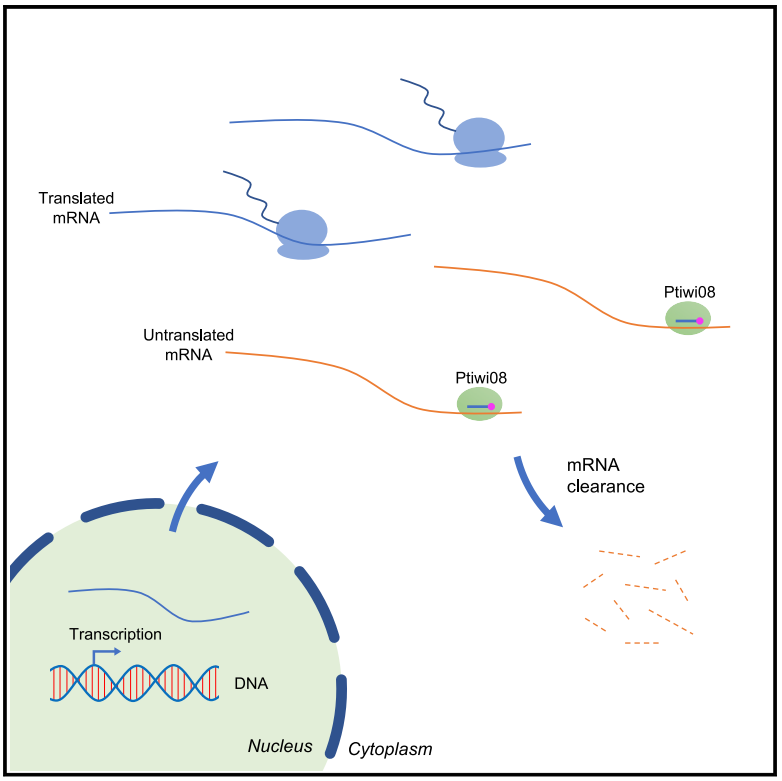
Cell Reports

Article

Developmental mRNA clearance by PIWI-bound endo-siRNAs in *Paramecium*

Graphical abstract



Authors

Therese Solberg, Victor Mason, Chundi Wang, Mariusz Nowacki

Correspondence

mariusz.nowacki@unibe.ch

In brief

Solberg et al. describe a small-RNA-mediated mRNA clearance pathway at the start of zygotic genome activation in the unicellular eukaryote *Paramecium tetraurelia*. In this pathway, the endo-siRNAs are strictly antisense to their target mRNAs, require Dicer1 for their production, and are bound by the PIWI-clade Argonaute protein Ptiwi08.

Highlights

- Discovery of a small-RNA-mediated mRNA clearance pathway in a unicellular eukaryote
- Endo-siRNAs display a strict strand bias and are antisense to their target mRNAs
- Endo-siRNAs are Dicer dependent and bound by a PIWI-clade Argonaute protein, Ptiwi08
- Untranslated mRNAs are likely targeted for degradation by the Ptiwi08 pathway

Solberg et al., 2023, Cell Reports *42*, 112213
March 28, 2023
https://doi.org/10.1016/j.celrep.2023.112213

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Developmental mRNA clearance by PIWI-bound endo-siRNAs in *Paramecium*

Therese Solberg,[1] Victor Mason,[1] Chundi Wang,[1,2] and Mariusz Nowacki[1,3,*]
[1]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland
[2]Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, China
[3]Lead contact
*Correspondence: mariusz.nowacki@unibe.ch
https://doi.org/10.1016/j.celrep.2023.112213

SUMMARY

The clearance of untranslated mRNAs by Argonaute proteins is essential for embryonic development in metazoans. However, it is currently unknown whether similar processes exist in unicellular eukaryotes. The ciliate *Paramecium tetraurelia* harbors a vast array of PIWI-clade Argonautes involved in various small RNA (sRNA) pathways, many of which have not yet been investigated. Here, we investigate the function of a PIWI protein, Ptiwi08, whose expression is limited to a narrow time window during development, concomitant with the start of zygotic transcription. We show that Ptiwi08 acts in an endogenous small interfering RNA (endo-siRNA) pathway involved in the clearance of untranslated mRNAs. These endo-siRNAs are found in clusters that are strictly antisense to their target mRNAs and are a subset of siRNA-producing clusters (SRCs). Furthermore, the endo-siRNAs are 2′-O-methylated by Hen1 and require Dcr1 for their biogenesis. Our findings suggest that sRNA-mediated developmental mRNA clearance extends beyond metazoans and may be a more widespread mechanism than previously anticipated.

INTRODUCTION

Argonaute proteins associate with small RNAs (sRNAs) to form complexes capable of sequence-specific targeting of complementary nucleic acids. These complexes have been shown to perform diverse roles ranging from translational repression and mRNA decay to heterochromatin formation guidance.[1,2] Recently, there has been much discussion in the field concerning the regulation of untranslated mRNAs by Argonaute proteins as well as the biogenesis of the sRNAs facilitating these processes.[3–6] The focus of most of these studies have been maternal mRNA clearance during the maternal-to-zygotic transition (MZT), when the zygotic genome gains transcriptional activity and maternally deposited mRNAs are degraded. A recent report links Argonaute 2 (Ago2) to this process in mouse early embryos and demonstrated that the degradation of over half of the maternal mRNAs is Ago2 dependent.[3] Moreover, the length of the 3′ UTR and the translational activity of the mRNA were shown to influence maternal mRNA clearance.[7] In *Caenorhabditis elegans*, antisense sRNAs bound to the Argonaute CSR1 are required for the removal of maternal mRNAs no longer engaged in translation.[4,5] In zebrafish, the microRNA miR-430 is required for clearance of maternal mRNAs and was recently shown to be required for heterochromatin establishment during the MZT.[6,8] Hence, there is increasing evidence that Argonaute proteins play a conserved role during early developmental stages in metazoans by ensuring the clearance of untranslated mRNAs.[4] While these reports are numerous, they have focused on multicellular organisms, and developmental mRNA clearance by such a pathway has not been described in any unicellular organism to date. This process is thought to be restricted to sexually reproducing multicellular eukaryotes carrying germline and somatic cells. However, single-celled eukaryotes such as ciliates also undergo sexual reproduction and separate the germline and soma in the form of two functionally distinct nuclei.

The ciliate *Paramecium tetraurelia* harbors a vast array of Argonaute proteins of the Piwi clade involved in various sRNA pathways, many of which have not yet been investigated. Although it has no proteins of the Argonaute clade, the large number of Ptiwi proteins (*P. tetraurelia* Piwi) fulfill both Argonaute and Piwi-like functions. The Ptiwi proteins investigated to date have been shown to perform diverse roles, including both exogenous and endogenous small interfering RNA (siRNA) pathways, as well as two sRNA pathways essential for DNA deletions during development as an extreme form of transposon silencing.[9–11] Hence, *Paramecium* is an excellent model for investigating the diversity of Argonaute proteins and sRNA pathways, yet the function of most Ptiwis still remains unknown. One of these is Ptiwi08, a protein that is expressed exclusively during development of a new somatic genome concomitant with the start of zygotic transcription.

Our analysis of Ptiwi08-bound endogenous (endo)-siRNAs revealed a strong antisense bias with complementarity to developmental mRNAs, which are negatively regulated by the expression of Ptiwi08. When mapped to the genome, these endo-siRNAs form clusters of high depth, which are a subset of the previously identified sRNA-producing clusters (SRCs) present during the *Paramecium* vegetative life stage.[12] While

Ptiwi08 is required for the stability of the siRNAs, neither Ptiwi08 nor its slicer activity is required for the antisense bias. Furthermore, we demonstrate that the biogenesis of the endo-siRNAs is Dicer1 (Dcr1) dependent and that they are 2′-O-methylated at their 3′ ends by methyltransferase Hen1, an interaction partner of Ptiwi08. We hypothesize that the entry into the Ptiwi08 pathway may be determined by translation potential and occurs post-transcriptionally in the cytoplasm in order to clear the cell of untranslated mRNAs during development. Our findings in this unicellular eukaryote are analogous to the maternal mRNA clearance during the MZT in metazoans, suggesting that the utilization of Argonaute proteins for the clearance of untranslated mRNAs during development may be a more widespread mechanism than previously anticipated and that its evolutionary origin may predate multicellularity.

RESULTS

Ptiwi08-bound sRNAs map to the genome in clusters

To understand the role of Ptiwi08 during development, Ptiwi08 was immunoprecipitated, and the co-precipitated sRNAs were extracted and sequenced. Cells were transformed with a FLAG-hemagglutinin (HA)-tagged Ptiwi08 under the control of Dicer-like (Dcl)5 regulatory regions, as described in a previous study of two late-expressed Ptiwi proteins due to the possibility of endo-Ptiwi08 being expressed from the new macronuclei (MACs).[10] Two timepoints were collected 6 h apart when the new MACs were clearly visible (late and late+), at the stages where Ptiwi08 has the highest expression according to a publicly available RNA sequencing (RNA-seq) expression profile (Figure 1A).[13,14] Immunoprecipitation successfully purified Ptiwi08-FLAG-HA, and co-purified sRNAs of 23 nt were visible on a UREA-PAGE, which was confirmed by deep sequencing (Figures 1B and 1C). 23 nt is also the size of siRNAs in *Paramecium* and a large fraction of the sRNAs mapped to the construct used for transformation (vector).[15] A closely related paralog of Ptiwi08, Ptiwi14, has been shown to be involved in transgene-induced silencing, and it is therefore plausible that Ptiwi08 might also have a similar role, as it appears to bind siRNAs originating from the introduced transgene construct.[9] Since our study focuses on the endogenous role of Ptiwi08 during development, we removed the vector-mapping reads before further analysis. The remaining sRNAs map with perfect complementarity to the macronuclear genome in short clusters with high depth, and the sRNAs display a very weak 5′ U signature in both timepoints (Figures 1D and 1E). Although the timepoints were taken 6 h apart, the SRCs are remarkably similar, both in relative abundance and distribution of sRNAs across the clusters. Therefore, we chose to only focus on one of these timepoints, "late," for subsequent studies. Using a sliding window analysis of Ptiwi08-bound 23 nt sRNAs, we identified SRCs with average per-base depth $\geq$50. Consecutive high-depth windows were concatenated, and cluster coordinates were subsequently refined informatically by removing all zero-depth bases from the ends of each cluster. Manual curation was necessary for some of the clusters, as low coverage regions can split one cluster into two. These cluster fractures often overlapped annotated introns, which had 19.6 times fewer mapped reads than coding sequences. This analysis identified 70 SRCs ranging from 45 to 814 bps (Table S1). We conclude that Ptiwi08-bound sRNAs map to short clusters in the genome.

A recent publication identified 2,602 endo-SRCs in vegetatively growing *Paramecium* cells.[12] Additionally, the closely related paralog of Ptiwi08, Ptiwi14, was shown to bind endo-siRNAs from SRCs, as well as siRNAs from all protein coding genes in the *Paramecium* genome.[11] Since we found Ptiwi08-bound sRNAs to also map in clusters in the genome, we crossreferenced our clusters with the previously published SRCs. Indeed, the majority of the clusters we identified in our analysis are SRCs (90%, 63 of 70 clusters), yet not all SRCs are Ptiwi08 bound (2.4%, 63 of 2,602 clusters). For the newly identified clusters, we followed the same naming convention and named them C2603–C2609. We then asked if there was anything different about these clusters, as Ptiwi08 appears to only bind a very small subset of the SRCs (70 vs. 2,602).

Many clusters overlap annotated genes in the *Paramecium* database, and we first analyzed the expression profiles of the genes overlapping the published SRC clusters and the Ptiwi08 clusters using a publicly available RNA-seq dataset.[13] Most genes overlapping SRCs have the highest expression peak during vegetative growth or at the latest developmental time point Dev4 (61%), suggesting that their expression is unrelated to sexual development (Figures S1A and S1B). In contrast to this, only 33% of genes overlapping Ptiwi08 clusters have the highest expression peak during vegetative growth or Dev4. Instead, 52% of genes have the highest expression peak in Dev1 and Dev2/3, the stages in which Ptiwi08 is also expressed. This number increases to 67% when only considering mRNAs (data not shown). In summary, our results indicate that Ptiwi08 may target development-specific mRNAs.

