Published online 22 December 2008

Nucleic Acids Research, 2009, Vol. 37, No. 3 ***903–915***
doi:10.1093/nar/gkn1018

Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*

Gersende Lepère[1,2], Mariusz Nowacki[1,2], Vincent Serrano[1,2], Jean-François Gout[3], Gérard Guglielmi[1,2], Sandra Duharcourt[1,2] and Eric Meyer[1,2,*]

[1]Ecole Normale Supérieure, Laboratoire de Génétique Moléculaire, [2]CNRS, UMR8541, 46 rue d'Ulm, 75005 Paris, [3]Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

Received September 11, 2008; Revised December 4, 2008; Accepted December 7, 2008

ABSTRACT

Distinct small RNA pathways are involved in the two types of homology-dependent effects described in Paramecium tetraurelia, as shown by a functional analysis of Dicer and Dicer-like genes and by the sequencing of small RNAs. The siRNAs that mediate post-transcriptional gene silencing when cells are fed with double-stranded RNA (dsRNA) were found to comprise two subclasses. *DCR1*-dependent cleavage of the inducing dsRNA generates ~23-nt primary siRNAs from both strands, while a different subclass of ~24-nt RNAs, characterized by a short untemplated poly-A tail, is strictly antisense to the targeted mRNA, suggestive of secondary siRNAs that depend on an RNA-dependent RNA polymerase. An entirely distinct pathway is responsible for homology-dependent regulation of developmental genome rearrangements after sexual reproduction. During early meiosis, the *DCL2* and *DCL3* genes are required for the production of a highly complex population of ~25-nt scnRNAs from all types of germline sequences, including both strands of exons, introns, intergenic regions, transposons and Internal Eliminated Sequences. A prominent 5′-UNG signature, and a minor fraction showing the complementary signature at positions 21–23, indicate that scnRNAs are cleaved from dsRNA precursors as duplexes with 2-nt 3′ overhangs at both ends, followed by preferential stabilization of the 5′-UNG strand.

INTRODUCTION

The hallmark of RNA interference (RNAi) pathways is the use of small RNAs (sRNAs), bound by proteins of the Argonaute family, as specificity subunits that allow the recognition of target sequences by pairing interactions. In most eukaryotes, RNAi will process aberrant transcripts or experimentally introduced double-stranded RNA (dsRNA) into small interfering RNAs (siRNAs) that target homologous mRNAs for degradation, resulting in post-transcriptional gene silencing (PTGS). In plants and animals, miRNAs are processed from endogenously encoded non-coding transcripts and mediate PTGS through mRNA degradation or translation inhibition. These related pathways depend on dsRNA-specific ribonucleases of the Dicer family, which produce sRNA duplexes of 21–24 nt with characteristic 2-nt 3′ overhangs at both ends (1). In *Caenorhabditis elegans*, however, the RNAi response to dsRNA contained in food bacteria was shown to involve not only Dicer-dependent primary siRNAs, but also secondary siRNAs that are not Dicer products but instead appear to be synthesized by an RNA-dependent RNA polymerase (RdRP) from the targeted mRNA (2–4).

Dicer proteins are further involved in specialized RNAi pathways that control gene expression at the transcriptional level. In plants and fungi, siRNAs are believed to target chromatin-modifying complexes to specific loci by pairing with nascent transcripts (5). In contrast, the biogenesis of the recently described piRNAs of metazoans appears to be Dicer-independent. PiRNAs are germline-specific sRNAs that associate with the Piwi subfamily of Argonaute proteins and, at least in vertebrates, are

*To whom correspondence should be addressed. Tel: +33 1 44 32 39 48; Fax: +33 1 44 32 39 41; Email: emeyer@biologie.ens.fr
Present addresses:
Gersende Lepère, Laboratoire de Biologie Cellulaire, Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique (INRA), 78026 Versailles Cedex, France
Mariusz Nowacki, Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
Vincent Serrano, Institut de Génétique Humaine, Centre National de la Recherche Scientifique, 34396 Montpellier Cedex 5, France

specifically produced during meiosis (6–10). Although their biological functions are not fully understood, they have been implicated in the repression of germline retrotransposons by mechanisms that include the targeting of chromatin modifications (8,11–14).

In ciliates, at least two distinct small RNA pathways appear to mediate different types of homology-dependent effects. PTGS was first shown to be induced in *Paramecium tetraurelia* by high-copy, untranslatable transgenes (15,16), or by feeding cells with bacteria producing dsRNA (17). In both cases, silencing of the targeted cellular gene correlates with the accumulation of homologous ~23-nt siRNAs (18,19). In the related *Tetrahymena thermophila*, a similar class of ~23–24-nt RNAs is produced endogenously from a limited number of genomic loci (20). That the latter represent endogenous siRNAs is suggested by the fact that experimentally induced hairpin transcripts, which cause the silencing of homologous genes, are processed into siRNAs of similar lengths (21). *In vitro* experiments have implicated Dcr2, one of two Dicer proteins in this organism, and the RdRP Rdr1 in the biogenesis of ~23–24-nt siRNAs (22).

The second small RNA pathway described in ciliates is restricted to the sexual phase of the life cycle. In these unicellular eukaryotes, germline and somatic functions are ensured by two different kinds of nuclei, the diploid micronucleus (MIC) and the highly polyploid macronucleus (MAC). Sexual events are initiated by meiosis of the MIC. Following fertilization, new MIC and MAC develop from copies of the zygotic nucleus. MAC development involves extensive rearrangements of the germline genome, including the elimination of transposable elements and other repeated sequences and the precise excision of numerous single-copy Internal Eliminated Sequences (IESs) (23–26). Elimination of MIC-specific DNA was shown to be targeted by histone H3 modifications in several species (27–30). In *T. thermophila*, meiosis-specific Dicer-like and Piwi-like proteins (Dcl1 and Twi1) are required for the accumulation of scnRNAs ('scan' RNAs), a complex class of ~26–31-nt sRNAs produced by meiotic MIC (20,31–33). These proteins are also required for DNA elimination in the developing MAC, suggesting that rearrangement patterns are determined by scnRNA-directed chromatin modifications.

A large body of experimental evidence has shown that alternative rearrangements of the same germline genome can be reproduced from maternal MAC to zygotic MAC, implying a *trans*-nuclear comparison of germline and somatic genomes (34,35). Homology-dependent maternal effects can inhibit the excision of a subset of *P. tetraurelia* IESs, called maternally controlled IESs (mcIESs) (36), and can also mediate the elimination of cellular genes from the MAC genome (18). To explain similar effects in *T. thermophila*, it has been proposed that scnRNAs, which appear to become progressively enriched in MIC-specific sequences after meiosis and to move from parental to zygotic MAC (37), are involved in this *trans*-nuclear comparison. ScnRNAs that can pair with homologous sequences in the parental MAC (MAC scnRNAs) would be degraded, while the MIC-specific ones would be exported to the zygotic MAC to target DNA elimination (38). In *P. tetraurelia*, the recognition and inactivation of MAC scnRNAs appears to be determined by pairing with non-coding transcripts from the maternal MAC, rather than with DNA (18,39), and this may also be true in *T. thermophila* (40,41).

Here, we show that different Dicer and Dicer-like genes are involved in the biosynthesis of *P. tetraurelia* siRNAs and scnRNAs, and present the first analysis of small RNA sequences from both classes.

MATERIALS AND METHODS

Paramecium strain and cultivation

The entirely homozygous strain 51 was grown at 27°C in a wheatgrass-powder infusion medium bacterized with *Klebsiella pneumoniae* and supplemented with 0.8 mg l^{-1} β-sitosterol as described (19). Conjugation was carried out with very young cells (≤5 divisions since the last meiosis) to minimize the occurrence of autogamy upon starvation. Briefly, old cell lines of both mating types were allowed to starve in 100 ml (~400 000 cells) to induce 100% autogamy (fragmentation of the old MAC was monitored by carmine red/fast green staining). Post-autogamous cells were fed with four volumes of bacterized medium to allow two divisions before starvation. Aliquots of starved cells were then fed with *K. pneumoniae* or *Escherichia coli* for ≤3 divisions; upon starvation, cells become reactive and ready for conjugation. Autogamy was induced by starvation of old cells (≥20 divisions). Unlike conjugation which can be initiated in a synchronous manner by the mixing of mating types, cells enter autogamy from a fixed point of the cell cycle, leading to a minimal asynchrony of ~6 h.

DNA and RNA extraction, Southern and northern blot analyses

These were as described (19), except that small-RNA northerns were washed in 2× SSC, 0.1% SDS at ≤25°C to allow the detection of scnRNAs.

5′-end labelling of small RNAs

Total RNA samples (10 μg for Fig. 1A) were dephosphorylated using Calf Intestine Phosphatase, extracted with acid phenol, and labelled using T4 Polynucleotide Kinase and [γ-^{32}P]ATP.

dsRNA feeding experiments

Target sequences were cloned into plasmid L4440. The sequences used for *ND7* and *NOWA1* silencing, and the growth and IPTG induction of transformed *E. coli* (HT115) clones, were as described (19). Sequences used for silencing of *DCR1*, *DCL2* and *DCL3* were segments 3019–3987, 434–1032, and 630–1468 of GSPATG00021751001, GSPATG00008494001 and GSPATG00027456001 (see ParameciumDB at http://paramecium.cgm.cnrs-gif.fr/), respectively.

