosperms have unusually large genomes, ranging in size from 2.1 to 37.0 Gb (1C values), in contrast to angiosperms, which have much smaller modal genome sizes of ~0.6 Gb (Leitch et al. 2005). Conifer genomes, in particular, while rarely polyploid, are larger than most animal and plant genomes, having a modal genome size of >15 Gb (Ahuja and Neale

2005). According to DNA reassociation kinetics data, 75% of these genomes are comprised of repetitive DNA that includes transposable and retroelements (Rake et al. 1980; Kriebel 1985). There is substantial evidence indicating that RNA silencing is responsible for keeping transposable and retroelements transcriptionally silenced in flowering plants, thereby limiting their propagation (Henderson and Jacobsen 2007). RNA interference in eukaryotic organisms other than plants has also been shown to play a role in packaging of repetitive sequences into chromosomes during mitosis/meiosis (Fukagawa et al. 2004; Martienssen et al. 2005), which ensures genomic integrity upon cell division. Given that eukaryotic genome size variation is accounted for by the differential amount of noncoding repetitive sequences (Kidwell 2002; Gregory 2005), and that transcriptional regulation of these sequences, whether RNA directed or not, affects genome stability (Grewal and Moazed 2003; Henderson and Jacobsen 2007), the difference in RNA silencing pathways that we have demonstrated to exist between conifers and angiosperms may reflect a variation in how these plants maintain and organize their differently sized genomes. Future studies of small RNA biogenesis and function in the gymnosperms will broaden our understanding of the evolution of RNA silencing in plants and its role in genome organization.

MATERIALS AND METHODS

Total RNA extraction

Total RNA from angiosperms, pteridophytes, *Ephedra sp.*, *Cycas revoluta*, and *Ginkgo biloba* was extracted from young, actively growing leaf buds. Total RNA from the rest of the gymnosperms was extracted from young needle tips. A detailed extraction protocol is available in Morin et al. (2008). The concentration of total RNA was measured using a UV-visible spectrophotometer (UV-1601; SHIMADZU).

RNA labeling

RNA extracts were 5'-phosphorylated at the concentrations of 60–200 ng/ μ L for 35 min at 37°C using 0.33 μ M [γ -³²P]ATP (NEN), 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 50 mM Tris-HCl (pH 7.9), and 0.5 U/ μ L PNK (New England Biolabs). Labeled RNA extracts were resolved on a 20% denaturing polyacrylamide gel. Synthetic 21- and 24-nt RNAs (Dharmacon) were used as size standards together with 70-nt and 125-nt T7 RNA polymerase transcripts.

Sodium periodate treatment of small RNAs

Radiolabeled total RNA was phenol/chloroform-extracted and resuspended in ddH₂O. A small fraction was then incubated in 10 mM HEPES (pH 7.4) and 100 mM sodium periodate for 10 min at 22°C. An equal volume of formamide loading dye (supplemented with 5 mM EDTA) was added, and β -elimination was facilitated by heating the reaction mixture for 30 min at 100°C. Products of the reaction were resolved on a denaturing 20% polyacrylamide gel. A synthetic 11-nt RNA marker was radio-

labeled as described above (RNA labeling section) and added to the plant sample as an internal control prior to periodate cleavage. Synthetic 21- and 24-nt RNAs (Dharmacon) were used as size standards.

Dicer alignment and phylogeny

DCL sequences from *A. thaliana* and *M. truncatula* were accessed via GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>). *O. sativa* DCL sequences were obtained from the Institute for Genomic Research (TIGR) rice database (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice). Conifer EST libraries were searched at the Genomic Biology database (<http://www.ncbi.nlm.nih.gov/Genomes/>) and at the Computational Biology and Functional Genomics Laboratory database (<http://compbio.dfci.harvard.edu/tgi/>) with Dicer protein sequences from *A. thaliana*. Tblastn and default parameters were used in similarity searches. Accession numbers were as follows: *A. thaliana*: NP_171612, NP_566199, NP_189978, NP_197532/DQ118423; *O. sativa*: Os03g02970, Os03g38740, Os09g14610, Os01g68120, Os10g34430, Os04g43050; *M. truncatula*: ABD32724, ABE81761; *P. sitchensis*: TC19075; *P. taeda*: DR160292, DR742056, DR743459, CO168838; *P. glauca*: DR563234; *Picea spp.*: TC37624.