The sRNAs used for these two analyses were not only extracted from different life stages (vegetative vs. developmental) but were also purified by different methods: sRNAs for the SRCs were isolated through total sRNA extraction, and our Ptiwi08 clusters came from a Ptiwi08 immunoprecipitation. This could either mean that Ptiwi08 only binds a subset of SRCs or that these are the only SRCs present in this life stage. To address this, we sequenced total sRNAs from the late stage of development and performed the same sliding window analysis to define all developmental SRCs. This analysis identified 128 SRCs in the total RNA sample from the late time point, and we refer to this subset as developmental SRCs (dSRCs) (Table S2). In addition to the previously described clusters, four newly identified clusters were given IDs C2610–C2613. While this can explain part of the drop in cluster number, only half of the dSRCs were identified in the Ptiwi08 sliding window analysis. Because the depth of sequencing and chosen threshold affects the number of clusters identified, we wondered whether this discrepancy might be a consequence of these factors rather than a true biological difference. We therefore examined sRNA read counts for all dSRCs in both samples to determine if Ptiwi08 truly only binds sRNAs from half of the identified clusters. This analysis revealed that the vast majority (98%) of sRNAs that map to the macronuclear genome fall within the previously identified Ptiwi08 clusters, whereas only a very small fraction (0.3%) of the reads mapped to the remaining dSRCs (Figure S1C).

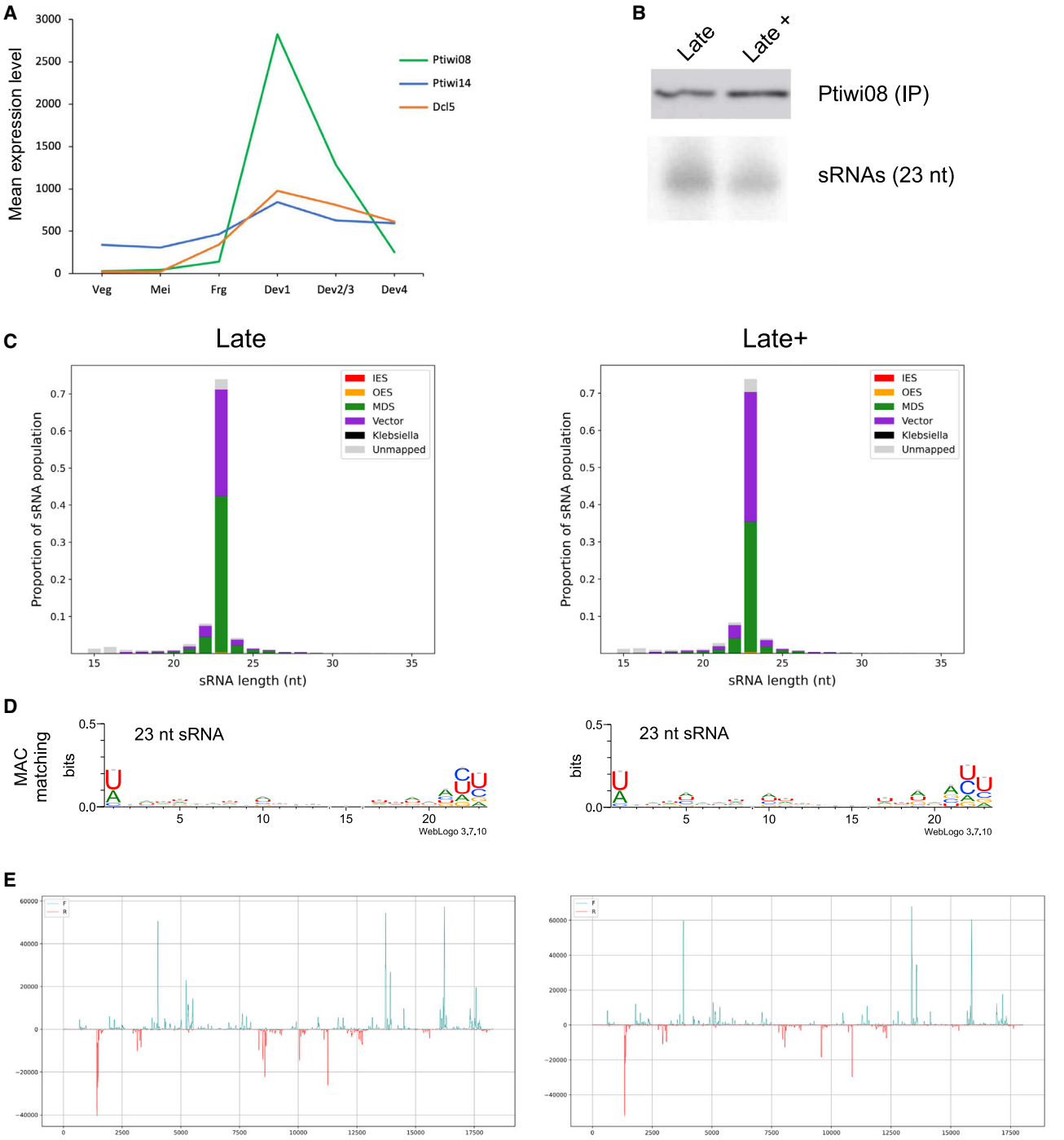


Figure 1. Features of Ptiwi08-bound small RNAs

(A) Expression profiles of Ptiwi08, Ptiwi14, and Dcl5 generated from published RNA-seq datasets.[13]

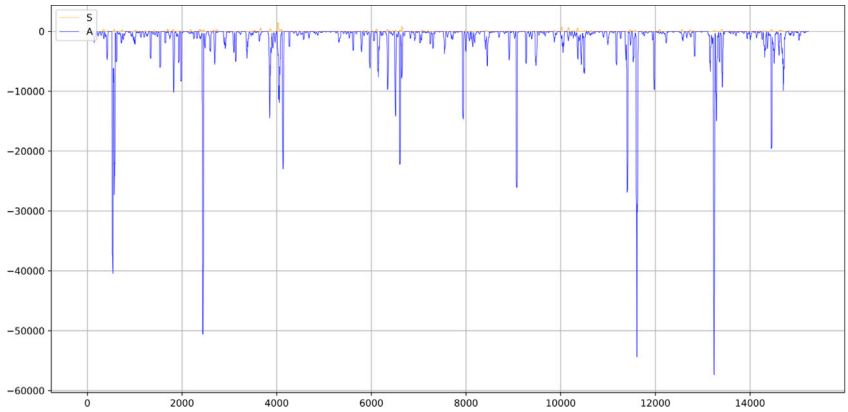
(B) Western blot of immunoprecipitated Ptiwi08 using an anti-HA antibody from late and late+ timepoints, and visualization of co-precipitated small RNAs on a 16% 7 M UREA-PAGE.

(C) Mapping and size distribution of small RNAs. IES, internally eliminated sequence; OES, other eliminated sequence; MDS, macronuclear destined sequence (MAC matching); TE, transposable element; vector, pGEM T and L4440 vectors; Klebsiella, *Klebsiella pneumoniae*.

(D) Sequence logos of small RNAs generated with Weblogo.[16] Note that the maximum y axis value has been adjusted to 0.5 instead of 2.

(E) Coverage of sRNAs mapping to the concatenated clusters. Cyan: forward; red: reverse.

A



B

	Feature	sRNAs mapping	GC percentage	Gene Type	sRNA count	SRC Length	sRNA orientation to mRNA
Class I	Coding	>99%	Low	Single copy	High	Long	Antisense
Class II	Non-coding	<1%	High	Multicopy	Low	Short	Sense or antisense

Figure 2. Ptiwi08-bound endo-siRNA clusters have an extreme strand bias and can be separated into two classes

(A) Coverage of small RNAs (sRNAs) mapping to the concatenated Ptiwi08 clusters after correction by mRNA sequencing. Yellow: sense orientation to mRNA; blue: antisense orientation to mRNA.

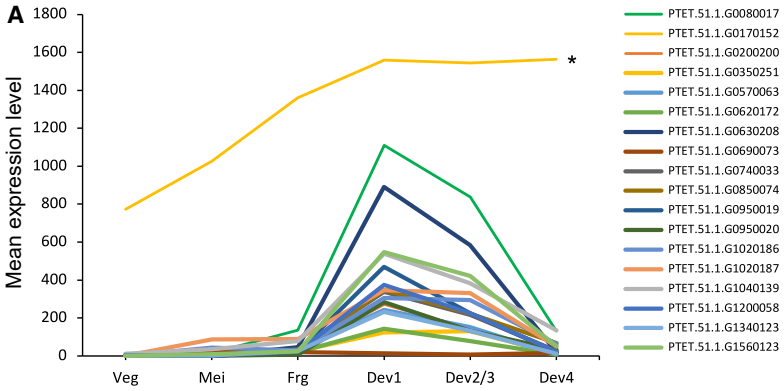
(B) Summary of features separating class I and class II clusters.

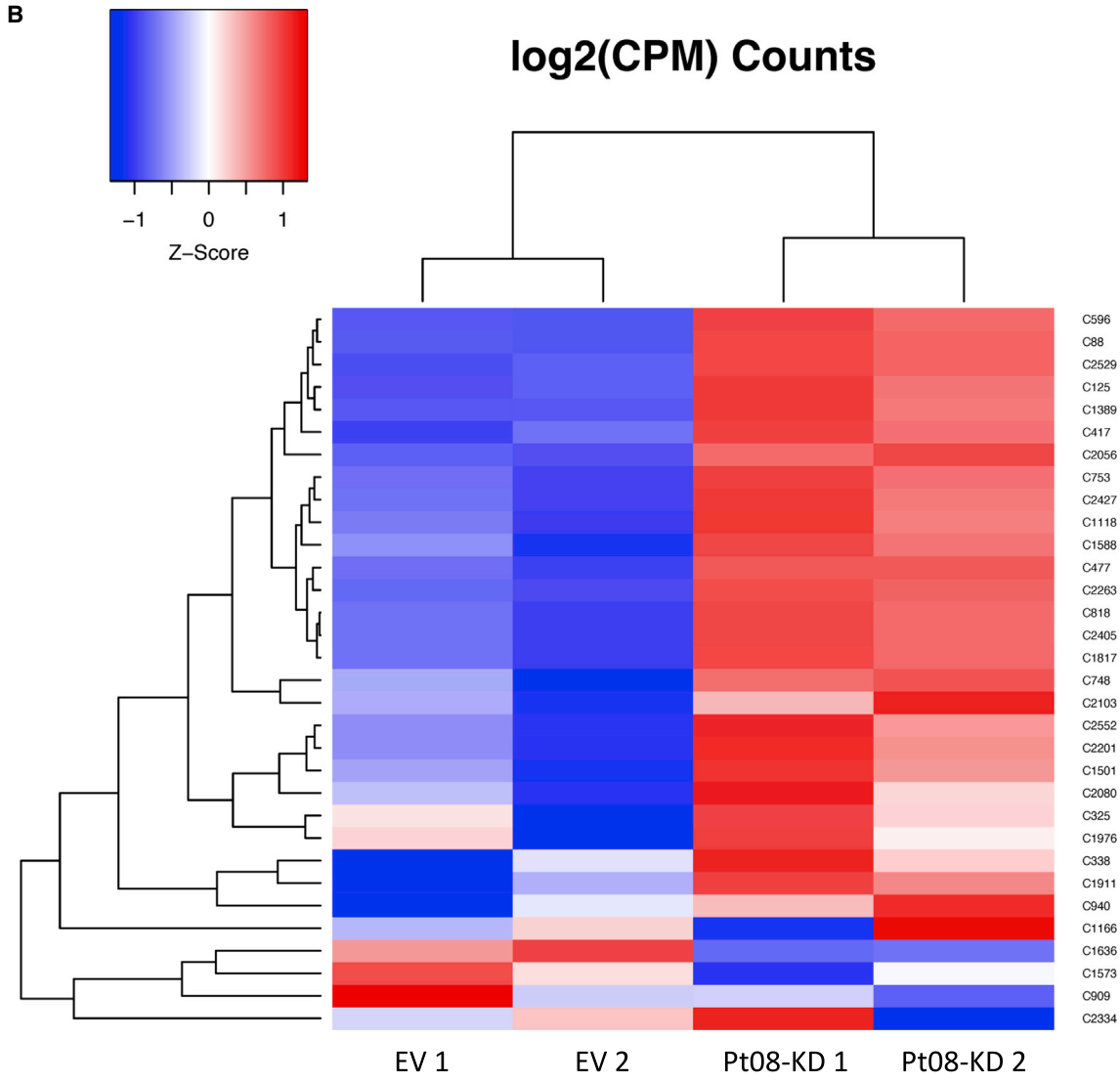
However, the same dynamics were also observed in the total sRNA sample: clusters with high sRNA coverage in the total sRNA sample were also high in the Ptiwi08 ribonucleoprotein immunoprecipitation (RIP) sample, and vice versa. Our data suggest that Ptiwi08 may bind endo-siRNAs from all dSRCs but that some clusters escape detection due to low sRNA abundance. Clusters with low read density have little support and may be unreliable and were therefore excluded from further analyses. Regardless, 98% of the sRNAs that mapped to the MAC genome were described as 70 clusters for the Ptiwi08 RIP sample, and we therefore chose to only focus on these clusters for the rest of the analyses. Taken together, our results suggest that Ptiwi08 binds endo-siRNAs mapping to dSRCs and may predominantly target developmentally regulated transcripts. Furthermore, we show that which SRCs are expressed depends on the stage of the cells and that there are fewer SRCs in late development than in vegetatively growing cells.