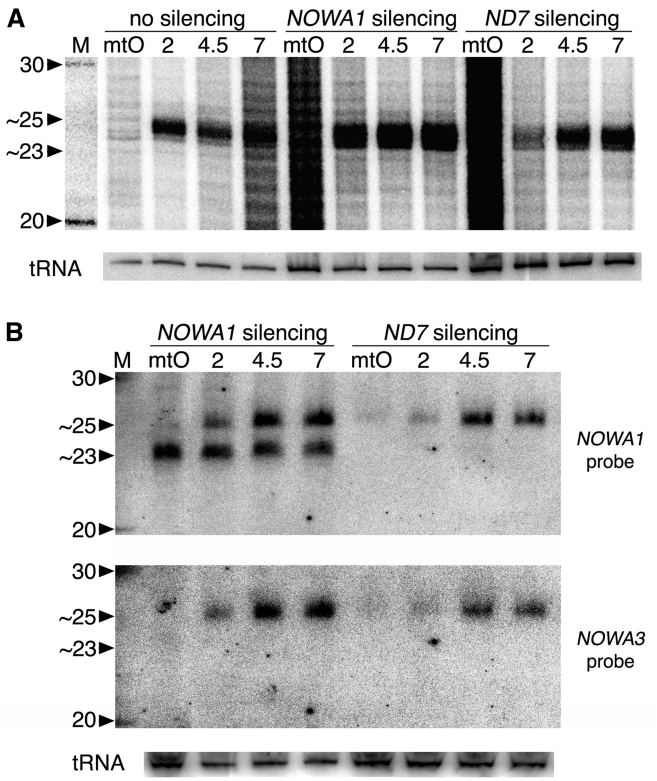

Figure 1. Evidence for meiosis-specific ~25-nt RNAs distinct from the ~23-nt siRNAs. (**A**) Total RNA samples from starved reactive cells (mtO, mating type O; mtE gave similar results) just before the mixing of mating types, or at different times post-mixing (2, 4.5 and 7 h), were 5′-end-labelled and run on a 15% polyacrylamide-urea gel. The size marker (M) shows the positions of 20 and 30-nt RNAs. Three different experiments are shown: before conjugation, cells were fed either with *Klebsiella* (no silencing) or with *E. coli* strains producing dsRNA homologous to parts of the *NOWA1* or *ND7* coding sequences, as indicated. The high background seen in reactive cells (mtO) silenced for these genes is likely due to bacterial RNA being degraded. The lower panel (tRNA) shows subsequent hybridization of the same membrane with a tRNA probe, as a loading control. (**B**) A similar gel of samples from the conjugation of *NOWA1*- and *ND7*-silenced cells was blotted and revealed successively with probes specific for the *NOWA1* dsRNA sequence, for the *NOWA3* gene, and for the same tRNA as in A, as indicated. The faint band seen at ~25 nt in reactive cells (mtO) prior to the initiation of conjugation is due to a small fraction of cells (<3%) undergoing autogamy, a self-fertilization sexual process.

Constructs, probes and oligonucleotides

The tRNA probe in Figure 1A is a 75-nt oligonucleotide complementary to a Gln tRNA (scaffold_39-tRNA1, see ParameciumDB), including CCA extension, end-labelled with T4 kinase. The NOWA1 probe in Figure 1B is the same portion of the coding sequence as used for dsRNA feeding, *i.e.* a PCR fragment covering positions 2401–3246 of AJ876761, labelled by random priming. The NOWA3 probe covered segment 507–1053 of GSPATG00006397001 (see ParameciumDB). The silencing A^{51} transgene, microinjection method, and A^{51}-specific probe have been described (18). Probes used for *DCR1*, *DCL2* and *DCL3* were a plasmid containing the entire *DCR1* gene and 735/184 bp of upstream/downstream sequences, and PCR fragments covering the entire *DCL2* and *DCL3* coding sequences.

Cloning of sRNAs and hybridization-selection

Small RNAs were purified on 15% polyacrylamide (acrylamide:bis 19:1)-7 M urea gels and cloned according to the procedure described in (42). For hybridization-selection, the sequences of interest (2–5 μg of PCR fragments) were fixed on strips of Hybond-N+ membranes (Pharmacia) with 0.4 N NaOH and hybridized in Church buffer with ~2 μg of denatured RT-PCR products from the 25-nt sRNA fractions, in the presence of a 100-fold molar excess of each of the PCR primers to avoid the concatemerization of all RT-PCR products that would result from the hybridization of the adaptors. DNA was recovered from small volumes of 2× SSC, 0.1% SDS used to wash the membranes with 5°C temperature increments and reamplified before cloning. After washing at 40–50°C and eluting at 100°C, 20–80% of reamplified sRNAs matched the target sequences.

Microarray expression data

Expression data were obtained from single channel Nimblegen® microarrays with six different 50-mer probes per gene. Signals from microarrays from four biological replicates were normalized together and $\log_2$-transformed. The expression level of each gene at each stage (vegetative growth and early autogamy) was taken as the mean of the signals from the six probes over the four biological replicates. Significant differences between means in the two stages were assessed with a two-sided Student's *t*-test.

Bioinformatic analyses

Small RNA sequences were searched in relevant databases using the fuzznuc and blast softwares. To classify the 641 25-nt scnRNA sequences, a position-specific weighting matrix was constructed using the frequency of each base observed at each position in the alignment of the whole set (known to contain an excess of guide strands), divided by the frequency of that base in the macronuclear genome (0.36 for A and U, 0.14 for G and C). The score of each sequence was defined as the sum of the $\log_2$ of values corresponding to each base for the first 23 positions. A scnRNA was classified as a guide if its score was greater than that of its reverse complement. Sequence logos were computed using the WebLogo (http://weblogo.berkeley.edu/logo.cgi) and RNA Structure Logo (http://www.cbs.dtu.dk/~gorodkin/appl/slogo.html#form) web sites.

RESULTS

Evidence for meiosis-specific ~25-nt scnRNAs in *P. tetraurelia*

To examine the population of endogenous sRNAs during sexual reproduction, total RNA was extracted at different time points from a culture undergoing synchronous conjugation. Total RNA samples were dephosphorylated, labelled with P^{32} using T4 kinase, and run on a 15% polyacrylamide-urea gel. In the 20–30-nt range, only background signal was visible in starved, sexually reactive cells of mating types O and E prior to their mixing,

which initiates the pairing of cells and conjugation (Figure 1A, no silencing, mtO). In contrast, a homogeneous class of ~25-nt RNAs was readily detected as soon as 2 h after the mixing, well before the first meiotic division of micronuclei. These RNAs were also easily seen by ethidium bromide staining of 20 µg of total RNA (data not shown), indicating that they are produced in massive amounts. To illustrate the size difference between these meiosis-specific sRNAs and the ~23-nt siRNAs associated with PTGS, conjugation was induced between reactive cells that had previously been grown on dsRNA-producing *E. coli* to silence the *NOWA1* or *ND7* genes. SiRNAs were not readily detected by 5′-end labelling, presumably because of their much lower abundance (Figure 1A). They could be specifically revealed, however, by northern blotting of the same samples after hybridization with homologous probes (Figure 1B). Such northern blots further showed that the ~25-nt RNAs contained sequences homologous to the *NOWA1* and *NOWA3* genes (Figure 1B), as well as a wide variety of other genomic sequences including MIC-specific elements such as transposons (data not shown) and IESs (39). Thus, a very complex population of ~25-nt RNAs appears to be produced by micronuclei at the very beginning of meiosis. These will be called scnRNAs because of their similarity to those described in *T. thermophila*.

Relative to the tRNA loading control, the total amount of scnRNAs appeared to decrease between 2 and 7 h in the wild-type conjugation, and to increase during the same period of time when the *NOWA1* or *ND7* genes were silenced. While this difference may be a consequence of the dsRNA feeding technique, it is not possible to tell whether it results from a true effect on scnRNA processing or from a more general effect on the efficiency or synchrony of conjugation. Similarly, scnRNA levels appeared to be higher in *NOWA1*-silenced than in wild-type or *ND7*-silenced cells (Figure 1A and B), but more experiments will be needed to determine whether this is linked to the putative role of the Nowa1 protein in scnRNA function (19).