Predicted amino acid coding sequences were generated using the ExPASy translate tool (<http://ca.expasy.org/tools/dna.html>), and the sequences with the longest open reading frames were used for further analysis. Protein domains of the translated ESTs were analyzed by scanning predicted protein sequences against the InterPro protein signature database (<http://www.ebi.ac.uk/InterProScan>). Domains were assigned according to pFAM predictions.

Alignment was generated using ClustalX and default parameters. The same program was used to produce the phylogenetic tree of Dicers in the seed-bearing plants using the neighbor-joining algorithm with 1000 rounds of bootstrapping. Positions with gaps were excluded. The phylogenetic tree was visualized using Phylodendron (<http://iubio.bio.indiana.edu/treeapp/treeprint-form.html>).

SUPPLEMENTAL DATA

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

ACKNOWLEDGMENTS

This work was supported by grants from the Canadian Institutes of Health Research (to P.J.U.) and the Michael Smith Foundation for Health Research (to P.J.U.). E.V.D. designed and performed experiments and contributed to the writing of the manuscript. R.D.M. performed bioinformatic analysis together with G.A. and S.C.S. V.M. and E.R.M. performed the 454 sequencing of the plant samples. J.M. provided plant and RNA samples and contributed to writing. P.J.U. designed experiments and contributed to the writing of the manuscript.

Received February 29, 2008; accepted April 25, 2008.

REFERENCES

Ahuja, M.R. and Neale, D.B. 2005. Evolution of genome size in conifers. *Silvae Genet.* **54**: 126–137.