Ptiwi08-bound endo-siRNA clusters have an extreme strand bias and can be separated into two classes

The majority of reads are found in clusters overlapping an annotated gene in the *Paramecium* database, whereas a smaller subset of reads overlaps non-coding RNAs, such as tRNAs and 5S rRNAs. Moreover, we found nearly all (68 of 70) clusters to be stranded ($\geq$75% of reads pointing in same direction) (Table S1). In the cases where there is an annotated gene overlapping the cluster, most clusters cover the entire mRNA in the antisense direction. However, we noticed that the information in the database often did not correctly annotate the mRNAs covering these loci: some had mRNA coverage and putative open reading frames (ORFs) but no annotation, and most were completely absent in the previous version of the database. Thus, we deemed the annotations in the database too unreliable and sequenced mRNAs from the late developmental stage with a strand-specific library preparation to determine the correct mRNA transcripts as well as their orientation (i.e., strandedness). This analysis confirmed our suspicions that the annotations in the database are incomplete and revealed that all clusters overlapping mRNAs are antisense (Figure 2A; Table S1). Moreover, the sRNAs cover the entire mRNA length (data not shown). Clusters that do not overlap mRNAs mainly map to tRNAs and 5S rRNAs, but the depth of these clusters is lower, and they represent less than 1% of all reads. Taken together, our results suggest that Ptiwi08 mainly binds sRNAs that are antisense to mRNAs.

To gain further insight into the clusters, we calculated their summary statistics. Increasing cluster length correlated with increased normalized mRNA and sRNA counts, while negatively correlating with GC percentage (Figure S2). For clusters overlapping an mRNA, the orientation of the sRNAs were always opposite (antisense) to the mRNA. These relationships are shown in a





(legend on next page)

pairwise scatterplot and the loading vectors of a principal-component analysis (PCA), which described 93.8% of the variance of 7 summary statistics in 3 dimensions (Figures S3A and S3B). These summary statistics differentiate two classes of SRCs, class I and class II (Figure 2B; Table S1). Class I clusters overlap mRNAs and contain the vast majority of reads (>99% of reads in 33 of 70 clusters). Members of these clusters correspond to unique genomic sequences (only 0.3% multiple mapping). Furthermore, they are generally longer, have a lower GC percentage (24.3% on average, reference GC), and are antisense to the mRNAs they overlap (96.3% of sRNA antisense to mRNA). Class II clusters mainly overlap non-coding RNAs such as tRNAs and 5S rRNAs but only account for a small proportion of reads (<1% of reads in 37 of 70 clusters). These clusters are generally shorter, have a higher GC percentage (46.4% on average, reference GC), and correspond to duplicated sequences (68.2% of reads have multiple mapping). This results in the overrepresentation of class II reads due to the presence of multiple copies in the genome, as we kept all reads also in the case of multimappers. Unlike the strict antisense bias of class I clusters, many class II clusters are sense to the feature they overlap. However, these clusters also display a strong strand bias. Taken together, our results show that Ptiwi08-bound endo-siRNA clusters can be separated into two distinct classes: class I, which accounts for the vast majority of reads and are antisense to mRNAs; and class II, which overlaps non-coding RNAs and to which only a small number of reads map.

Ptiwi08 negatively regulates developmental mRNAs at the post-transcriptional level

Argonaute proteins are known to repress mRNAs that are complementary to their bound sRNAs.[1,2] Given that >99% of the endo-siRNAs map with perfect complementarity to mRNAs, we hypothesized that Ptiwi08 may directly repress these mRNAs at the transcriptional or post-transcriptional level. As already established, most of these mRNAs have the highest expression peak in the late stage of development, at the same stage as Ptiwi08 expression peaks. The previous analysis included putative gene annotations that we found to lack mRNA coverage in the late stage of development by mRNA-seq, as well as overlapping transcription units where only one unit is covered by sRNAs. Our mRNA-seq analysis in the previous section identified the most likely targets of these sRNAs, and we therefore used the annotations remaining after correction by mRNA-seq for further analyses. We first observed the dynamics of mRNA expression throughout development and found that most of the genes display a clear peak in Dev1, before rapidly disappearing (Figure 3A). Only one gene did not follow the same dynamics, the constitutively expressed gene PTET.51.1.G0170152 (overlapping cluster C909). As the sRNAs display a strong antisense bias (>90% in each cluster), they have the potential to target the mRNAs directly through base pairing and may regulate their expression. To explore this possibility, we performed mRNA-seq upon Ptiwi08 silencing. We also co-silenced Ptiwi14 for these analyses, as we are unable to silence only Ptiwi08 due to a high sequence similarity, and as previously mentioned, they may have partly overlapping functions. However, since Ptiwi14 may have additional roles, we focused only on the mRNAs overlapping the clusters we identified, which allows us to interrogate the functions of Ptiwi08. Our analysis confirmed that most mRNAs overlapping class I clusters are significantly upregulated in the absence of Ptiwi08, indicating a suppressive role of these sRNAs (Figure 3B; Table S3).

We then sought to determine the mode of action of Ptiwi08. Argonaute proteins post-transcriptionally targeting mRNAs are generally localized in the cytoplasm, whereas nuclear-localized Argonautes usually transcriptionally regulate its targets. The perfect complementarity and conservation of the catalytic triad in the PIWI domain of Ptiwi08 indicates an active slicer activity, which may suggest that Ptiwi08 negatively regulates these mRNAs at the post-transcriptional level. To test this hypothesis, we observed the localization of a GFP-tagged Ptiwi08 to determine its subcellular localization. These results confirmed a cytoplasmic localization, suggesting that Ptiwi08 likely targets the mRNA post-transcriptionally (Figure S4).

Ptiwi08 targets do not appear to produce proteins

The identification of a dedicated Ptiwi protein regulating developmental mRNAs indicates a need for their specific degradation. To find out more about the function of these mRNAs, we first attempted to investigate the putative proteins encoded in the mRNAs with no success: attempts at expressing the proteins failed under their endogenous promoter, as well as under the promoter of Dcl5, which is normally expressed at the same time as Ptiwi08 (data not shown). We next asked if there are protein products at all in the cells, so we extracted total proteins from the late time point of a control (empty vector [EV]) as well as Ptiwi08/14 silenced culture, which were sent for mass spectrometry. We reasoned that since Ptiwi08 regulates these mRNAs, we might find more protein products when it is silenced. As the putative protein products are small, they were size selected (<30 kDa) before mass spectrometry. We found no hits for any of the putative proteins these mRNAs encode (Table S4). However, we cannot exclude that they are present in low abundance and may escape detection in our approach. Additionally, as mentioned previously, the gene annotations are poor representatives of the clusters and the mRNAs they overlap. Therefore, the peptides used for identification in the mass spectrometry analysis (for known and annotated genes) may not be in the correct reading frame or even from the correct mRNA. Some annotations have convergent transcription units, split annotations with mRNA coverage along the entire region, or lack annotations altogether despite consistent mRNA coverage (Figures S5A–S5C). Therefore, we hypothesized that the mRNAs may be aberrantly transcribed and untranslated. We next examined these putative protein sequences more in

Figure 3. Ptiwi08 negatively regulates developmental mRNAs at the post-transcriptional level

(A) Expression profiles of annotated coding genes overlapping the Ptiwi08 clusters generated from published RNA-seq datasets.[13] Asterisk indicates PTET.51.1.G0170152 (overlapping cluster C909).

(B) Heatmap of the expression of all mRNAs overlapping class I clusters in control (EV) and Ptiwi08-KD (Pt08-KD) samples. Numbers indicate replicates.

detail. They are very small, with the majority being less than 20 kDa and having no annotated domains. The *Paramecium* genome is very AT rich, with only 28% G + C in the macronuclear genome, and we found the putative protein-coding clusters (class I) to have a significantly lower G + C percentage (median, mean = 23.8, 24.3) than the pool of annotated protein-coding genes (median, mean = 29.2, 29.4) (Figure S5D).[17] We also noted that none of the genes have any paralogs, which is uncommon for *Paramecium* genes after three whole-genome duplication events, nor were we able to find any homology to known proteins (data not shown).[17] Based on these findings, we deem it unlikely that these mRNAs encode proteins. The closely related paralog of Ptiwi08, Ptiwi14, has been implicated in the regulation of untranslatable transgenes, and we therefore hypothesize that the targets of Ptiwi08 may be chosen based on translation potential.

We noted that the mRNA expression of the targets coincides with the start of zygotic transcription when the new MACs gain transcriptional activity. The derepression of the zygotic genome occurs at the time where massive genome rearrangements take place, which may lead to aberrant transcription, to which a dedicated machinery featuring Ptiwi08 must respond. Taken together, our results suggests that these mRNAs are aberrantly transcribed at the start of zygotic genome activation but do not encode for functional proteins. Hence, this endo-siRNA pathway might have evolved to repress the transcripts of non-functional mRNAs in this fragile stage of development.

Ptiwi08-bound sRNAs are Dicer dependent

All Ptiwi-bound sRNAs in *Paramecium* that have been investigated so far are produced in a Dicer-dependent manner and map with perfect complementarity to their targets.[9–11,15,18] Hence, their biogenesis more resembles siRNAs than PIWI-interacting RNAs (piRNAs). The strong strand bias of the Ptiwi08-bound endo-siRNAs may suggest that they derive from single-stranded precursors or that they are generated from double-stranded precursors and undergo strong strand selection after production. To find out more about the biogenesis of these endo-siRNAs, we decided to look more closely into this. Although there are only few sense sRNAs mapping to class I clusters, calculating the 5′ to 5′ distance on opposite genomic strands revealed that 1.2% of reads from the Ptiwi08 RIP and 2% of the reads from the total sRNA sample display a 21 bp overlap with 2 nt 5′ overhangs, characteristics of RNase III-mediated cleavage (Figure 4A). This suggests that these sRNAs may also be dicer products. *Paramecium* has 3 Dicers (Dcrs) as well as 5 Dcl proteins encoded in its genome.[15,18] Since our endo-siRNA clusters are a subset of the vegetative SRCs, we found it unlikely that a separate Dicer would produce them in vegetative and developmental stages, which excluded the Dicer-like proteins exclusively expressed during development. Our attention was drawn to Dcr1, as this is the only constitutively expressed Dicer protein harboring the conserved residues necessary for cleavage in the RNaseIII domains.[15] Dcr1 generates 23 nt siRNAs participating in post-transcriptional gene silencing and transgene-induced silencing mechanisms in vegetatively growing *Paramecium* cells. However, a potential role during development has not been investigated. Therefore, we silenced Dcr1 and performed sequencing of sRNAs extracted from the late stage of development (Figure S6). We observed an overall decrease in quantity of 23 nt sRNAs in the Dcr1-silenced sample (2.7-fold reduction), as expected if this size class is generated by Dcr1 (Figure 4B). Next, we examined the 23 nt sRNAs mapping to the class I clusters we identified in the Ptiwi08 RIP sample. We observed a 3.9-fold decrease in 23 nt sRNAs in class I SRCs in the Dcr1-silenced sample compared with the control (Figure 4C).

Calculating the 5′ to 5′ distance on opposite genomic strands revealed an increase in 21 bp overlaps (i.e., Dicer signal) from 2.3% in the control to 4.3% in the Dcr1 knockdown (KD) (Figure 4D). This increase mainly originates from cluster C1166 (cluster 22) and, to a minor extent, C909, C2056, and C2529 (Figure 4D). The remaining clusters do not display a difference in the Dicer signal strength compared with the control. It is striking that the exception is C1166 (cluster 22), a cluster first reported in 2009 by small-scale sRNA sequencing.[15] Of all class I clusters, C1166 displays the highest proportion of sense sRNAs (9.82%) and may therefore be considered a semi-dual-strand cluster. Examining sRNAs from the Ptiwi08 RIP and the overall sRNA pool in the late time point (EV late) revealed that this cluster is overrepresented in the latter and accounts for the majority of the increase in Dicer signal (from 1.2% in Ptiwi08 RIP to 2% in the EV late) (Figure 4A). These results suggest that sRNAs mapping to C1166 are only bound to a minor extent by Ptiwi08 (compared with the total sRNA pool) and that another Ptiwi protein might preferentially stabilize these sRNAs. Additionally, it appears that sRNAs mapping to C1166 do not display the same reduction as most of the Ptiwi08 clusters when Dcr1 is depleted; thus, the increase in Dicer signal strength may be a consequence of the selective reduction of sRNAs mapping to unistrand clusters. This is consistent with a previous study of RNAi factors in vegetatively growing *Paramecium* cells, in which C1166 endo-siRNAs were not affected by Dcr1 mutations or depletion.[19] Hence, endo-siRNAs mapping to C1166 may be both Ptiwi08 as well as Dcr1 independent. Taken together, our results suggest that the Ptiwi08-bound endo-siRNAs are Dcr1 dependent, despite the strict strand bias.