Functional analysis of Dicer and Dicer-like genes

The *P. tetraurelia* MAC genome (43) contains seven genes encoding proteins with RNaseIII domains (Figure 2A). Three of these (*DCR1-3*) are typical Dicer genes, containing the two DExH helicase subdomains, followed by a DUF283 domain, and two RNaseIII domains near the C-terminus. However, alignment of protein sequences suggests that Dcr2 and Dcr3 are not catalytically active, as is the case of *T. thermophila*'s Dcr1: the four catalytic residues of both RNaseIII domains are all mutated in these 3 proteins (Supplementary Figure S1). It is also unclear whether the ciliate Dicers contain functional PAZ and DSRM domains, because sequences are quite divergent in these regions. The remaining four genes only contain two RNAseIII domains and a ~350-aa N-terminal extension with no homology in databases, except to each other. These are structurally closer to *T. thermophila*'s *DCL1* and were called *DCL1-4*; all RNaseIII domains contain the four catalytic residues except for the second domain of

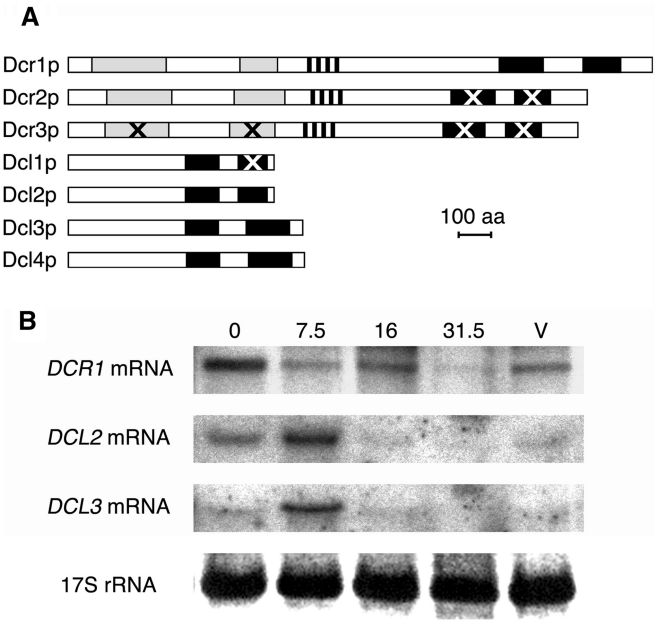

Figure 2. Dicer and Dicer-like genes in *P. tetraurelia*. (**A**) Schematic representation of protein domains. Grey boxes, helicase and helicase C-terminal domains. Hatched boxes, DUF283. Black boxes, RNaseIII domains. Black or white Xs mark domains that are unlikely to be catalytically active because of non-conservative substitutions at key residues. For accession numbers, see Supplementary Figures S1 and S2. (**B**) Northern blot analysis of the expression of the *DCR1*, *DCL2* and *DCL3* mRNAs in total RNA samples extracted at different times during autogamy. The 0 time point was arbitrarily set when 12% of cells in the culture had undergone meiosis, as indicated by their fragmented MAC. That percentage was 50% at 7.5 h, 98% at 16 h, and 100% at 31.5 h. Developing new MAC was visible in 20% of cells at 16 h and 98% at 31.5 h. V, total RNA extracted from starved vegetative cells at 84 h, after refeeding the culture for a few vegetative divisions. The bottom panel shows 17S rRNA (negative of ethidium bromide stain, before blotting) as a loading control.

Dcl1 (Supplementary Figure S2). Expression patterns were determined for some of these genes by probing a northern blot of RNA samples extracted at different times from a culture undergoing autogamy, a self-fertilization sexual process (Figure 2B). The *DCR1* mRNA appeared to be expressed throughout the life cycle, which would be consistent with a role in the vegetative siRNA pathway, whereas *DCL2* and *DCL3* mRNAs were upregulated during early autogamy, suggesting they may be involved in scnRNA biogenesis. These expression patterns were confirmed by microarray data (Table 1). Microinjection of a *DCL2-GFP* fusion construct further showed that the fusion protein accumulated in the MIC, but not in the MAC, at the beginning of meiosis; by the end of meiosis I, it was excluded from all nuclei (Supplementary Figure S3).

To test the implication of the *DCR1* gene in siRNA-mediated PTGS, cells were microinjected with an *ND7* transgene lacking the 3′UTR (15) to silence the endogenous *ND7* gene, resulting in a trichocyst non-discharge phenotype. *ND7* silencing was suppressed within 48 h when transformed clones were fed an *E. coli* strain producing *DCR1* dsRNA, but not when they were fed *DCL1*, *DCL2* or *DCL3* dsRNAs. Similarly, double-feeding experiments

Table 1. Expression levels of Dicer and Dicer-like genes during the life cycle

Gene	Vegetative	Autogamy	*P*-value
DCR1	11.4	11.0	NS (0.4)
DCR2	10.4	9.6	$\mathbf{5 \times 10^{-4}}$
DCR3	9.6	9.3	NS (0.2)
DCL1	10.0	9.9	NS (0.6)
DCL2	9.5	11.1	$\mathbf{2 \times 10^{-3}}$
DCL3	9.3	12.2	$\mathbf{2 \times 10^{-7}}$
DCL4	10.5	10.3	NS (0.4)

The table gives normalized $\log_2$ values from microarrays hybridized with cDNA from vegetatively growing cells (Vegetative) or from early autogamy (Autogamy) (see Materials and methods section). The *P*-values in bold indicate significant differences in expression levels between these two stages.

in which wild-type cells were fed equal amounts of *ND7* and other dsRNAs showed that the *ND7*/*DCR1* combination resulted the next day in a transient *ND7* silencing phenotype that reverted to wild type after 48 h, while other combinations gave a stable non-discharge phenotype. That *DCR1* is involved in the synthesis of ~23-nt siRNAs was directly verified using clones transformed with a 3′-truncated A^{51} transgene (18) at moderately high copy numbers. ~23-nt A^{51} siRNAs could be detected in total RNAs extracted from three such clones grown in normal conditions, but not when the same clones were fed *DCR1* dsRNA (Figure 3A). Conversely, long A^{51} transcripts (~8000 nt) accumulated only in the latter case (Figure 3B). Additional experiments (not shown) indicated that dsRNA-induced silencing of *DCR1*, *DCR2*, or *DCR3* during autogamy did not prevent the development of a functional new MAC. However, the induction of gene deletions in the developing MAC by RNAi (18) appeared to require the expression of *DCR1* (data not shown).

To identify genes involved in the scnRNA pathway, cells were fed bacteria producing dsRNA homologous to each of the 4 *DCL* genes, either alone or in all possible combinations, for at least three divisions prior to induction of meiosis. Examination of post-autogamous progeny showed that all silencing combinations including either *DCL2* or *DCL3* had subtle effects on MAC development, including mating-type change from O to E and reversal of maternally inherited deletions of cellular genes in the MAC, while combinations involving both of these genes resulted in non-viable sexual progeny (data not shown). In the latter case cell death occurred within a few divisions after refeeding autogamous cells, suggesting new MACs were not functional. To determine whether these phenotypes were due to defects in scnRNA production or function, total RNA was extracted at different time points from a large-scale culture undergoing autogamy after dsRNA-induced silencing of *DCL2* and *DCL3*. In contrast with the non-silenced control culture, 5′-end labelling after dephosphorylation failed to reveal any accumulation of ~25-nt scnRNAs (Figure 3C). A small amount of slightly longer RNAs was detected at late time points, but similar bands were also present in late samples from the control, suggesting these could represent a different

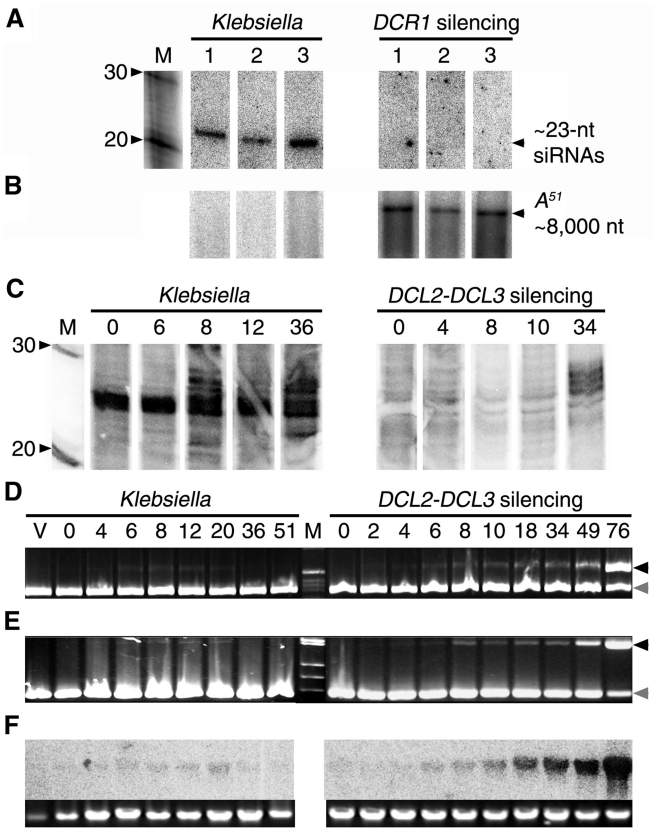

Figure 3. Functional analysis of Dicer and Dicer-like genes. (**A, B**) Northern blot analysis of total RNA from 3 clones transformed with a 3′-truncated A^{51} transgene (lanes 1–3), grown on the normal food bacterium (*Klebsiella*) or on *E. coli* producing *DCR1* dsRNA (*DCR1* silencing). RNAs were resolved on 15% polyacrylamide (A) or 1% agarose (B) denaturing gels, blotted and hybridized with an A^{51}-specific probe. The size marker (M) shows the positions of 20 and 30-nt RNAs. (**C**) 5′-end-labelled total RNA from different time points (as indicated in hours) during autogamy of cultures fed with *Klebsiella* or with *E. coli* producing *DCL2* and *DCL3* dsRNA prior to meiosis. The t = 0 time points were arbitrarily set when 5% (*Klebsiella*) or 8% (*DCL2-DCL3* silencing) of cells showed fragmented macronuclei; that fraction reached 54% at 6 and 4 h, and 99% at 20 and 18 h, respectively. (**D, E**) Total DNA samples from the same autogamy time courses, extracted at the indicated times (hours), were amplified with primers flanking IESs 51G2832 (D) or 51G4404 (E) (36). V, vegetative sample before induction of meiosis. IES-containing and IES-excised PCR fragments (black and grey arrowheads, respectively) are co-amplified in each reaction, allowing semi-quantitative assessment of the accumulation of unexcised IES copies in developing MAC. (**F**) A Southern blot of the same DNA samples (uncut) is hybridized with a probe specific for the *Sardine* transposon. The lower panel shows ethidium bromide staining as a loading control.

class of small RNAs accumulating at later stages of MAC development. Total DNA samples were also taken at different time points and subjected to PCR tests to monitor the excision of IESs during MAC development: silencing of *DCL2* and *DCL3* resulted in the accumulation of unexcised IESs, at least for two mcIESs (Figure 3D–E). A Southern blot analysis also revealed that a DNA transposon failed to be eliminated in these conditions (Figure 3F). Thus, *DCL2* and *DCL3* are required for developmental elimination of at least some MIC-specific DNA sequences.