- Axtell, M.J., Snyder, J.A., and Bartell, D.P. 2007. Common functions for diverse small RNAs of land plants. *Plant Cell* **19**: 1750–1769.
- Borsani, O., Zhu, J.H., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**: 1279–1291.
- Brodersen, P. and Voinnet, O. 2006. The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**: 268–280.
- Cerutti, H. and Casas-Mollano, J.A. 2006. On the origin and functions of RNA-mediated silencing: From protists to man. *Curr. Genet.* **50**: 81–99.
- Chan, S.W.L., Zilberman, D., Xie, Z.X., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. 2004. RNA silencing genes control *de novo* DNA methylation. *Science* **303**: 1336.
- Deleris, A., Gallego-Bartolome, J., Bao, J.S., Kasschau, K.D., Carrington, J.C., and Voinnet, O. 2006. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* **313**: 68–71.
- Dlagic, M. 2006. DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics* **22**: 2711–2714.
- Ebhardt, H.A., Thi, E.P., Wang, M.B., and Unrau, P.J. 2005. Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. *Proc. Natl. Acad. Sci.* **102**: 13398–13403.
- Freitag, M., Lee, D.W., Kothe, G.O., Pratt, R.J., Aramayo, R., and Selker, E.U. 2004. DNA methylation is independent of RNA interference in *Neurospora*. *Science* **304**: 1939.
- Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. 2004. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat. Cell Biol.* **6**: 784–791.
- Gascioli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr. Biol.* **15**: 1494–1500.
- Gregory, R.T. 2005. The evolution of the genome. Elsevier AP, Boston.
- Grewal, S.I.S. and Moazed, D. 2003. Heterochromatin and epigenetic control of gene expression. *Science* **301**: 798–802.
- Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. 2002. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**: 4671–4679.
- Hammond, S.M. 2005. Dicing and slicing—The core machinery of the RNA interference pathway. *FEBS Lett.* **579**: 5822–5829.
- Henderson, I.R. and Jacobsen, S.E. 2007. Epigenetic inheritance in plants. *Nature* **447**: 418–424.
- Henderson, I.R., Zhang, X.Y., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**: 721–725.
- Jia, S.T., Noma, K., and Grewal, S.I.S. 2004. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**: 1971–1976.
- Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**: 19–53.
- Kaller, M., Euteneuer, U., and Nellen, W. 2006. Differential effects of heterochromatin protein 1 isoforms on mitotic chromosome distribution and growth in *Dictyostelium discoideum*. *Eukaryot. Cell* **5**: 530–543.
- Kenrick, P. and Crane, P.R. 1997. The origin and early evolution of plants on land. *Nature* **389**: 33–39.
- Kidwell, M.G. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**: 49–63.
- Kriebel, H.B. 1985. DNA sequence components of the *Pinus strobus* nuclear genome. *Can. J. For. Res.* **15**: 1–4.
- Leitch, I.J., Soltis, D.E., Soltis, P.S., and Bennett, M.D. 2005. Evolution of DNA amounts across land plants (embryophyta). *Ann. Bot. (Lond.)* **95**: 207–217.
- Lu, C., Tej, S.S., Luo, S.J., Haudenschild, C.D., Meyers, B.C., and Green, P.J. 2005. Elucidation of the small RNA component of the transcriptome. *Science* **309**: 1567–1569.
- MacRae, I.J. and Doudna, J.A. 2007. Ribonuclease revisited: Structural insights into ribonuclease III family enzymes. *Curr. Opin. Struct. Biol.* **17**: 138–145.
- MacRae, I.J., Zhou, K.H., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. 2006. Structural basis for double-stranded RNA processing by dicer. *Science* **311**: 195–198.
- Margis, R., Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J., and Waterhouse, P.M. 2006. The evolution and diversification of Dicers in plants. *FEBS Lett.* **580**: 2442–2450.
- Martienssen, R.A., Zaratiegui, M., and Goto, D.B. 2005. RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. *Trends Genet.* **21**: 450–456.
- Matzke, M.A. and Birchler, J.A. 2005. RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet.* **6**: 24–35.
- Morin, R.D., Askay, G., Dolgosheina, D., Ebhardt, H.A., Magrinio, V., Mardiso, E.R., Cenk Sahinalp, S.C., and Unrau, P.J. 2008. Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res.* **18**: 571–584.
- Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Dev.* **20**: 3407–3425.
- Rake, A.V., Miksche, J.P., Hall, R.B., and Hansen, K.M. 1980. DNA reassociation kinetics of four conifers. *Can. J. Genet. Cytol.* **22**: 69–79.
- Schneider, H., Schuettpelz, E., Pryer, K.M., Cranfill, R., Magallon, S., and Lupia, R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* **428**: 553–557.
- Wassenegger, M. 2005. The role of the RNAi machinery in heterochromatin formation. *Cell* **122**: 13–16.
- Xie, Z.X., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**: 642–652.
- Xie, Z.X., Allen, E., Wilken, A., and Carrington, J.C. 2005. DICER-LIKE 4 functions in *trans*-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **102**: 12984–12989.
- Yu, B., Yang, Z.Y., Li, J.J., Minakhina, S., Yang, M.C., Padgett, R.W., Steward, R., and Chen, X.M. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**: 932–935.



RNA

A PUBLICATION OF THE RNA SOCIETY

Conifers have a unique small RNA silencing signature

Elena V. Dolgosheina, Ryan D. Morin, Gozde Aksay, et al.

RNA 2008 14: 1508-1515

**Supplemental
Material**

<http://rnajournal.cshlp.org/content/suppl/2008/06/20/rna.1052008.DC1.html>

References

This article cites 38 articles, 17 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/14/8/1508.full.html#ref-list-1>

**Email Alerting
Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



Rudi Micheletti uses LNA™
GapmeRs to silence cardiac lncRNAs
www.exiqon.com/gapmers

EXIQON

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