Endo-siRNAs are likely single stranded before loading onto Ptiwi08

Dicer-dependent sRNA duplexes associate with Argonaute proteins, which remove the passenger strand to generate a single-stranded sRNA capable of sequence-specific targeting.[20–22] This was also shown for the Dicer-dependent sRNAs loaded onto Piwi-clade proteins involved in DNA elimination in *Paramecium* and *Tetrahymena*.[10,23] The strong antisense bias we observed for the Dcr1-dependent endo-siRNAs may indicate strand selection by Ptiwi08. If strand selection occurs after loading onto the Argonaute, the sRNAs should remain double stranded in the absence of Ptiwi08. To determine if this was the case, we sequenced sRNAs from cells depleted of Ptiwi08. As in previous experiments, Ptiwi14 was also co-silenced. We observed an overall decrease in the quantity of 23 nt sRNAs mapping to most clusters in the Ptiwi08-KD culture yet a modest increase in Dicer signal from 2.267% in the EV control to 2.970% in the Ptiwi08 KD (Figures 5A and 5B). Closer inspection of the clusters contributing to this signature revealed that the increase

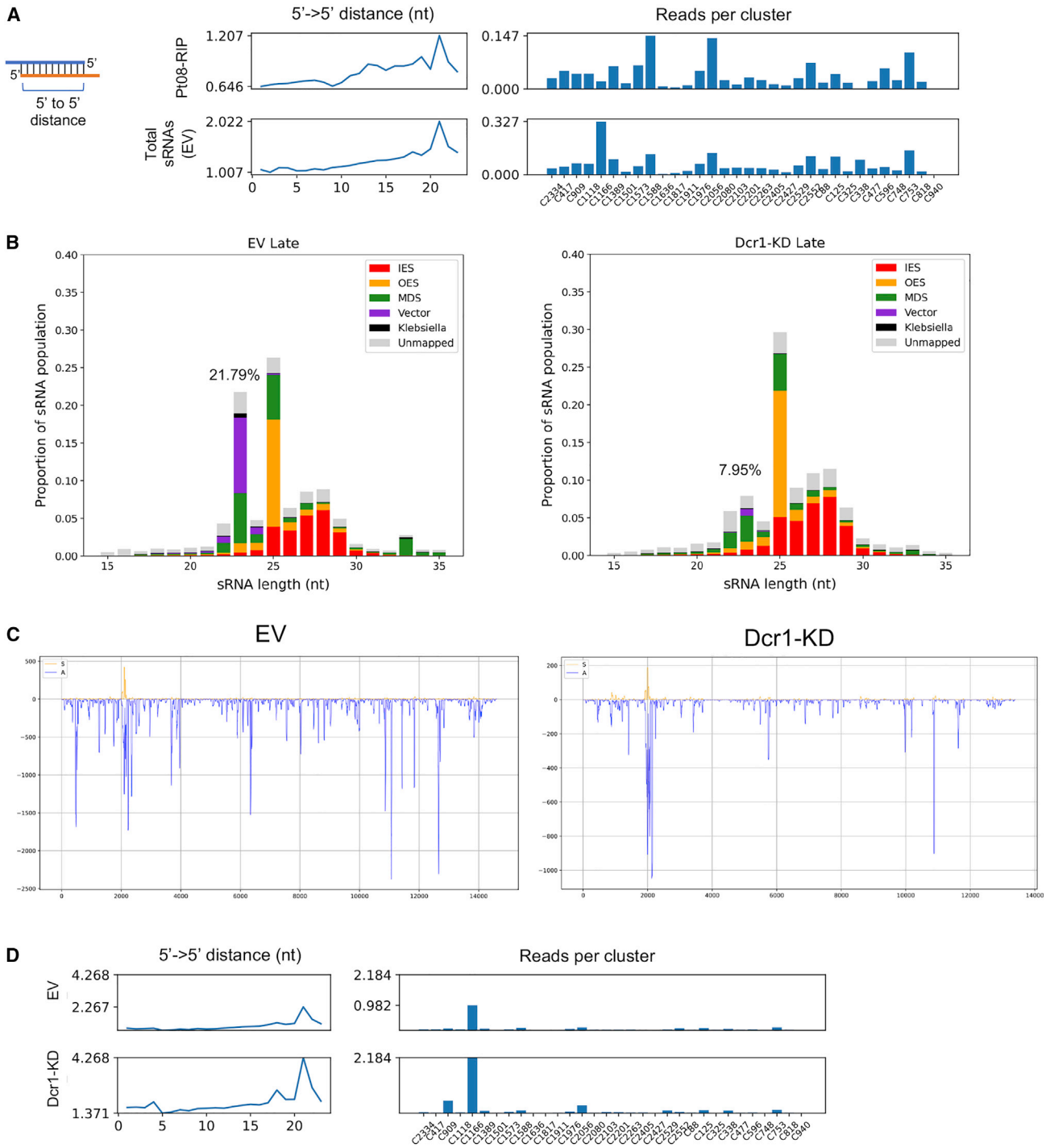


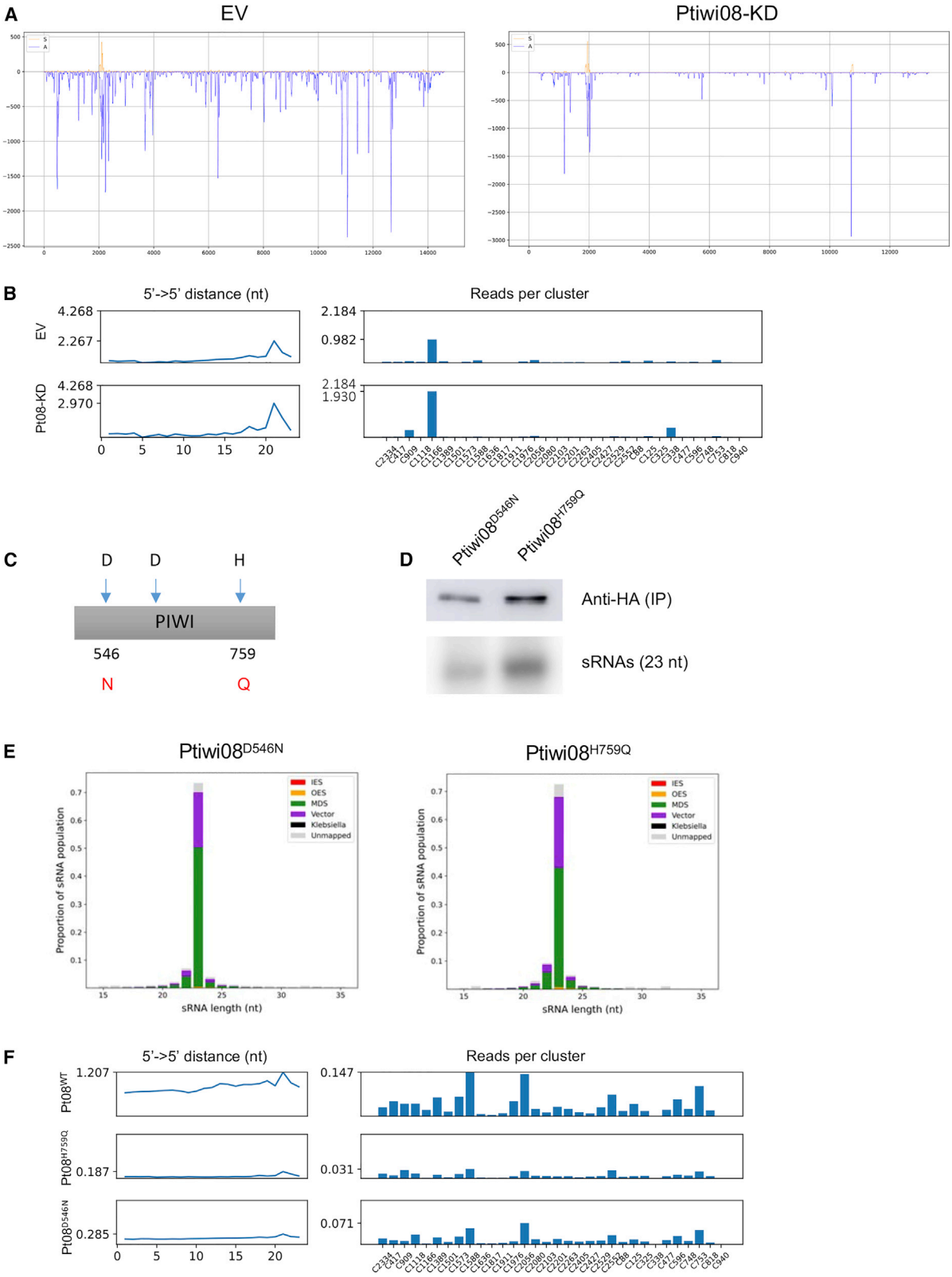
Figure 4. Ptiwi08-bound sRNAs are Dicer dependent

(A) Left: schematic depicts calculation of 5′ to 5′ distance on opposite genomic strands. Middle: percentage of sRNAs with complementarity (5′ to 5′ distance on opposite genomic strands) mapping to Ptiwi08 clusters in Ptiwi08 RIP late and total sRNAs from EV late datasets. Right: contribution of each cluster to the Dicer signal (peak at 21).

(B) Mapping and size distribution of sRNAs. Percentage denotes fraction of 23 nt sRNAs. IES, internally eliminated sequence; OES, other eliminated sequence; MDS, macronuclear destined sequence (MAC matching); TE, transposable element; vector, pGEM T and L4440 vectors; Klebsiella, *Klebsiella pneumoniae*.

(C) Coverage of sRNAs mapping to class I clusters. Yellow: sense orientation to mRNA; blue: antisense orientation to mRNA.

(D) Percentage of sRNAs with complementarity (5′ to 5′ distance on opposite genomic strands) mapping to Ptiwi08 clusters in EV late and Dcr1-KD late datasets, as well as each clusters contribution to the Dicer signal (peak at 21).



(legend on next page)

is almost entirely derived from cluster C1166 (and to a minor extent C909 and C338), similar to the case we observed for Dcr1 KD in the previous section. As discussed, this cluster is likely both Ptiwi08 as well as Dcr1 independent. For the rest of the clusters, we did not observe an increase in Dicer signal in the absence of Ptiwi08. These results may suggest that the sRNAs were single stranded before loading onto Ptiwi08, or that this is a consequence of the lack of a complete silencing, since we are unable to fully remove the protein.

Catalytically active Argonaute proteins use their slicer activity to cleave the passenger strand of perfect or nearly perfect duplexes generated by Dicer.[20–22] The slicer activity is found in the RNase-H-like fold harboring the catalytic residues within the PIWI domain.[24,25] Mutations in these residues generate slicer-dead mutants, which are unable to cleave target RNAs.[24] Alignment of Ptiwi08 to other eukaryotic Argonautes revealed that the PIWI domain harbors the catalytic residues required for slicer activity.[9] It is therefore possible that this activity is used for removal of the passenger strand after loading duplex sRNAs. To investigate this possibility, we generated two slicer mutants of Ptiwi08 by introducing two single amino acid mutations in the catalytic residues in the PIWI domain (Figure 5C). Using these mutants, we performed immunoprecipitation and isolated co-purified sRNAs. We could immunopurify the mutant proteins and their associated sRNAs to the same extent as wild-type Ptiwi08, suggesting that the mutations did not affect the stability of the protein nor sRNA loading (Figures 5D and 5E). Contrary to the hypothesis, the mutants did not load double-stranded duplexes (Figure 5F). Instead, the 21 bp overlap decreased in the mutants from 1.207% in the Ptiwi08 wild type to 0.187% and 0.285% in the Ptiwi08^{H759Q} and Ptiwi08^{D546N} mutations, respectively. These results suggest that the sRNAs are single stranded before loading onto Ptiwi08 and that the 21 bp overlap may at least partly occur after Ptiwi08 slicing. Moreover, the lack of sequence specificity of Ptiwi08-bound sRNAs suggests that Ptiwi08 is unlikely to choose a strand based on the sequence alone, as it does not display any nucleotide preference; hence, the selection is unlikely to be explained by its thermodynamic properties (Figure 1D). We conclude that the antisense bias is independent of Ptiwi08 and its slicer activity and that the sRNAs may be single stranded before they are loaded into Ptiwi08.