Table 2. Number and percentage of sRNAs in each sequence category for each of the 8 size-selected fractions

Sample	*ND7* silencing	MAC genome	RIBO	EXO	No match	Total
2hWT—23 nt	0 (0%)	5 (5%)	32 (34%)	48 (52%)	8 (9%)	93 (100%)
2hWT—24 nt	0 (0%)	20 (18%)	45 (39%)	36 (32%)	13 (11%)	114 (100%)
2hWT—25 nt	0 (0%)	63 (45%)	23 (16%)	25 (18%)	30 (21%)	141 (100%)
2hWT—26 nt	0 (0%)	5 (7%)	44 (61%)	15 (21%)	8 (11%)	72 (100%)
7h*ND7*—23 nt	26 (20%)	20 (16%)	37 (29%)	30 (24%)	14 (11%)	127 (100%)
7h*ND7*—24 nt	19 (11%)	44 (24%)	50 (28%)	24 (13%)	43 (24%)	180 (100%)
7h*ND7*—25 nt	1 (1%)	97 (49%)	10 (5%)	3 (2%)	89 (45%)	200 (100%)
7h*ND7*—26 nt	0 (0%)	22 (23%)	32 (33%)	5 (5%)	38 (39%)	97 (100%)

Cloning and sequencing of size-selected small RNAs

To better characterize silencing-associated siRNAs and meiosis-specific scnRNAs, the small RNAs from two conjugation samples (wild-type 2 h and *ND7*-silenced 7 h time points, hereafter called 2hWT and 7h*ND7*, see Figure 1A) were purified on a 15% polyacrylamide-urea gel. For each sample, RNAs contained in four adjacent, 1-nt wide gel slices (23, 24, 25 and 26 nt) were eluted independently and cloned by a procedure involving dephosphorylation (42). A total of 1024 sequences were obtained by bulk sequencing of these eight fractions. Sequences were compared to the MAC genome, to the full set of predicted spliced mRNAs, to other *P. tetraurelia* sequences (rDNA and other abundant RNAs, mitochondrial genome, all known MIC-specific sequences), and to the *ND7* feeding vector, tolerating one mismatch to allow for RT-PCR-induced substitutions (Supplementary Tables 1 and 2). Because a significant fraction of the sRNAs eluted at 23 and 24 nt carried a 3′ polyA tail (see below), the search was repeated for unidentified sequences after removal of 3′-terminal As. Finally, the remaining unmatched sequences were searched in public databases. Table 2 gives the number of sequences identified in each of these categories for the eight fractions. As expected, sequences matching the *ND7* feeding vector were found only in the 7h*ND7* sample, where they represent 20% and 11% of the 23- and 24-nt fractions, respectively. A significant percentage of sequences in the 23, 24 and 26-nt fractions (28–61%) represented fragments of abundant non-coding RNAs such as rRNAs, tRNAs, snRNAs or mitochondrial transcripts (RIBO). In the 25-nt fractions, however, the RIBO category accounted for only 5–16% of sequences, reflecting the abundance of genomic scnRNAs. Similarly, between 2% and 52% of sequences appeared to be derived from RNAs of other organisms, mostly bacterial rRNAs (EXO). About one half of the remaining sequences matched the MAC genome. Those for which no match could be found could in principle arise from the unsequenced genomes of the MIC or of other organisms.

Polyadenylation of RIBO and EXO sRNAs

A suspiciously large number of sequences cloned from the 23-nt fraction of the 7h*ND7* sample contained stretches of As at their 3′ end and did not match any known sequence. Re-mapping the sequences after removal of the terminal As confirmed that the 3′ A stretches were untemplated polyA tails present in a significant fraction of sequences of all categories: 68% of RIBO, 20% of EXO, 50% of silencing-associated and 60% of genomic sRNAs (Supplementary Table 2). However, polyadenylation is unlikely to have the same biological significance in all cases. Indeed, all RIBO and EXO sequences were on the plus strand, suggesting that these sRNAs arise from the degradation of abundant cellular RNAs, mostly rRNAs and tRNAs. These polyA+ sRNAs may represent degradation intermediates that are targeted to the exosome by the addition of short polyA tails, as shown in other eukaryotes (44,45). In contrast, polyA+ sRNAs homologous to the *ND7* gene cannot be degradation products of the *ND7* mRNA since they were all on the antisense strand (see below).

PolyA+ RIBO and EXO sRNAs were much less frequent in the 24-nt fraction of the 7h*ND7* sample (14% and 4%, respectively) and virtually inexistent in higher fractions. This could reflect either the preferential polyadenylation or the slower degradation of shorter molecules. PolyA+ sRNAs were also found among the sequences cloned from the 2hWT sample, albeit at a lower frequency (only 9% of RIBO in the 23-nt fraction, very few elsewhere) (Supplementary Table 1). The higher frequency seen in the 7h*ND7* sample suggests that dsRNA feeding does increase the steady-state amount of polyA+ sRNAs, possibly through the saturation of some degradation pathway. Of note, the fact that the polyadenylation of EXO sRNAs is similar to that of fragments of *Paramecium* rRNAs and tRNAs suggests that single-stranded RNAs from food bacteria or from the wheat-grass infusion medium (the most likely source of RNAs from plants, fungi, or bacteria other than *E. coli*) can pass from phagosomes to cytoplasm, like the dsRNAs used to induce RNAi.

Analysis of ND7-silencing siRNAs reveals two subclasses

Before the re-mapping of polyA+ sRNAs, only eight of the sRNAs randomly sequenced from the 7h*ND7* sample were found to match the *ND7* dsRNA used for silencing with $\leq$1 mismatch (Supplementary Table 3A). As expected for primary siRNAs cleaved by Dcr1 from the dsRNA, these sRNAs were equally distributed on both strands. Half of them were 23 nt in length, and the other half (24–25 nt) ended with As and/or mismatched nucleotides. A larger number (23) of sRNAs matched the vector

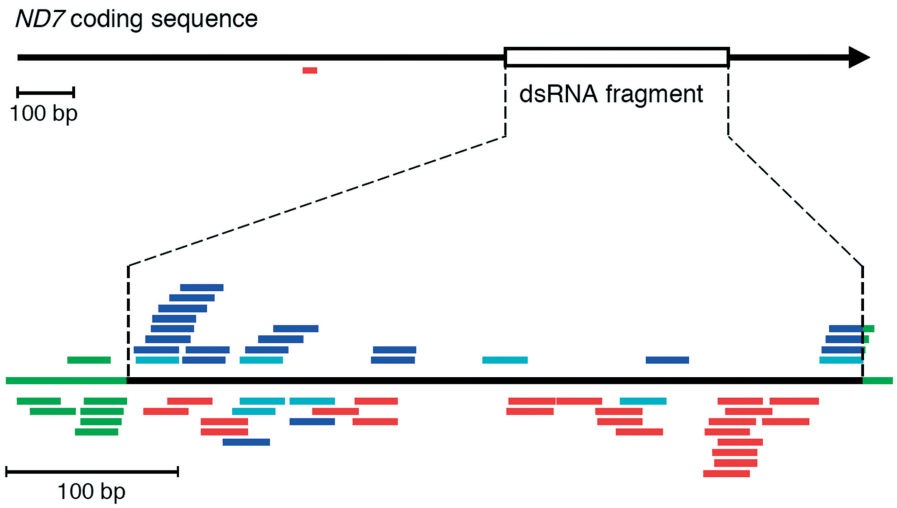

Figure 4. Distribution of *ND7* siRNAs. The black arrow represents the full length of the *ND7* coding sequence; the open box indicates the position of the 397-bp fragment cloned into the feeding vector for dsRNA production. The bottom line shows the full length of the dsRNA molecule produced in *E. coli*; green extensions on both sides of the *ND7* fragment are vector sequences located between the convergent T7 promoters. Sense and antisense siRNAs are aligned above and below the line, respectively. Primary siRNAs obtained by random sequencing of the 23-nt fraction are shown in green and light blue; those obtained by hybridization-selection from the 25-nt fraction are shown in dark blue. PolyA+ siRNAs obtained by random sequencing or by selective reamplification of polyA+ molecules are shown in red; one was found to map in the *ND7* coding sequence upstream of the dsRNA fragment (top).

part of the *ND7* feeding plasmid on both strands. All but two were exactly 23-nt long, although most were found in the 24-nt fraction (Supplementary Table 3B). To determine whether sRNAs related to dsRNA feeding could also be found in the 25-nt fraction, where their presence would likely be masked by the vast excess of scnRNAs, we used a hybridization-selection method to enrich that RT-PCR library in sequences homologous to the entire *ND7* gene, prior to concatemerization and cloning. This procedure yielded 20 non-redundant *ND7* sRNAs from the 25-nt fraction (Supplementary Table 3C). All of them mapped to the *ND7* dsRNA fragment and three overlapped the junction between *ND7* and vector sequences, indicating that they were silencing-associated siRNAs rather than genomic scnRNAs. All but two were 23 or 24 nt in length (11 and 7, respectively); thus, selecting for slow-migrating *ND7* sRNAs did not yield significantly longer molecules, but instead selected for sense-strand siRNAs (18 of 20), which migrate more slowly because of the high purine bias of the sense strand (64% purines over the *ND7* dsRNA fragment). Indeed, two 23-nt RNA oligonucleotides with 78% and 26% purines migrated with an apparent size difference of ~2 nt, the purine-rich one being the slower (data not shown). Thus, primary siRNAs cleaved from the dsRNA are predominantly 23 nt in length.