Endo-siRNAs are 2′-O-methylated by Hen1, an interaction partner of Ptiwi08

Our data so far suggest that the sRNAs are generated by Dcr1 and loaded onto Ptiwi08, which post-transcriptionally regulates untranslated mRNAs. However, the origin of the strong strand bias remains elusive. Since we found Dicer dependence, but no indication of strand selection by Ptiwi08 itself, we investigated putative modifications that may aid in selection. A candidate for this is 2′-O-methylation on the 3′ ends of sRNAs, which protects sRNAs from degradation and thus aid in their stability. This modification is widely found in a several sRNA classes including piRNAs, plant microRNAs (miRNAs), and fly endo-siRNAs.[26–32] In a related ciliate, *Tetrahymena*, 2′-O-methylation has been demonstrated to be set on scan RNAs (scnRNAs; involved in DNA elimination) but not siRNAs.[33] In favor of our hypothesis, we found this modification on the Ptiwi08 co-precipitated sRNAs (Figure 6A). We hypothesized that this modification may be preferentially set on the antisense strand of the duplex and that this may be a prerequisite for loading onto Ptiwi08. Therefore, we looked for the presence of the methyltransferase responsible for setting this mark, Hen1. One Hen1 homolog is encoded in the *Paramecium* genome, which is constitutively expressed based on RNA-seq data and a northern blot time course (Figures S7A and S7B).[14] To test our hypothesis, we extracted and sequenced sRNAs from cells depleted of Hen1 (Figure S5C). We could extract the same amount of 23 nt sRNAs as in the control, suggesting that the modification is not required for the stability of the sRNAs (Figure 6B). However, we did not observe an increase in strand bias or Dicer signal strength of the sRNAs mapping to the clusters when Hen1 was depleted (Figures 6C and 6D). On the contrary, the proportion of reads displaying a 21 bp overlap decreased from 2.022% to 0.953%, suggesting that the SRCs are in fact "more" single stranded in the absence of Hen1. Hence, this modification is unlikely to be the mode of selection.

Upon closer inspection of Hen1, we found the double-stranded RNA (dsRNA)-binding domain, which is present in the plant Hen1 protein but absent in metazoans, to also be absent in *Paramecium* Hen1. The absence of this domain has implications for the sRNAs that Hen1 can methylate: plant Hen1 exclusively modifies dsRNA duplexes, and fly Hen1 exclusively modifies single-stranded RNAs (ssRNAs).[29,31,32] Thus, Hen1 in *Paramecium* likely modifies ssRNAs. Consequently, it is unlikely that this modification is used as a way for preferential strand selection, and the sRNAs are likely modified after binding to the Ptiwi protein. In support of this, we found Hen1 and Ptiwi08 to interact in both native (non-crosslinked) and crosslinked conditions, suggesting that the modification is set on single-stranded sRNAs complexed with Ptiwi08 (Figure 6E). In conclusion, Ptiwi08 interacts with Hen1, and Hen1 2′-O-methylates endo-siRNAs at their 3′ ends. While we do not currently know the importance of this modification, our results indicate that it is dispensable for the strand selection of endo-siRNAs.

Figure 5. Endo-siRNAs are likely single stranded before loading onto Ptiwi08

(A) Coverage of sRNAs mapping to class I clusters. Yellow: sense orientation to mRNA; blue: antisense orientation to mRNA.

(B) Percentage of sRNAs with complementarity (5′ to 5′ distance on opposite genomic strands) mapping to Ptiwi08 clusters in EV late and Ptiwi08-KD late datasets, as well as each clusters contribution to the Dicer signal (peak at 21).

(C) Positions of mutated residues in the PIWI domain of Ptiwi08.

(D) Western blot of immunoprecipitated Ptiwi08 slicer mutants using an anti-HA antibody, and visualization of co-precipitated sRNAs on a 16% 7 M UREA-PAGE.

(E) Mapping and size distribution of sRNAs. IES, internally eliminated sequence; OES, other eliminated sequence; MDS, macronuclear destined sequence (MAC matching); TE, transposable element; vector, pGEM T and L4440 vectors; Klebsiella, *Klebsiella pneumoniae*.

(F) Percentage of sRNAs with complementarity (5′ to 5′ distance on opposite genomic strands) mapping to Ptiwi08 clusters after immunoprecipitation of Ptiwi08^{WT} and two slicer mutants (Ptiwi08^{D546N} and Ptiwi08^{H759Q}), as well as each cluster's contribution to the Dicer signal (peak at 21).

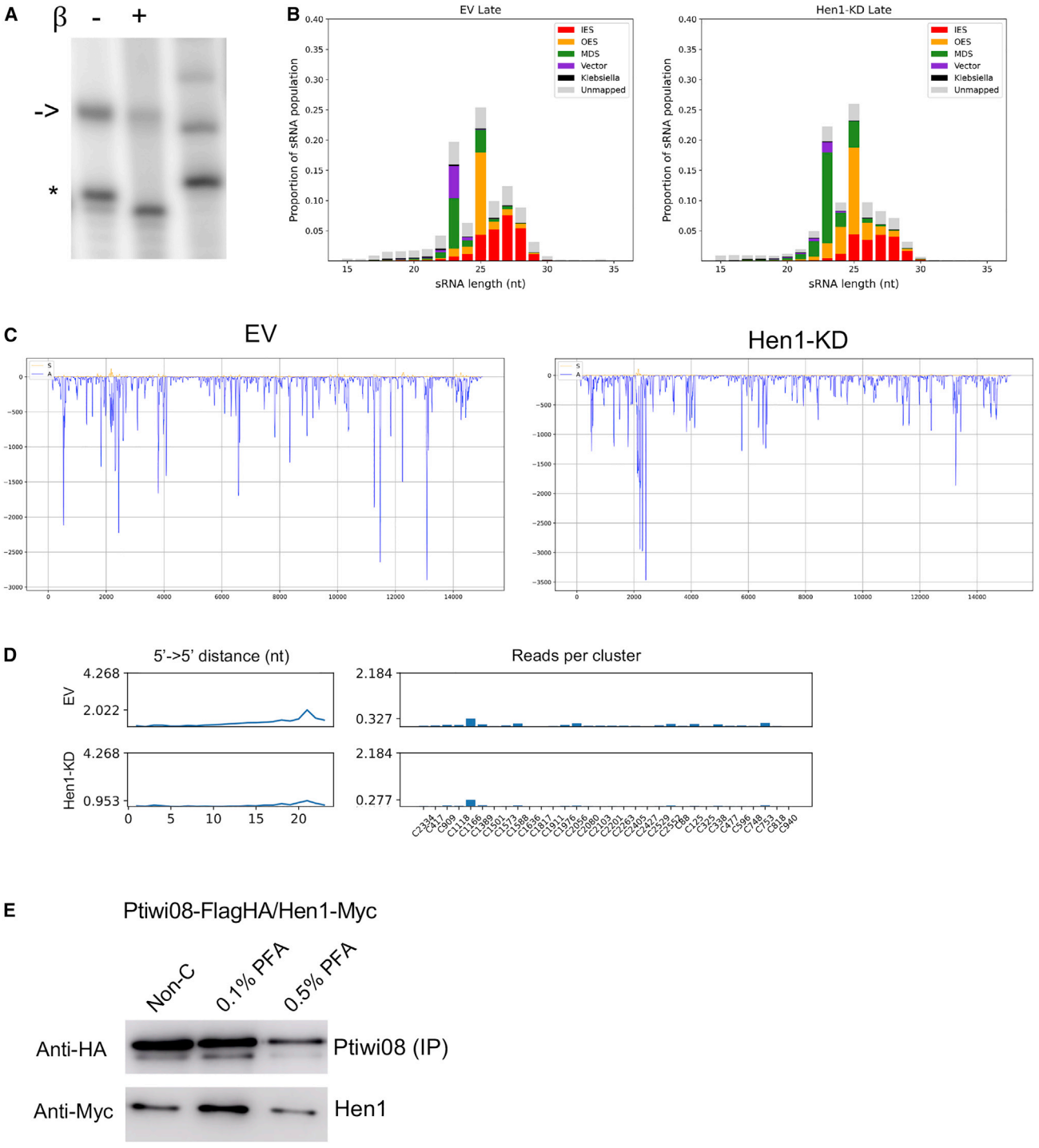


Figure 6. The endo-siRNAs are 2′-O-methylated by Hen1, an interaction partner of Ptiwi08

(A) Beta elimination of Ptiwi08-FLAG-HA co-precipitated sRNAs mixed with an unmodified 17 nt RNA oligonucleotide (oligo). Arrow indicates the Ptiwi08-bound sRNAs and the asterisk the unmodified 17 nt RNA oligo.

(B) Mapping and size distribution of small RNAs. IES, internally eliminated sequence; OES, other eliminated sequence; MDS, macronuclear destined sequence (MAC matching); TE, transposable element; vector, pGEM T and L4440 vectors; Klebsiella, *Klebsiella pneumoniae*.

(C) Coverage of sRNAs mapping to class I clusters. Yellow: sense orientation to mRNA; blue: antisense orientation to mRNA.

(legend continued on next page)

DISCUSSION

In this work, we discovered an endo-siRNA-mediated developmental mRNA clearance pathway in the unicellular eukaryote *P. tetraurelia* (Figure 7). We demonstrated that these endo-siRNAs are bound by an Argonaute protein, Ptiwi08, and interact with the methyltransferase Hen1, which is required for 2′-O-methylation on siRNA 3′ ends. Furthermore, we found the endo-siRNAs to be Dcr1 dependent despite displaying a strict strand bias. Calculating summary statistics for these clusters allowed us to distinguish two classes of clusters, class I and class II, based on their characteristics. By sequencing mRNAs upon silencing of Ptiwi08, we provided evidence that class I clusters, which account for 98% of all Ptiwi08 RIP reads that map to the macronuclear genome, are negatively regulated by the Ptiwi08 pathway at the post-transcriptional level. Finally, we hypothesize that these mRNAs are likely untranslated and postulate that the main role of Ptiwi08 is the regulation of aberrant, untranslated mRNAs during development or development-specific degradation of "unwanted" transcripts that would otherwise interfere with the correct developmental progression.

The endo-siRNA pathway described in this study displays an unusual combination of an extreme strand bias and Dicer dependence, which is seemingly independent of the Argonaute that binds it. Notably, similar SRCs were found in the related ciliate *Tetrahymena thermophila*.[34,35] Also here, the sRNAs are found in clusters of high depth and are bound by a Piwi protein, Twi2. The SRCs display an extreme strand bias that is antisense to genes, which the authors propose are pseudogenes. The putative proteins encoded on the antisense strand of the Twi2-bound SRCs share many similarities to the ones described in this study, and neither appears to encode functional proteins. Hence, we believe that the function of both pathways is the degradation of aberrant, untranslatable transcripts. Similarly, the closely related paralog of Ptiwi08, Ptiwi14, is constitutively expressed and has been implicated in transgene-induced silencing (i.e., the regulation of untranslatable transgenes).[9] Due to the similarity with Ptiwi14, Ptiwi08 has been hypothesized to bind the same sRNAs as Ptiwi14 and bring them into the zygotic nucleus as a means of transgenerational inheritance.[36] Here, we demonstrate that while this is not strictly true, there are striking similarities between the regulation of untranslatable transgenes and the developmental mRNA clearance we described in this study. The same key players appear to be involved (Ptiwi, Dcr1), and our results suggest that the mRNAs targeted by the Ptiwi08 pathway may also be untranslated, as is the trigger for transgene-induced silencing by Ptiwi14. In favor of this hypothesis, similar regulation was recently reported in various animals for maternal mRNA clearance during early stages of embryogenesis, where untranslated mRNAs devoid of ribosomes are degraded in an Argonaute-dependent manner.[3–5,7] Although there are similarities between these systems, there are also clear differences, perhaps the most important of which is that the role of these pathways in ciliates appears to be the clearance of non-functional, aberrantly transcribed mRNAs, not to clear mRNAs that have fulfilled their functional role in the early stages of development. Nonetheless, we hypothesize that a common prerequisite for targeting by these pathways is untranslatable mRNAs.

Despite the strong antisense bias of the endo-siRNAs and lack of Ptiwi08 dependence for their selection, we found the sRNAs to be Dcr1 dependent. In other systems, Argonautes bind dsRNA duplexes generated by Dicer cleavage and remove the passenger strand using their slicer activity, which we showed was not the case for Ptiwi08 with the use of slicer mutants. Neither did we observe any increase in duplexes for these clusters when Ptiwi08 was silenced, and the selection appears independent of 2′-O-methylation. Hence, we believe the endo-siRNAs are single stranded before loading onto Ptiwi08, which suggests that they may derive from single-stranded precursors. Further work is required to determine the origin of the strict antisense bias of these sRNAs.