After removal of 3′-terminal A stretches (2–9 nt), the re-mapping of unidentified sequences obtained by random sequencing of the 7h*ND7* sample yielded 15 additional sRNAs homologous to the *ND7* dsRNA, none of which contained any vector-derived sequence (Supplementary Table 3D). Two additional features appeared to distinguish these polyA+ sRNAs from the primary *ND7* siRNAs described above. First, they showed a strong preference for a 5′U (11 of 15 sRNAs, versus only 6 of 28 primary siRNAs). Secondly and quite strikingly, they were all antisense to the *ND7* gene (Figure 4 and Supplementary Figure S4). Thus polyA+ *ND7* sRNAs cannot be mRNA degradation products, which also stands in contrast to polyA+ RIBO and EXO sRNAs. Although most polyA+ *ND7* sRNAs had total lengths of 24 or 25 nt, 13 were from the 23-nt fraction and 2 from the 24-nt fraction, indicating that they migrated faster than would be expected from their size. Indeed, the average size of those cloned from the 23-nt fraction was 24.5 nt, compared to 23.3 nt for primary siRNAs isolated from the same fraction (Supplementary Table 4). The lack of any untemplated polyA tail among the 20 *ND7* siRNAs cloned by hybridization-selection from the 25-nt fraction further suggests that polyA+ *ND7* sRNAs are rare in higher fractions.

To confirm this point, polyA+ sRNAs were specifically reamplified from the 23 and 26-nt RT-PCR libraries using a downstream primer extended at its 3′ end by 5 or 10 Ts. This yielded 13 *ND7* sRNAs out of 94 sequences from the 23-nt fraction (14%, Supplementary Table 3E), but none among 71 sequences from the 26-nt fraction. Thus, polyA+ *ND7* sRNAs indeed appear to be almost entirely confined to the 23-nt fraction. However, because the 3′ polyT extension of the downstream primer could possibly pair with A-rich genomic sequences, some of the sRNAs reamplified in this way may not be truly polyadenylated. This was likely the case for two of the *ND7* sRNAs that were on the sense strand of the *ND7* dsRNA, since 4 of 5 and 8 of 10 nucleotides facing the polyT extensions were natural As in the *ND7* sequence. To circumvent this specificity problem, only those sequences with untemplated polyA tails longer than the primer's polyT extension were considered to be unambiguously polyA+. Of the 13 sequences, eight satisfied this criterion; like the 15

Table 3. Distributions of all 25-nt segments of the MAC genome beginning with 5′-TNG, and of 5′-UNG, 25-nt MAC scnRNAs, over annotated features of the MAC genome

Features	5′TNG in MAC genome		5′UNG MAC scnRNAs		Statistical test	
	Numbers	Frequencies	Observed	Expected (e)	χ^2	*P*-value
Intergenic regions[a]	1265012	0.169	31	29.3	0.06	0.80
Coding exons,[b] sense str.	2922337	0.391	71	67.6	0.21	0.65
Introns,[c] sense str.	246029	0.033	5	5.7	0.01	0.94
Coding exons,[b] antisense str.	2859122	0.382	61	66.1	0.52	0.47
Introns,[c] antisense str.	181929	0.024	4	4.2	0.02	0.89
Other[d]	6469	0.001	1	0.1	0.82	0.36
Total	7480898	1.000	173	173.0		

[a]Many scnRNAs mapping to intergenic regions may be part of non-coding exons, because intergenic regions are very short (352 bp on average) and most 5′ and 3′UTRs have not been annotated.
[b]scnRNAs mapping in coding exons, including those overlapping 5′ or 3′UTRs, but excluding those overlapping introns.
[c]scnRNAs overlapping intronic sequences. Because introns are very short (25 nt on average), very few scnRNAs are expected to be entirely within introns.
[d]All other cases. The one case observed overlaps the coding sequences of two closely spaced, convergent genes.
[e]Random expectation is based on the actual distribution of 25 nt, 5′-TNG segments of the MAC genome. The non-significant *P*-values in the last column (Pearson's χ^2 test with Yates' continuity correction) indicate that the fractions of scnRNAs mapping to the different types of annotated features are consistent with a random distribution (confirmed by a χ^2 test comparing observed and expected distributions, $P = 0.35$).

polyA+ *ND7* sRNAs obtained by bulk sequencing, these were all antisense (Figure 4 and Supplementary Figure S4) and showed a 5′U bias (seven of eight). Interestingly, one of the 13 sRNAs was found to map ~350 bp upstream of the dsRNA segment in the *ND7* coding sequence, in the antisense orientation (Figure 4). Although this observation would need to be confirmed by deeper sequencing of siRNAs to obtain more examples, it raises the possibility that dsRNA feeding results in low levels of transitive siRNAs outside the targeted region of the mRNA sequence, as observed in *C. elegans* (3,4).

A fraction of the sRNAs from 23-nt fractions that matched the MAC genome were also polyadenylated (20% in 2hWT and 65% in 7h*ND7*; Supplementary Table 1). The absence of a clear 5′UNG signature (see below) suggests that many of these are not meiosis-specific scnRNAs and may represent endogenous siRNAs. Despite the small number of such sequences, nine sRNAs were found to cluster within 200 bp in a 1.1 kb intergenic region of scaffold 22, between convergent genes (Supplementary Figure S5). At least five of these sRNAs carried a polyA tail, and seven started with a 5′U. Furthermore, all were on the same strand, suggesting they could be endogenous siRNAs produced by the same mechanism as polyA+ *ND7* sRNAs. Northern blot analyses with a probe specific for this region revealed that homologous ~23–24-nt sRNAs are present in vegetative cells as well as during sexual events, confirming that they are not meiosis-specific scnRNAs (not shown). The clustering and absolute strand bias of sequenced molecules are also reminiscent of the genomic distribution of endogenous 23–24-nt endogenous sRNAs in *T. thermophila* (22).

25-nt scnRNAs with a 5′-UNG signature cover all types of micronuclear sequences

As expected from the abundance of scnRNAs revealed by 5′-end labelling, the fraction of sRNAs matching the MAC genome was much higher in the 25-nt fractions than in other fractions (Table 2). Analysis of 160 sequences from 2hWT and 7h*ND7* showed the following characteristics: (i) a predominant length of 25 nt (77%); (ii) a very high frequency of 5′U (96%); (iii) a very high frequency of G at the third position (73%). Sequences with no identified match were also very abundant (119 sRNAs) in these fractions and shared the same characteristics: (i) 72%; (ii) 92%; (iii) 64%. While these slightly lower figures are consistent with a small contamination by EXO sequences from unknown genomes, the vast majority of these sequences likely represent scnRNAs from MIC-specific regions of the *P. tetraurelia* genome. Indeed, they show the same nucleotide composition as MAC scnRNAs (71% A+T), and two sequences were found to match the MIC-specific class-II transposon Sardine.

Each of the MAC or putative MIC-specific sRNAs from the 25-nt fractions was cloned only once, suggesting high complexity. In contrast with putative endogenous siRNAs, for which one local cluster was detected with fewer sequences, their mapping suggests an even distribution all along the entire set of MAC chromosomes on both strands. To study the distribution of MAC scnRNAs over annotated features in more detail, we constituted a set of 173 sequences of exactly 25 nt starting with 5′UNG by pooling all those obtained from bulk sequencing of size-selected fractions together with additional sequences obtained during preliminary tests of the hybridization-selection method (which yielded no enrichment). When compared to the distribution of all 25-nt sequences starting with 5′TNG in the MAC genome, the distribution of these 173 scnRNAs was not statistically different from random expectation, with the same proportions matching intergenic regions and the sense or antisense strands of coding exons and introns (Table 3). A larger set of 641 MAC scnRNAs (not necessarily 5′UNG, see below) contained very similar proportions of sequences overlapping

introns (18 on the sense strand and 14 on the antisense strand, Supplementary Table 5). All sRNAs sequenced in this study (including the 344 25-nt, putative MIC-specific scnRNAs, Supplementary Table 6) were also compared to the entire set of predicted mRNAs, but none of the putative scnRNAs was found to match an exon-exon junction. We conclude that scnRNAs are processed from unspliced precursors.

Because of the high complexity of the scnRNA population, identifying one scnRNA matching one of the few known IESs by random sequencing would require a very large number of sequences. To examine the distribution of scnRNAs over known MIC loci, we therefore used the MIC versions of the *ND7* and G^{51} genes, containing several IESs each, to enrich RT-PCR products from the 25-nt fraction of the 2hWT sample by hybridization-selection before cloning. Although the enriched libraries were not sequenced to saturation, the diversity of homologous scnRNAs obtained in this manner confirmed that scnRNAs densely cover all types of micronuclear sequences on both strands: intergenic regions, exons, introns, and IESs (Supplementary Figure S6).