Argonaute proteins in several animals were recently shown to be involved in the clearance of maternal mRNAs accompanying the transcriptional activation of the zygotic genome, known as the MZT.[3–8] In *Paramecium*, the analogous process occurs in the late stage of development as the developing zygote gains transcriptional activity. This developmental stage features large-scale genome rearrangements and DNA elimination to form a streamlined new soma free of transposons and transposon remnants.[37] Here, we demonstrated that this stage also utilizes an Argonaute protein, Ptiwi08, as a means of clearing developmental mRNAs. The loss of transcriptional repression concomitant with massive chromatin changes poses a unique challenge, which might necessitate post-transcriptional silencing to counteract aberrant transcription. Though there are clear differences between the maternal mRNA clearance during the MZT occurring in multicellular organisms and what we reported here for a unicellular eukaryote, we hypothesize that both systems evolved with a common goal: degradation of mRNAs not engaged in translation at a fragile stage of development. This suggests that the use of Argonaute proteins may be a widespread mechanism for mRNA clearance during development, both in multicellular as well as unicellular eukaryotes.

Limitations of the study

This study discovered a sRNA-mediated mRNA clearance pathway in which the PIWI clade Argonaute protein Ptiwi08 negatively regulates a set of developmental transcripts during sexual development. While our results support the notion that these transcripts are untranslated, we do not yet have conclusive proof that they do not encode proteins. We refer to these transcripts as mRNAs because of the presence of putative ORFs and untranslated regions and because they have been annotated as putative protein-coding genes in the Paramecium

(D) Percentage of sRNAs with complementarity (5′ to 5′ distance on opposite genomic strands) mapping to Ptiwi08 clusters in EV late and Hen1-KD late datasets, as well as each clusters contribution to the Dicer signal (peak at 21).
(E) Western blots after immunoprecipitation of Ptiwi08-FLAG-HA in Ptiwi08-FLAG-HA/Hen1-Myc co-transformed cultures, using anti-HA or anti-Myc antibodies to detect Ptiwi08 and Hen1.

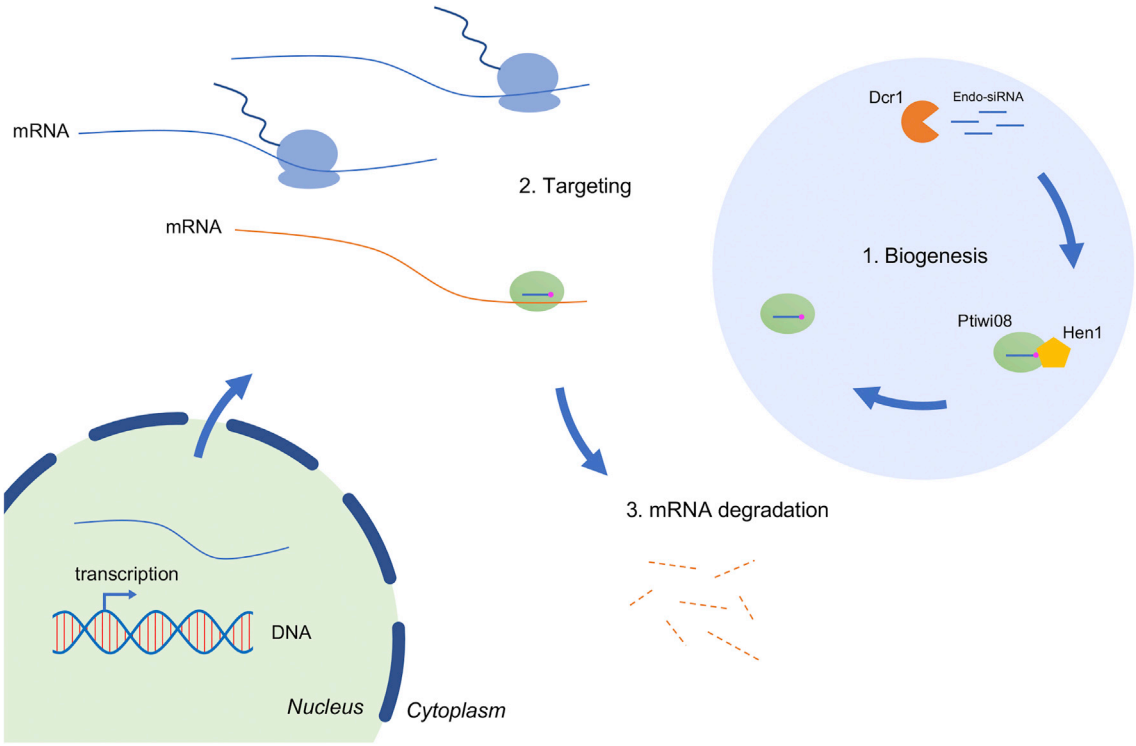


Figure 7. Model of the Ptiwi08 pathway

Biogenesis: endo-siRNAs are generated in a Dcr1-dependent manner and loaded onto Ptiwi08, after which Hen1 sets 2′-O-methylation on endo-siRNA 3′ ends. Targeting: mRNAs not engaged in translation are targeted by the Ptiwi08 pathway. mRNA degradation: targeted mRNAs are post-transcriptionally degraded by the Ptiwi08 pathway.

Genome database. However, one cannot exclude the possibility that at least some of the transcripts are non-coding RNAs. Further work is needed to fully understand the nature of these transcripts and whether they are simply byproducts of transcriptional activation or have an unidentified role with biological significance.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Constructs
  - Cytological staging
  - Macronuclear transformation by microinjection
  - GFP localization and imaging
  - Silencing by feeding
  - Total RNA extraction and sequencing
  - Northern blot
  - RNA immunoprecipitation
  - Co-immunoprecipitation
  - Small RNA 5′ end labeling
  - Beta elimination
  - Western blot
  - Mass spectrometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatic methods
  - Sequence processing and alignment
  - Cluster identification and characteristics
  - Iterative read alignment and subtraction
  - Introns vs CDS features
  - mRNAseq sequence alignment
  - Differential expression

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112213.

ACKNOWLEDGMENTS

We thank Dr. Nassikhat Stahlberger for technical support and the Proteomics and Mass Spectrometry Core Facility (PMSCF) at the Department for Biomedical Research (DBMR), University of Bern, Switzerland, for mass spectrometry analyses. This work was supported by European Research Council grants (ERC) 260358 "EPIGENOME" and 681178 "G-EDIT"; Swiss National Science Foundation grants 31003A_146257, 31003A_166407, and 310030_184680; and grants from the National Center of Competence in Research (NCCR) RNA and Disease to M.N.

AUTHOR CONTRIBUTIONS

T.S. and M.N. designed the experiments. T.S. performed experiments with help from C.W. V.M. and T.S. performed the bioinformatic analyses. V.M. wrote python and R code. T.S. wrote the manuscript with input from the other authors. M.N. supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: October 6, 2022
Revised: December 21, 2022
Accepted: February 17, 2023

REFERENCES

1. Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2019). PIWI-interacting RNAs: small RNAs with big functions. Nat. Rev. Genet. *20*, 89–108. https://doi.org/10.1038/s41576-018-0073-3.
2. Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat. Rev. Genet. *12*, 19–31. https://doi.org/10.1038/nrg2916.
3. Zhang, J.M., Hou, W.B., Du, J.W., Zong, M., Zheng, K.L., Wang, W.J., Wang, J.Q., Zhang, H., Mu, Y.S., Yin, Z., et al. (2020). Argonaute 2 is a key regulator of maternal mRNA degradation in mouse early embryos. Cell Death Discov. *6*, 133. https://doi.org/10.1038/s41420-020-00368-x.
4. Quarato, P., Singh, M., Cornes, E., Li, B., Bourdon, L., Mueller, F., Didier, C., and Cecere, G. (2021). Germline inherited small RNAs facilitate the clearance of untranslated maternal mRNAs in C. elegans embryos. Nat. Commun. *12*, 1441. https://doi.org/10.1038/s41467-021-21691-6.
5. Singh, M., Cornes, E., Li, B., Quarato, P., Bourdon, L., Dingli, F., Loew, D., Proccacia, S., and Cecere, G. (2021). Translation and codon usage regulate Argonaute slicer activity to trigger small RNA biogenesis. Nat. Commun. *12*, 3492. https://doi.org/10.1038/s41467-021-23615-w.
6. Laue, K., Rajshekar, S., Courtney, A.J., Lewis, Z.A., and Goll, M.G. (2019). The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. Nat. Commun. *10*, 1551. https://doi.org/10.1038/s41467-019-09582-3.
7. Sha, Q.Q., Zhu, Y.Z., Li, S., Jiang, Y., Chen, L., Sun, X.H., Shen, L., Ou, X.H., and Fan, H.Y. (2020). Characterization of zygotic genome activation-dependent maternal mRNA clearance in mouse. Nucleic Acids Res. *48*, 879–894. https://doi.org/10.1093/nar/gkz1111.
8. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79. https://doi.org/10.1126/science.1122689.
9. Bouhouche, K., Gout, J.F., Kapusta, A., Bétermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling. Nucleic Acids Res. *39*, 4249–4264. https://doi.org/10.1093/nar/gkq1283.
10. Furrer, D.I., Swart, E.C., Kraft, M.F., Sandoval, P.Y., and Nowacki, M. (2017). Two sets of Piwi proteins are involved in distinct sRNA pathways leading to elimination of germline-specific DNA. Cell Rep. *20*, 505–520. https://doi.org/10.1016/j.celrep.2017.06.050.
11. Drews, F., Karunanithi, S., Götz, U., Marker, S., deWijn, R., Pirritano, M., Rodrigues-Viana, A.M., Jung, M., Gasparoni, G., Schulz, M.H., and Simon, M. (2021). Two Piwis with Ago-like functions silence somatic genes at the chromatin level. RNA Biol. *18*, 757–769. https://doi.org/10.1080/15476286.2021.1991114.
12. Karunanithi, S., Oruganti, V., Marker, S., Rodriguez-Viana, A.M., Drews, F., Pirritano, M., Nordström, K., Simon, M., and Schulz, M.H. (2019). Exogenous RNAi mechanisms contribute to transcriptome adaptation by phased siRNA clusters in Paramecium. Nucleic Acids Res. *47*, 8036–8049. https://doi.org/10.1093/nar/gkz553.
13. Arnaiz, O., Van Dijk, E., Bétermier, M., Lhuillier-Akakpo, M., de Vanssay, A., Duharcourt, S., Sallet, E., Gouzy, J., and Sperling, L. (2017). Improved methods and resources for paramecium genomics: transcription units, gene annotation and gene expression. BMC Genom. *18*, 483. https://doi.org/10.1186/s12864-017-3887-z.
14. Arnaiz, O., Meyer, E., and Sperling, L. (2020). ParameciumDB 2019: integrating genomic data across the genus for functional and evolutionary biology. Nucleic Acids Res. *48*, D599–D605. https://doi.org/10.1093/nar/gkz948.
15. Lepère, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. Nucleic Acids Res. *37*, 903–915. https://doi.org/10.1093/nar/gkn1018.
16. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190. https://doi.org/10.1101/gr.849004.
17. Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, 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. https://doi.org/10.1038/nature05230.
18. Sandoval, P.Y., Swart, E.C., Arambasic, M., and Nowacki, M. (2014). Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. Dev. Cell *28*, 174–188. https://doi.org/10.1016/j.devcel.2013.12.010.
19. Marker, S., Carradec, Q., Tanty, V., Arnaiz, O., and Meyer, E. (2014). A forward genetic screen reveals essential and non-essential RNAi factors in Paramecium tetraurelia. Nucleic Acids Res. *42*, 7268–7280. https://doi.org/10.1093/nar/gku223.
20. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev. *19*, 2837–2848. https://doi.org/10.1101/gad.1370605.
21. Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621–629. https://doi.org/10.1016/j.cell.2005.10.020.
22. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607–620. https://doi.org/10.1016/j.cell.2005.08.044.
23. Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., DeSouza, L.V., Siu, K.W.M., Pearlman, R.E., Gorovsky, M.A., and Mochizuki, K. (2010). The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. Cell *140*, 692–703. https://doi.org/10.1016/j.cell.2010.02.010.
24. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437–1441. https://doi.org/10.1126/science.1102513.
25. Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. *12*, 340–349. https://doi.org/10.1038/nsmb918.
26. Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. Nat. Struct. Mol. Biol. *14*, 347–348. https://doi.org/10.1038/nsmb1218.
27. Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., and Suzuki, T. (2007). The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated. Nat. Struct. Mol. Biol. *14*, 349–350. https://doi.org/10.1038/nsmb1220.

28. Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabildopsis. Curr. Biol. *15*, 1501–1507. https://doi.org/10.1016/j.cub.2005.07.029.

29. Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. Nucleic Acids Res. *34*, 667–675. https://doi.org/10.1093/nar/gkj474.

30. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932–935. https://doi.org/10.1126/science.1107130.

31. Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr. Biol. *17*, 1265–1272. https://doi.org/10.1016/j.cub.2007.06.030.

32. Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of PIWI-interacting RNAs at their 3′ ends. Genes Dev. *21*, 1603–1608. https://doi.org/10.1101/gad.1563607.

33. Kurth, H.M., and Mochizuki, K. (2009). 2′-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. RNA *15*, 675–685. https://doi.org/10.1261/rna.1455509.

34. Lee, S.R., and Collins, K. (2006). Two classes of endogenous small RNAs in Tetrahymena thermophila. Genes Dev. *20*, 28–33. https://doi.org/10.1101/gad.1377006.

35. Couvillion, M.T., Lee, S.R., Hogstad, B., Malone, C.D., Tonkin, L.A., Sachidanandam, R., Hannon, G.J., and Collins, K. (2009). Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of Tetrahymena thermophila. Genes Dev. *23*, 2016–2032. https://doi.org/10.1101/gad.1821209.

36. Götz, U., Marker, S., Cheaib, M., Andresen, K., Shrestha, S., Durai, D.A., Nordström, K.J., Schulz, M.H., and Simon, M. (2016). Two sets of RNAi components are required for heterochromatin formation in trans triggered by truncated transgenes. Nucleic Acids Res. *44*, 5908–5923. https://doi.org/10.1093/nar/gkw267.

37. Rzeszutek, I., Maurer-Alcalá, X.X., and Nowacki, M. (2020). Programmed genome rearrangements in ciliates. Cell. Mol. Life Sci. *77*, 4615–4629. https://doi.org/10.1007/s00018-020-03555-2.

38. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

39. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

40. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

41. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

42. Beisson, J., Bétermier, M., Bré, M.H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Jr., and Sperling, L. (2010). DNA microinjection into the macronucleus of paramecium. Cold Spring Harb. Protoc. *2010*. https://doi.org/10.1101/pdb.prot5364.

43. Beisson, J., Bétermier, M., Bré, M.H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Jr., and Sperling, L. (2010). Silencing specific Paramecium tetraurelia genes by feeding double-stranded RNA. Cold Spring Harb. Protoc. *2010*. https://doi.org/10.1101/pdb.prot5363.

44. Gunasekera, K., Wüthrich, D., Braga-Lagache, S., Heller, M., and Ochsenreiter, T. (2012). Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. BMC Genom. *13*, 556. https://doi.org/10.1186/1471-2164-13-556.

45. Eng, J.K., Hoopmann, M.R., Jahan, T.A., Egertson, J.D., Noble, W.S., and MacCoss, M.J. (2015). A deeper look into comet-implementation and features. J. Am. Soc. Mass Spectrom. *26*, 1865–1874. https://doi.org/10.1007/s13361-015-1179-x.

46. Craig, R., and Beavis, R.C. (2003). A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid Commun. Mass Spectrom. *17*, 2310–2316. https://doi.org/10.1002/rcm.1198.

47. Kim, S., and Pevzner, P.A. (2014). MS-GF plus makes progress towards a universal database search tool for proteomics. Nat. Commun. *5*, 5277. https://doi.org/10.1038/ncomms6277.

48. Tabb, D.L., Fernando, C.G., and Chambers, M.C. (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J. Proteome Res. *6*, 654–661. https://doi.org/10.1021/pr0604054.

49. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat. Methods *14*, 513–520. https://doi.org/10.1038/nmeth.4256.

50. Choi, H., Ghosh, D., and Nesvizhskii, A.I. (2008). Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. J. Proteome Res. *7*, 286–292. https://doi.org/10.1021/pr7006818.

51. Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., et al. (2010). A guided tour of the Trans-Proteomic Pipeline. Proteomics *10*, 1150–1159. https://doi.org/10.1002/pmic.200900375.

52. Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol. Cell. Proteomics *10*. https://doi.org/10.1074/mcp.M111.007690.

53. Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658. https://doi.org/10.1021/ac0341261.

54. Zybailov, B.L., Florens, L., and Washburn, M.P. (2007). Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. Mol. Biosyst. *3*, 354–360. https://doi.org/10.1039/b701483j.

55. Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L. (2010). Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal. Chem. *82*, 2272–2281. https://doi.org/10.1021/ac9023999.

56. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

57. Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K., and Zamore, P.D. (2018). A single mechanism of biogenesis, initiated and directed by PIWI proteins, explains piRNA production in most animals. Mol. Cell *71*, 775–790.e5. https://doi.org/10.1016/j.molcel.2018.08.007.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-HA-Tag Monoclonal antibody	Cell Signaling Technology	Cat# 3724; RRID:AB_1549585
Mouse Anti-Myc-Tag Monoclonal Antibody	Cell Signaling Technology	Cat# 2276; RRID:AB_331783
Anti-HA Affinity Matrix	Roche	Cat# 11815016001; RRID:AB_390914
Bacterial and virus strains		
Klebsiella pneumoniae non-virulent strain	Gift from Eric Meyer (ENS, Paris)	N/A
Escherichia coli strain HT115 (DE3)	Gift from Eric Meyer (ENS, Paris)	HT115
Endura electrocompetent *Escherichia coli*	Lucigen	Cat# 60242-0
Chemicals, peptides, and recombinant proteins		
Wheat grass powder	Pines International, Lawrence, KS	N/A
β-sitosterol	Calbiochem, Millipore	Cat# 567152
Tri Reagent	Sigma-Aldrich	Cat# T9424
Complete EDTA-free protease inhibitor cocktail tablet	Roche	Cat# C762Q78
Critical commercial assays		
Q5 high fidelity DNA polymerase	NEB	Cat# M0491L
GoTaq G2 DNA polymerase	Promega	Cat# M7848
RadPrime DNA Labeling System	Invitrogen	Cat# 18428011
QIAGEN Plasmid Midi Kit	QIAGEN	Cat# 12143
Illumina TruSeq stranded mRNA Kit	Illumina	Cat# 20020594
Illumina TruSeq small RNA kit	Illumina	Cat# RS-200-0012
Deposited data		
Small RNA sequencing from immunoprecipitation: Ptiwi08 Late, Ptiwi08 Late+, Ptiwi08^{HQ} Late, Ptiwi08^{DN} Late	This paper	PRJNA881722 (NCBI)
Small RNA sequencing from EV control and Ptiwi08/14-KD, EV control and Dcr1-KD, EV control and Hen1-KD	This paper	PRJNA881722 (NCBI)
RNA-seq from EV control (2 replicates) and Ptiwi08/14-KD (2 replicates)	This paper	PRJNA881722 (NCBI)
Mass spectrometry data total proteins <25 kDa, EV control and Ptiwi08/14-KD	This paper	PXD038984 (PRIDE)
Small RNA sequencing from WT vegetative stage	Karunanithi et al.[12]	PRJEB25903 (NCBI)
Experimental models: Organisms/strains		
Paramecium tetraurelia, strain 51	Gift from Eric Meyer (ENS, Paris)	N/A
Recombinant DNA		
Ptiwi08-GFP-pGEM T (Dcl5 promoter)	This paper	N/A
Ptiwi08-FlagHA-pGEM T (Dcl5 promoter)	This paper	N/A
Ptiwi08^{HQ}-FlagHA-pGEM T (Dcl5 promoter)	This paper	N/A
Ptiwi08^{DN}-FlagHA-pGEM T (Dcl5 promoter)	This paper	N/A
Ptiwi08-Myc-pGEM T (Dcl5 promoter)	This paper	N/A
Hen1-FlagHA-pGEM T	This paper	N/A
Hen1-Myc-pGEM T	This paper	N/A
Ptiwi08 silencing construct (in L4440)	This paper	N/A
Ptiwi14 silencing construct (in L4440)	This paper	N/A
Hen1 silencing construct (in L4440)	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dcr1 silencing construct (in L4440)	This paper	N/A
Software and algorithms		
Cluster identification and relevant python code	This paper	https://github.com/VCMason/Solberg2023; https://doi.org/10.5281/zenodo.7643023
IGV	Robinson et al.[38]	http://www.broadinstitute.org/igv/
EdgeR	Robinson et al.[39]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
LIMMA	Ritchie et al.[40]	https://bioconductor.org/packages/release/bioc/html/limma.html
Hisat2	Kim et al.[41]	https://github.com/DaehwanKimLab/hisat2
Other		
Amicon Ultra-2 Centrifugation Filter Unit	Millipore	Cat# UFC200324
Ultrafree-MC Centrifugal Filter	Millipore	Cat# UFC30GV25

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents and resources should be directed to the lead contact, Mariusz Nowacki (mariusz.nowacki@unibe.ch).

Materials availability

This study did not generate any unique reagents.

Data and code availability

All the RNA-Seq data have been uploaded to NCBI under the BioProject ID PRJNA881722. The mass spectrometry data has been uploaded to PRIDE under the project accession number PXD038984. Python code used in this manuscript is available on Github https://github.com/VCMason/Solberg2023 (https://doi.org/10.5281/zenodo.7643023). Accession numbers are listed in the key resources table. This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experiments were performed with *Paramecium tetraurelia* strain 51 (mating type seven). Cells were cultured at 27°C in wheat grass powder (WGP) medium (Pines international, Lawrence, KS) bacterized with *Klebsiella pneumoniae* and supplemented with 0.8 mg/mL β-sitosterol (Merck). Autogamy was induced by starvation.

METHOD DETAILS

Constructs

For expression of tagged constructs, the gene of Hen1 including putative 5′ and -3′ UTRs (349 bp upstream and 144 bp downstream), were cloned into pGEM T-easy vector and subsequently N-terminally tagged with FlagHA, GFP or Myc. The gene of Ptiwi08 was cloned between Dcl5 5′ and -3′ UTRs (352 bp upstream and 108 bp downstream) and N-terminally tagged with FlagHA, GFP or Myc. Ptiwi08 slicer mutants were generated by point mutations of the Ptiwi08-FlagHA-pGEM T construct in the following positions: Ptiwi08^{H759Q} T- > A (nt 2197) and Ptiwi08^{D546N} G- > A (nt 1685).

Silencing constructs of Ptiwi08 (1283–2449 bp), Ptiwi14 (1110–2428 bp), Hen1 (4–1041 bp) and Dcr1 (2999–4258 bp) were generated by insertion between two inverted T7 promoters of the L4440 vector. The empty L4440 vector (EV) was used as a negative control.

Cytological staging

Autogamy was induced by starvation and the developmental timepoints assessed by 4,6-diamidino-2-phenylindole (DAPI)-staining and microscopy. A minimum of 100 cells were staged per time point. Seven timepoints were used in this study: Vegetative, Starved, Early (30–40% fragmentation), 100% Fragmentation, Late (Nearly all cells have visible new macronuclei), Late+ (6 h after "Late"), and Post-autogamous.

Macronuclear transformation by microinjection

Constructs for macronuclear transformation were linearized in the backbone of the vector and purified by phenol-chloroform followed by centrifugation using an Ultrafree-MC Centrifugal Filter (Millipore). Transformation was performed by microinjection as described.[42] Successful transformation was confirmed by dot blot following standard protocols.

GFP localization and imaging

300,000 cells were collected, washed twice in 10 mM Tris-HCl (pH 7.4), fixed with 70% Ethanol, and stored at 4°C. For imaging, the cells were washed twice with 1x PBS (pH 7.4) before staining with DAPI. Imaging was performed on a phase-contrast inverted microscope (Axiovert A1, Zeiss) and the ZEN2 software (Zeiss) was used for image processing.

Silencing by feeding

Escherichia coli strain HT115 (DE3) expressing double-stranded RNA against target genes (EV, Dcr1, Ptiwi08, Ptiwi14 and Hen1 (See "constructs")) were used for silencing by feeding as previously described.[43]

Total RNA extraction and sequencing

Total RNA from 600,000 cells was extracted with TRIzol (Sigma-Aldrich) following the TRIzol reagent BD protocol.

mRNA sequencing: Library preparation and sequencing was performed by the Next Generation Sequencing (NGS) Platform at the University of Bern, Switzerland. Library preparation used the illumina TruSeq stranded mRNA kit and paired-end 2x150 bp sequencing was done on a NovaSeq.

Small RNA sequencing: Size selection, library preparation and sequencing of total RNAs were performed by Fasteris (Geneva, Switzerland). Polyacrylamide gel purification was used to size select small RNAs in the range of 17–35 nt, and library preparation used the illumina TruSeq small RNA kit. Single-end 1x75 bp sequencing was performed on a NextSeq.

Northern blot

10 μg of total RNA was resolved in a 1.2% denaturing agarose gel containing 2.2M formaldehyde. The RNA was transferred to a charged nylon membrane (Amersham Hybond-XL, GE Healthcare Life Sciences) by capillary blotting in 20X saline sodium citrate (SSC) buffer. After transfer, the membrane was UV crosslinked twice at 120,000 μJ/cm^2, washed once for 10 min in 2X SSC buffer and pre-hybridized in Church buffer (1% BSA, 0.5M $NaPO_4$ (pH 7.2), 7% SDS, 1mM EDTA) for 2 h at 60°C. Hybridization was performed at 60°C overnight using Church buffer containing the radioactive probe. The following day, the membrane was washed twice with 2X SSC containing 0.1% SDS and exposed on a phosphor screen (Amersham) followed by visualization on a Typhoon FLA 7000 (GE Healthcare).