The complementary signature implies scnRNA duplexes with 2-nt 3′ overhangs

Because Dicer-like proteins appear to be involved in their biogenesis, scnRNAs may be expected to be produced from longer dsRNAs as 25-nt duplexes with 2-nt 3′ overhangs. The 5′UNG signature of one strand should therefore be mirrored by the complementary CNA at positions 21–23 of the other strand. To look for this complementary signature, a larger data set of 641 sequences was constituted by selecting the sRNAs that were exactly 25 nt in length from all random sequencing or hybridization-selection experiments (after elimination of redundant sequences), keeping only those that matched the MAC genome to make the set more reliable (Supplementary Table 5). The compositional profile of the whole set showed no sign of a CNA signature at positions 21–23 (Figure 5A). However, that signature was obvious in the small subset (6%) of sRNAs that did not start with a 5′U (Figure 5B). That subset also showed an elevated frequency of C at position 22, which was mirrored by an elevated frequency of G at position two among sRNAs starting with a 5′U. This observation provides the first direct evidence that scnRNA duplexes are indeed cleaved from dsRNA with a typical RNaseIII geometry, i.e. 2-nt 3′ overhangs, at both ends. It further implies that the 5′UNG strands are preferentially stabilized, possibly through preferential loading into the two Piwi proteins involved in programming genome rearrangements (K. Bouhouche and E. Meyer, unpublished). To reflect this asymmetry, the more abundant 5′UNG sequences were tentatively called ‘guide’ strands, and the less frequent CNA sequences ‘passenger’ strands.

Preferential stabilization of one scnRNA strand does not appear to follow the rule shown to determine strand selection from siRNA and miRNA duplexes in metazoans, the preferential loading of the strand whose 5′ end has the lower pairing stability (46). Indeed, the pairing stability of

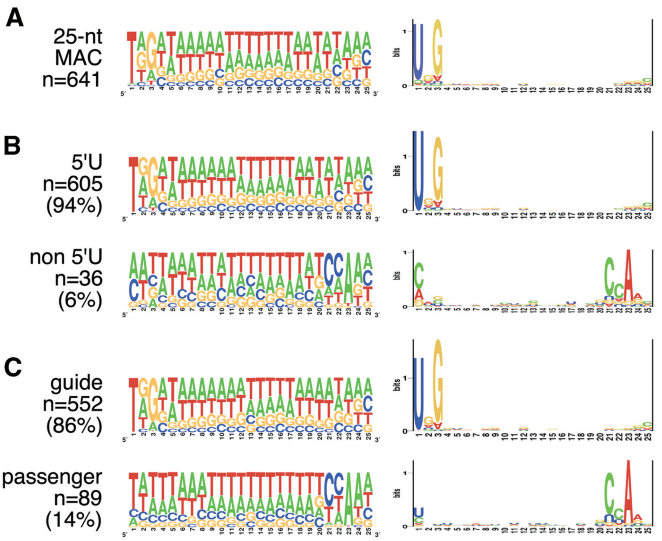

Figure 5. Compositional profile of MAC 25-nt scnRNAs. (**A**) The whole set of 641 non-redundant sequences. In the logo on the left, the height of each nucleotide symbol is proportional to its frequency at each position. In the logo on the right, the height of each symbol represents the relative entropy between the observed frequency of the nucleotide at that position and its *a priori* probability (0.36 for A and U, 0.14 for C and G). (**B**) The same representations are provided for two subsets of the 641 sequences: those beginning with a 5′U and those beginning with another nucleotide. (**C**) The 641 sequences are here classified into ‘guide’ or ‘passenger’ subsets by the automatic procedure described in the text.

the ends of perfectly matched scnRNA duplexes should depend only on base composition, so that one would expect the first 4 bases (positions 1–4) of most scnRNAs to have a lower G + C content than the last 4 bases of the duplex portion (positions 20–23). However, the opposite trend is observed in the sample of 641 MAC scnRNAs: on average, positions 1–4 have 0.28 more G/C bases than positions 20–23. This is likely due to the high frequency of G at position 3, which, given the 72% A + T content of the MAC genome, is more significant than the high frequency of U at position 1: positions 1–4 have 0.31 more G/C bases than positions 20–23 among scnRNAs starting with a 5′U, but 0.25 less G/C bases among scnRNAs not starting with a 5′U. Preferential loading of the 5′UNG strands would be better explained by base-specific contacts of the 5′ end-binding pocket of the relevant Piwi proteins (or of loading factors), as postulated to account for strong biases in the 5′ nucleotides of sRNAs associated with several Argonaute proteins in *Arabidopsis thaliana* (47–49).

A 5′U may not be the sole determinant of strand selection, since some non-CNA sequences had a G at the third position but no U at the 5′end; other sequences had neither the UNG nor the CNA signatures, while some had both. We therefore attempted to classify all sequences in a more objective manner. A position-specific weighting matrix was computed for the first 23 nt (the double-stranded portion of duplexes) of the whole set of 641 sRNAs, which contains an excess of guide strands. The matrix was then used to score each sequence and its reverse complement, thus allowing a preliminary

classification as guide or passenger. A new matrix was then computed from all sequences after reversing and complementing the passengers, and the process was repeated until the classification was stable (four rounds). This procedure increased the fraction of putative passenger strands to 14% (Figure 5C). The final matrix was also used to score a non-redundant set of 344 25-nt, putative or confirmed MIC scnRNAs, yielding a similar fraction of passenger strands (16%) (Supplementary Table 6).

DISCUSSION

Primary and secondary siRNAs associated with dsRNA-induced RNAi

Homology-dependent gene silencing in *P. tetraurelia* was previously shown to correlate with accumulation of ~23-nt siRNAs during both vegetative growth and sexual events (18,19,39). The results presented here indicate that *DCR1*, the only gene likely to encode a catalytically active Dicer, is involved in their production. The closest homolog in public databases is *T. thermophila*'s *DCR2*, which has been implicated in the biosynthesis of endogenous 23–24-nt sRNAs throughout the life cycle (20,22). The siRNA pathway thus appears to be conserved in these distantly related oligohymenophorean ciliates. The conservation of catalytically inactive Dicer paralogs in both species (*DCR2* and *DCR3* in *Paramecium*, and *DCR1* in *Tetrahymena*) is intriguing, but their functions, if any, remain to be determined.

Random sequencing of the 23 and 24-nt fractions of the 7h*ND7* sample yielded 45 siRNAs associated with *ND7* silencing and 64 sRNAs matching other sequences of the MAC genome, suggesting that the dsRNA feeding technique results in siRNA levels much higher than those of putative endogenous siRNAs produced by any locus. Unexpectedly, the analysis of dsRNA-induced siRNAs revealed the existence of two distinct subclasses. In this silencing technique, cells are fed with an RNaseIII-deficient *E. coli* strain engineered to produce dsRNA homologous to a segment of the target gene, which is cloned between convergent T7 promoters in the L4440 plasmid (17). To be processed by the *Paramecium* RNAi machinery, the dsRNA must escape the phagosomes during the digestion of bacteria and presumably enter the cytoplasm. One subset of sequenced sRNAs, predominantly 23 nt in length, appears to correspond to primary siRNAs that are directly produced by cleavage of the inducing dsRNA. Indeed it contained sequences matching both strands of the *ND7* gene segment as well as the L4440 vector, and included sequences that overlap the *ND7*-vector junction. These primary siRNAs did not show any particular nucleotide bias at the 5′ end, suggesting that Dcr1 has no sequence specificity, and did not contain any untemplated polyA tail.

A second subset of *ND7* sRNAs, clearly associated with dsRNA-induced silencing since they clustered in the targeted segment, was characterized by the presence of short, untemplated polyA tails (2–9 nt). Although similar polyA tails were also found in RIBO and EXO sRNAs from the 7h*ND7* sample, the latter were all on the plus strand of abundant RNAs such as rRNAs and tRNAs and therefore appear to be degradation intermediates, possibly tagged by a variant polyA-polymerase of the Cid14/Trf4/Trf5 family for digestion by the exosome (44,45). Because polyA+ RIBO and EXO sRNAs were much less frequent (but not completely absent) in the 2hWT sample, it is tempting to speculate that the dsRNA feeding technique is responsible for the stabilization of polyA+ intermediates in the 7h*ND7* sample, perhaps through saturation of the exosome degradation pathway. PolyA+ *ND7* sRNAs, on the other hand, cannot be mRNA breakdown products since they were all antisense to the gene. The fact that they were almost entirely confined to the 23-nt fraction also makes it unlikely that they could be random degradation intermediates from longer antisense transcripts. Indeed, none of the *ND7* sequences cloned by hybridization-selection from the 25-nt fraction carried a poly A tail, and the specific reamplification of polyA+ molecules from the 26-nt fraction failed to yield any *ND7* sRNA.

We cannot formally exclude that polyA+ *ND7* sRNAs are produced by partial degradation of primary siRNAs from their 3′ end, followed by polyadenylation. However, their strong preference for a 5′U is not observed among primary siRNAs. Furthermore, unlike the latter they do not include any vector-derived sequence and have a strict antisense polarity. Thus, if they were processed from the symmetrical dsRNA, the polyadenylation mechanism would have to be sensitive to the 5′ nucleotide and to discriminate between endogenous and exogenous sequences, as well as between sense and antisense strands of *ND7* sequences. An alternative possibility is suggested by the finding that, in *C. elegans*, induction of RNAi by dsRNA feeding (3) or by expression of a single primary siRNA (4) results in the production of secondary siRNAs that are synthesized by an RdRP from the targeted mRNA. In this organism, synthesis of secondary siRNAs was shown to be Dicer-independent and to rely on unprimed RdRP activity producing 5′-triphosphate sRNAs of variable lengths with a strong 5′G bias (2). A similar unprimed polymerization mode of the QDE-1 RdRP from *Neurospora crassa* can produce 9–21-nt sRNAs scattered along the length of a ssRNA template *in vitro* (50). Such a mechanism would perfectly explain the antisense polarity of polyA+ *ND7* sRNAs, and the lack of any vector-derived sequence among them. It could also account for the antisense sRNA found to map outside the feeding segment in the *ND7* coding sequence, which is consistent with the occurrence of transitive RNAi in *P. tetraurelia*. Thus we favour the hypothesis that polyA+ *ND7* sRNAs constitute a distinct subclass of secondary siRNAs. As observed in the nematode, these would be more abundant in the gene segment used for dsRNA feeding than in upstream or downstream regions (3).

An RdRP-dependent RNAi mechanism in *P. tetraurelia* is independently suggested by the unexpected observation that the two putatively active RdRP genes in the genome are required for efficient silencing by dsRNA feeding (A. Le Mouël and EM, unpublished; S. Marker and M. Simon, personal communication). The closest

homolog of these proteins, *T. thermophila*'s Rdr1, is thought to be involved in the synthesis of endogenous 23–24-nt siRNAs through unprimed copying of single-stranded transcripts (22). In the proposed model, however, the 3′ ends of siRNAs are cleaved from the resulting dsRNAs by Dcr2. The absolute strand bias of siRNA clusters is assumed to result from a processing polarity imposed by Rdr1 through its physical interaction with Dcr2; Rdr1 would also serve to provide a 5′-triphosphate end which appears to be required for Dcr2 to cleave dsRNA into precisely sized 23–24-nt siRNAs. In *P. tetraurelia*, Dcr1-mediated cleavage of the 3′ end of secondary siRNAs would imply 3′-end trimming before polyadenylation; Dicer-independent biogenesis, as observed in the nematode, is a simpler model for both secondary siRNAs associated with dsRNA feeding and the similar strand-biased cluster of putative endogenous siRNAs. Whatever the case may be, secondary siRNAs should have 5′-triphosphate ends defined by RdRP polymerization starts. This would fit the observation that they migrate faster than expected from their sizes, since additional phosphates result in faster migration in the type of gel used; the 5′U bias would then reflect some RdRP specificity.

The biological significance of secondary siRNA polyadenylation remains unclear. Further work is necessary to determine whether polyA+ siRNAs represent degradation intermediates stabilized by the saturation of the exosome or other nucleases, as may be the case for RIBO and EXO sRNAs, or whether they associate with any of the 14 *Paramecium* Piwi-like proteins and have silencing activity. Small RNA polyadenylation is reminiscent of the polyuridylation of *A. thaliana* miRNAs and siRNAs that is observed in the absence of Hen1-mediated methylation of the last ribose (51). Secondary siRNA polyadenylation might result from saturation of the *P. tetraurelia* Hen1 homolog, due to dsRNA feeding and/or to intense RNA metabolism during conjugation. However, primary siRNAs and scnRNAs do not appear to be significantly polyadenylated in the same sample. One possible explanation for this specificity is suggested by the finding that *T. thermophila*'s Rdr1 associates in a complex similar to the RDRC of *Schizosaccharomyces pombe* with Cid12-like polyA polymerases that have close homologs in the *P. tetraurelia* genome (22,52). The exact function of Cid12-like proteins in other systems is currently unclear; the *P. tetraurelia* homologs might be responsible for the specific polyadenylation of RdRP-dependent siRNAs.

Biogenesis and function of meiosis-specific scnRNAs

Our analysis of small RNAs during conjugation has shown that a highly complex population of ~25-nt scnRNAs is produced during early meiosis in *P. tetraurelia*. Their biogenesis appears to involve the developmentally regulated Dcl2 and Dcl3 proteins, which contain two apparently functional RNaseIII domains but are more similar to metazoan Drosha than to prototypical Dicers in that they lack an RNA helicase domain. These proteins are also structurally similar to *T. thermophila*'s Dcl1 (31,33), indicating that the scnRNA pathway is at least partially conserved in these two ciliates. Our preliminary screen did not suggest any possible function of the *DCL1* and *DCL4* genes during sexual reproduction. The *Paramecium DCL* gene family is an interesting case of functional diversification after whole-genome duplications (43): while only *DCL2* and *DCL3* are up-regulated during meiosis, *DCL1* is a close paralog of *DCL2* dating from the last duplication, and the *DCL3*/*DCL4* pair dates from the previous duplication.

We have previously shown that microinjection of synthetic RNA duplexes mimicking the inferred structure of *Paramecium* scnRNAs can target the elimination of homologous sequences in the developing MAC (39), as had been proposed in *T. thermophila* to explain the effects of knocking out several genes in the scnRNA pathway (31–33,40). Silencing both the *DCL2* and *DCL3* genes during sexual reproduction was here shown to result in unviable progeny. Cell death only occurred after cytologically normal development of the zygotic MAC and is likely due to genome rearrangement defects, since unexcised copies of IESs were shown to accumulate during development. This further supports the essential role of the scnRNA pathway in programming rearrangements. *DCL2* and *DCL3* may have redundant functions, since knocking down either one alone does not compromise viability but results in the same subtle phenotypes. These phenotypes are also similar to those observed after partial depletion of the meiosis-specific RNA-binding proteins Nowa1/2, confirming previous speculation that the latter are involved in the scnRNA pathway (19).

The sequencing of several hundreds of 25-nt scnRNAs offers further insight into their biogenesis. The vast majority of them present a 5′-UNG signature, while a small minority shows the complementary CNA motif at positions 21–23. This observation provides independent evidence that scnRNAs are cleaved from dsRNA precursors as duplexes with 2-nt 3′ overhangs at both ends, as would be expected for typical RNaseIII cleavage products. It also indicates that the CNA strand is rapidly eliminated, or that it is specifically modified in a way that interferes with its cloning by the method used. Preferential stabilization of 5′UNG strands ('guide' strands) *in vivo* could result from base-specific interactions with the two Piwi-like proteins implicated in genome rearrangements (K. Bouhouche and EM, unpublished results), or with specialized loading factors. If non-5′U scnRNAs represent the steady state of unstable passenger strands, the fact that they frequently show the CNA motif, rather than totally random sequences, suggests that the 5′UNG signature is due to some sequence specificity of one of the two Dicer-like cuts.

The identification of a few scnRNAs matching known MIC-specific sequences, as well as the localization of Dcl2 in the MIC during meiosis I prophase, indicates that scnRNAs are produced from the MIC. Although the detection of local clusters or barren regions would require deep sequencing, the distribution of scnRNAs over the known MAC genome appears to be homogeneous and random with respect to annotation, suggesting they may be produced from the whole germline genome. The fraction of putative MIC-specific scnRNAs was found to be higher in 7h*ND7* than in 2hWT, which is consistent with a

progressive enrichment during early conjugation. A firm conclusion, however, must await the complete sequencing of the MIC genome and the detailed analysis of different time points from the same synchronous conjugation. Ciliate scnRNAs resemble metazoan piRNAs in their developmental regulation and germline restriction, and at least partially in their function, since they are involved in epigenetic control of transposons and IESs, which are thought to have evolved from transposon insertions (23,25). However, scnRNAs are clearly produced by Dicer-like cleavage of dsRNA precursors, and no evidence was found for a 'ping-pong' mechanism of Dicer-independent amplification, as proposed in the fly (53,54) and in vertebrates (12,55) piRNAs, or for upstream motifs suggestive of individual transcription promoters, as observed in the nematode (56). While the similarities suggest that meiosis- and/or germline-specific small RNAs may be very ancient in eukaryotes, the present results highlight important differences in biogenesis mechanisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Mireille Bétermier for the gift of *ND7* micronuclear sequences, Laurent Duret for help with bioinformatic analyses, and Linda Sperling, Mireille Bétermier and all lab members for comments on the manuscript.

FUNDING

G.L., M.N. and V.S. are recipients of fellowships from the Ministère de l'Education Nationale and from the Association pour la Recherche sur le Cancer, and VS from the Ligue Nationale contre le Cancer and from the Fondation de la Recherche Médicale. Research was funded by the CNRS, the Ministère de l'Education Nationale, de la Recherche et de la Technologie (ACI BCMS 2004 #BCMS287), the Association pour la Recherche sur le Cancer (grant #3608), and contract NT05-2_41522 from the Agence Nationale de la Recherche. Funding for open access charge: Agence Nationale de la Recherche.

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Aoki,K., Moriguchi,H., Yoshioka,T., Okawa,K. and Tabara,H. (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *EMBO J.*, **26**, 5007–5019.
3. Pak,J. and Fire,A. (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science*, **315**, 241–244.
4. Sijen,T., Steiner,F.A., Thijssen,K.L. and Plasterk,R.H. (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, **315**, 244–247.
5. Matzke,M.A. and Birchler,J.A. (2005) RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet.*, **6**, 24–35.
6. Batista,P.J., Ruby,J.G., Claycomb,J.M., Chiang,R., Fahlgren,N., Kasschau,K.D., Chaves,D.A., Gu,W., Vasale,J.J., Duan,S. *et al.* (2008) PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol. Cell*, **31**, 67–78.
7. Das,P.P., Bagijn,M.P., Goldstein,L.D., Woolford,J.R., Lehrbach,N.J., Sapetschnig,A., Buhecha,H.R., Gilchrist,M.J., Howe,K.L., Stark,R. *et al.* (2008) Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol. Cell*, **31**, 79–90.
8. Hartig,J.V., Tomari,Y. and Forstemann,K. (2007) piRNAs–the ancient hunters of genome invaders. *Genes Dev.*, **21**, 1707–1713.
9. Klattenhoff,C. and Theurkauf,W. (2008) Biogenesis and germline functions of piRNAs. *Development*, **135**, 3–9.
10. O'Donnell,K.A. and Boeke,J.D. (2007) Mighty Piwis defend the germline against genome intruders. *Cell*, **129**, 37–44.
11. Aravin,A.A. and Bourc'his,D. (2008) Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev.*, **22**, 970–975.
12. Aravin,A.A., Sachidanandam,R., Girard,A., Fejes-Toth,K. and Hannon,G.J. (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, **316**, 744–747.
13. Carmell,M.A., Girard,A., van de Kant,H.J., Bourc'his,D., Bestor,T.H., de Rooij,D.G. and Hannon,G.J. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell*, **12**, 503–514.
14. Klenov,M.S., Lavrov,S.A., Stolyarenko,A.D., Ryazansky,S.S., Aravin,A.A., Tuschl,T. and Gvozdev,V.A. (2007) Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. *Nucleic Acids Res.*, **35**, 5430–5438.
15. Galvani,A. and Sperling,L. (2001) Transgene-mediated post-transcriptional gene silencing is inhibited by 3′ non-coding sequences in *Paramecium*. *Nucleic Acids Res.*, **29**, 4387–4394.
16. Ruiz,F., Vayssie,L., Klotz,C., Sperling,L. and Madeddu,L. (1998) Homology-dependent gene silencing in *Paramecium*. *Mol. Biol. Cell*, **9**, 931–943.
17. Galvani,A. and Sperling,L. (2002) RNA interference by feeding in *Paramecium*. *Trends Genet.*, **18**, 11–12.
18. Garnier,O., Serrano,V., Duharcourt,S. and Meyer,E. (2004) RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia. *Mol Cell Biol*, **24**, 7370–7379.
19. Nowacki,M., Zagorski-Ostoja,W. and Meyer,E. (2005) Nowa1p and Nowa2p: Novel Putative RNA Binding Proteins Involved in trans-Nuclear Crosstalk in Paramecium tetraurelia. *Curr Biol.*, **15**, 1616–1628.
20. Lee,S.R. and Collins,K. (2006) Two classes of endogenous small RNAs in Tetrahymena thermophila. *Genes Dev.*, **20**, 28–33.
21. Howard-Till,R.A. and Yao,M.C. (2006) Induction of gene silencing by hairpin RNA expression in Tetrahymena thermophila reveals a second small RNA pathway. *Mol. Cell Biol.*, **26**, 8731–8742.
22. Lee,S.R. and Collins,K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat. Struct. Mol. Biol.*, **14**, 604–610.
23. Bétermier,M. (2004) Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.*, **155**, 399–408.
24. Jahn,C.L. and Klobutcher,L.A. (2002) Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.*, **56**, 489–520.
25. Klobutcher,L.A. and Herrick,G. (1997) Developmental genome reorganization in ciliated protozoa: the transposon link. *Progr. Nucleic Acid Res. Mol. Biol.*, **56**, 1–62.
26. Yao,M.C., Duharcourt,S. and Chalker,D.L. (2002) In Craig,N., Craigie,R., Gellert,M. and Lambowitz,A. (eds), *Mobile DNA II*. Academic Press, New York, pp. 730–758.
27. Juranek,S.A., Rupprecht,S., Postberg,J. and Lipps,H.J. (2005) snRNA and heterochromatin formation are involved in DNA excision during macronuclear development in stichotrichous ciliates. *Eukaryot. Cell*, **4**, 1934–1941.
28. Liu,Y., Mochizuki,K. and Gorovsky,M.A. (2004) Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. *Proc. Natl Acad. Sci. USA*, **101**, 1679–1684. Epub 2004 Jan 1630.
29. Liu,Y., Taverna,S.D., Muratore,T.L., Shabanowitz,J., Hunt,D.F. and Allis,C.D. (2007) RNAi-dependent H3K27 methylation is

required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes Dev.*, **21**, 1530–1545.
30. Taverna,S.D., Coyne,R.S. and Allis,C.D. (2002) Methylation of histone H3 at lysine 9 targets programmed DNA elimination in Tetrahymena. *Cell*, **110**, 701–711.
31. Malone,C.D., Anderson,A.M., Motl,J.A., Rexer,C.H. and Chalker,D.L. (2005) Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. *Mol. Cell Biol.*, **25**, 9151–9164.
32. Mochizuki,K., Fine,N.A., Fujisawa,T. and Gorovsky,M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell*, **110**, 689–699.
33. Mochizuki,K. and Gorovsky,M.A. (2005) A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev.*, **19**, 77–89.
34. Meyer,E. and Chalker,D.L. (2007) In Allis,C.D., Jenuwein,T., Reinberg,D. and Caparros,M.C. (eds), *Epigenetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 127–150.
35. Yao,M.C. and Chao,J.L. (2005) RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. *Annu. Rev. Genet.*, **39**, 537–559.
36. Duharcourt,S., Keller,A.M. and Meyer,E. (1998) Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell Biol.*, **18**, 7075–7085.
37. Mochizuki,K. and Gorovsky,M.A. (2004) Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.*, **18**, 2068–2073.
38. Mochizuki,K. and Gorovsky,M.A. (2004) Small RNAs in genome rearrangement in Tetrahymena. *Curr. Opin. Genet. Dev.*, **14**, 181–187.
39. Lepere,G., Betermier,M., Meyer,E. and Duharcourt,S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. *Genes Dev.*, **22**, 1501–1512.
40. Aronica,L., Bednenko,J., Noto,T., Desouza,L.V., Siu,K.W., Loidl,J., Pearlman,R.E., Gorovsky,M.A. and Mochizuki,K. (2008) Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena. *Genes Dev.*, **22**, 2228–2241.
41. Chalker,D.L., Fuller,P. and Yao,M.C. (2005) Communication between parental and developing genomes during Tetrahymena nuclear differentiation is likely mediated by homologous RNAs. *Genetics*, **169**, 149–160.
42. Pfeffer,S., Lagos-Quintana,M. and Tuschl,T. (2003) In Ausubel,F., Kingston,R., Moore,D., Seidman,J., Smith,J. and Struhl,K. (eds), *Current protocols in molecular biology*. Wiley Interscience, New York, pp. 26.24.21–26.24.18.
43. Aury,J.M., Jaillon,O., Duret,L., Noel,B., Jubin,C., Porcel,B.M., Segurens,B., Daubin,V., Anthouard,V., Aiach,N. *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature*, **444**, 171–178.
44. Houseley,J., LaCava,J. and Tollervey,D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.*, **7**, 529–539.
45. Wang,S.W., Stevenson,A.L., Kearsey,S.E., Watt,S. and Bahler,J. (2008) Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. *Mol. Cell Biol.*, **28**, 656–665.
46. Schwarz,D.S., Hutvagner,G., Du,T., Xu,Z., Aronin,N. and Zamore,P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.
47. Mi,S., Cai,T., Hu,Y., Chen,Y., Hodges,E., Ni,F., Wu,L., Li,S., Zhou,H., Long,C. *et al.* (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. *Cell*, **133**, 116–127.
48. Montgomery,T.A., Howell,M.D., Cuperus,J.T., Li,D., Hansen,J.E., Alexander,A.L., Chapman,E.J., Fahlgren,N., Allen,E. and Carrington,J.C. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, **133**, 128–141.
49. Takeda,A., Iwasaki,S., Watanabe,T., Utsumi,M. and Watanabe,Y. (2008) The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol.*, **49**, 493–500.
50. Makeyev,E.V. and Bamford,D.H. (2002) Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell*, **10**, 1417–1427.
51. Li,J., Yang,Z., Yu,B., Liu,J. and Chen,X. (2005) Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. *Curr. Biol.*, **15**, 1501–1507.
52. Motamedi,M.R., Verdel,A., Colmenares,S.U., Gerber,S.A., Gygi,S.P. and Moazed,D. (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell*, **119**, 789–802.
53. Brennecke,J., Aravin,A.A., Stark,A., Dus,M., Kellis,M., Sachidanandam,R. and Hannon,G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, **128**, 1089–1103.
54. Gunawardane,L.S., Saito,K., Nishida,K.M., Miyoshi,K., Kawamura,Y., Nagami,T., Siomi,H. and Siomi,M.C. (2007) A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. *Science*, **315**, 1587–1590.
55. Houwing,S., Kamminga,L.M., Berezikov,E., Cronembold,D., Girard,A., van den Elst,H., Filippov,D.V., Blaser,H., Raz,E., Moens,C.B. *et al.* (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, **129**, 69–82.
56. Ruby,J.G., Jan,C., Player,C., Axtell,M.J., Lee,W., Nusbaum,C., Ge,H. and Bartel,D.P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell*, **127**, 1193–1207.