Probes specific to Hen1 (1–1044 bp) and Dcr1 (4851–5391 bp) were 3′ end labeled with α-P^{33} dATP using the RadPrime DNA Labeling System (Invitrogen). A probe complementary to the 17S rRNA was 5′ end labeled with γ-P^{33} dATP using T4 Polynucleotide Kinase (Thermo Scientific) and served as a loading control.

RNA immunoprecipitation

1.2 million cells were collected and washed twice in 10 mM Tris-HCl (pH 7.4) before flash freezing in liquid nitrogen. Frozen cell pellets were homogenized in 2 mL of lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl; 5 mM $MgCl_2$, 1 mM Dithiothreitol (DTT), 0.5% Sodium deoxycholate, 1% Triton X-100, 1x Protease inhibitor complete tablet (Roche), 2 mM Vanadyl ribonucleoside complex (Sigma-Aldrich), 10% glycerol) in a Dounce homogenizer. After clearing by centrifugation at 13,000 g for 20 min, 1 mL of the cleared lysate was incubated overnight at 4°C with 50 μL of Anti-HA affinity matrix beads (Roche), previously pre-equilibrated thrice with IP buffer (10 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM $MgCl_2$, 5% glycerol, 0.01% NP40). After incubation, the beads were washed five times with IP buffer and the supernatant removed. 10% of the solution was aliquoted for western blotting and boiled with 1x SDS-PAGE buffer. The rest was used for RNA extraction. For RNA extraction, the beads were incubated with 300 μL of 1x Proteinase K buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA pH 8.0, 0.5% SDS) and 1 μL of Proteinase K (20 mg/mL) for 20 min at 42°C. The RNA was extracted by phenol-chloroform and chloroform extraction, followed by ethanol precipitation. Finally, the RNA was resuspended in 10 μL of nuclease-free water. One μl was 5′ end labeled and visualized on a 16% polyacrylamide-7M urea gel (See "small RNA 5′ end labeling"). The rest was used for library preparation and sequencing. Size selection, library preparation and sequencing of the purified small RNAs were performed at Fasteris (Geneva, Switzerland) as described for total small RNAs.

Co-immunoprecipitation

1.2 million cells were collected and washed twice in 10 mM Tris-HCl (pH 7.4). Crosslinked samples were incubated with 1x PBS containing paraformaldehyde for 10 min at room temperature and quenched with 125 mM Glycine for 5 min at room temperature. Cells were washed twice with 1x PBS (pH 7.4) and resuspended in 2 mL of lysis buffer (50mM Tris (pH 8.0), 150mM NaCl, 5mM $MgCl_2$, 1 mM DTT, 1x complete EDTA-free protease inhibitor cocktail tablet (Roche), 1% Triton X-100, 10% glycerol). Lysis was performed using a Branson digital sonifier 250 at 55% amplitude for 15 s, and the lysate cleared for 30 min at 13,000 g and 4°C. 1 mL of the cleared lysate was incubated overnight at 4°C with 50 μL of pre-equilibrated Anti-HA affinity matrix beads (Roche). After incubation, beads were washed five times with IP buffer and boiled for 20 min in 1x SDS-PAGE buffer.

Small RNA 5′ end labeling

One μl of purified small RNAs from RNA immunoprecipitation experiments were 5′-end-labeled by the exchange reaction of T4 polynucleotide kinase (Thermo scientific) with γ-32P dATP (0.4 Mbq). The radioactively labeled RNAs were separated in a 16% polyacrylamide-7M urea gel, exposed on a phosphor screen (Amersham) and scanned on a Typhoon FLA 7000 (GE Healthcare).

Beta elimination

Four μl of purified small RNAs from RNA immunoprecipitation experiments were mixed with 4 ng of an unmodified 17 nt RNA oligo (5′-UCUAGUUCGGCUUACAC-3′). 17.5 μL of borax/boric acid buffer (120 mM, pH 8.6) and 2.5 μL of Sodium Periodate (200 mM) was added to the RNA and it was incubated for 30 min at room temperature in the dark. Two μl of 100% glycerol was added and the reaction incubated for another 10 min at room temperature in the dark. The pH was adjusted with 2 μL of Sodium hydroxide (500 mM) to raise the pH to 9.5 and the reaction was incubated for 90 min at 45°C. The RNA was purified by phenol-chloroform extraction followed by ethanol precipitation and the RNA visualized on a 20% polyacrylamide-7M Urea gel, exposed on a phosphor screen (Amersham) and scanned on a Typhoon FLA 7000 (GE Healthcare).

Western blot

Proteins from immunoprecipitation experiments were separated on SDS-PAGE gels and transferred onto a 0.45 μm nitrocellulose membrane (Amersham Protran, GE Healthcare) by wet transfer. After transfer, the membrane was blocked in blocking buffer (10% skim milk in 1x PBS with 0.1% Tween 20) for 1 h at room temperature and incubated overnight at 4°C with the primary antibody (rabbit anti-HA 3724S or mouse anti-Myc, 2276S, Cell Signaling Technology) at a dilution of 1:1,000 in blocking buffer. The membrane was washed three times for 5 min in PBST and incubated for 1 h at room temperature with the secondary antibody (goat anti-Mouse IgG-HRP, sc-2005, Santa Cruz) at a dilution of 1:10,000 in blocking buffer. Following three washes of 5 min each in PBST, the HRP substrate (Immobilion Forte, Millipore) was applied to the membrane, and it was scanned with the Amersham 600 (GE Healthcare Life Sciences).

Mass spectrometry

Total proteins from EV and Ptiwi08-KD Late were resolved on a 12% SDS-PAGE gel and stained with InstantBlue (Expedeon) for 1 h at room temperature. Sizes under 30kDa was cut into cubes and stored in 20% Ethanol at 4°C until processing. Proteins were in-gel digested as described elsewhere.[44] Digests were loaded onto a pre-column (C18 PepMap 100, 5μm, 100A, 300μm i.d. x 5mm length) at a flow rate of 10 μL/min with solvent C (0.05% TFA in water/acetonitrile 98:2). After loading, peptides were eluted in back flush mode onto a home-made pack C18 CSH Waters column (1.7 μm, 130 Å, 75 μm × 20 cm) by applying a 60-min gradient of 5% acetonitrile to 40% in water, 0.1% formic acid, at a flow rate of 250 nL/min. The column effluent was directly coupled to a Fusion LUMOS mass spectrometer (Thermo Fischer, Bremen; Germany) via a nano-spray ESI source. Data acquisition was made in data dependent mode with precursor ion scans recorded in the orbitrap with resolution of 120'000 (at m/z = 250) parallel to top speed fragment spectra of the most intense precursor ions in the Linear trap for a cycle time of 3 s maximum.

The mass spectrometry data was searched with Comet,[45] Xtandem,[46] MSGF,[47] Myrimatch[48] and MSFragger[49] against the strain 51 *Paramecium tetraurelia* database v2.0.[14] Reverse decoy sequences were concatenated to the database. Search parameters were set to orbitrap/iontrap parameters, namely 10 ppm and 0.4 Da for the peptide and fragment mass tolerance, respectively; cleavage sites were set to N terminal KR, allowing for 3 missed cleavages. Allowed modifications were fixed carbamidomethylation of cysteines, variable oxidation of methionines, deamidation of asparagine and glutamine and acetylation of protein N-termini. Each search was followed by the application of the PeptideProphet[50] tool from the Transproteomics pipeline (TPP),[51] and then by the application of iprophet[52] from TPP in order to combine the search results, which were filtered at the false discovery rate of 0.01; furthermore, the identification was only accepted if at least three of the search engines agreed on the identification. Protein inference was performed with ProteinProphet[53] from TPP. For those protein groups accepted by a false discovery rate filter of 0.01, a Normalized Spectral Abundance Factor (NSAF)[54] was calculated based on the peptide to spectrum match count; shared peptides were accounted for by the method of Zhang et al.[55]

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic methods

Relevant python code is available on Github https://github.com/VCMason/Solberg2023 (https://doi.org/10.5281/zenodo.7643023).

Sequence processing and alignment

Samtools v1.9 was used for all alignment file filtration, sorting, and indexing. RNA sequence quality control was assessed with Fastqc (v0.11.9).[56] Illumina adapters, bad reads, and poor-quality bases were trimmed with Fastp (v0.20.0). Sequences were filtered to 23 nt sRNAs, which represented the vast majority of the Ptiwi08 RIP. All subsequent analyses to identify Ptiwi08 RIP sRNA clusters only analyzed 23 nt RNA. RNA reads were aligned with Hisat2 (v2.1.0).[41] Sequences that aligned to the vector sequence were removed. Ptiwi08 RIP sRNA were aligned to the *Paramecium tetraurelia* strain 51 mac genome v1.0.[14] Annotated features for the *Paramecium tetraurelia* strain 51 mac genome were identified with the gff3 annotation file v2.0.[14]

Cluster identification and characteristics

Reads that aligned to annotated Ptiwi08 and Ptiwi14 were excluded from alignment files with samtools prior to cluster identification and analyses. A sliding window analysis (window size = 100, step = 25) was performed with python script ClusterIdentifierWithStrandedmRNA.py to identify Ptiwi08 sRNA clusters (VCMason Github). All consecutive windows with an average per depth $\geq$ 50 were kept. The first and last coordinates of these consecutive windows were used as the cluster start and end coordinates. Cluster coordinates were then refined by removing all zero depth bases from the start and end of each cluster. IGV was used to visualize the clusters.[38]

Bam files of aligned Ptiwi08 RIP sRNAs were analyzed with python script ClusterIndentifierWithStrandedmRNA.py to collect and calculate cluster summary information: sequence, length, reference GC percent, read GC percent, strandedness, sRNA orientation, mRNA orientation, sRNA normalized count, mRNA normalized count, annotated gene overlap, stranded mRNA overlap, polarity of sRNA to mRNA (sense/antisense). sRNA and mRNA orientation were calculated as the percentage of reads in the forward direction. The cluster was stranded if the forward percent of forward reads in the cluster was $\geq$ 75% or $\leq$ 25%. mRNA libraries were created with the Truseq stranded mRNA library prep procedure. Therefore, the read 2 (R2) sequencing file determined the orientation of the mRNA molecule. mRNA reads were only counted if they mapped in the proper pair, and only the R2 reads were counted. When both sRNA and mRNA were stranded and had opposing orientations, they were considered antisense.

Additionally, sam files of aligned Ptiwi08 RIP sRNAs were analyzed with DoMyRNATasteLikePi.py. This script performed the phasing and ping-pong analyses. These analyses search for known piRNA-like characteristics present in animals as described in Gainetdinov et al. 2018.[57] Phasing calculates the frequency of distances between between 3′ and 5′ ends of sRNAs on the same strand. Ping-pong calculates the frequency of all distances between 5′ ends of sRNAs on opposite strands within 23 bps. Our ping-pong analysis differed from Gainetdinov et al. 2018 because only the lower depth value between the forward and reverse strands was recorded. This ensured a 1 to 1 ratio of forward and reverse reads were compared, and therefore this analysis compared reads between potentially double stranded products. This was important in our analysis as our data was biased to include reads only on a single strand. The number of overlapping sequences with opposing orientations where 5′ to 5′ distances equaled 21 was calculated for each cluster as the dicer signal strength. This was done with DoMyRNATasteLikePi.py and SummarizePingPongSignal.py. Single stranded RNA analyses plot the depth of sequence for sRNAs for both forward and reverse strands.

Iterative read alignment and subtraction

Proportion of all Ptiwi08 sRNA RIP reads that originated from different sources: *Klebsiella pneumoniae*, vector, and *Paramecium tetraurelia* (MAC, IES, and OES) genomes were identified through iterative read alignment with Hisat2 and read subtraction. Ptiwi08 sRNA reads were aligned to the first genome sequence, aligned reads were then removed, the remaining reads were aligned to the next genome, then the aligned reads were removed, and this process repeated until all genomes were completed. Reads that did not align to any genome were labeled unmapped.

Introns vs CDS features

Introns were annotated using AnnotateIntrons.py. This script assumes that the gaps between CDS features in mRNA are introns. Read counts per feature were calculated with SeqsPerFeatureAndFeatureTypeFixed.py. All features were shortened by 6 total bases (3 bases from each feature end) to reduce counting of reads with short overlaps across exons and neighboring introns. This was done by setting the featurebuffer variable to 3. The RPKM of all Ptiwi08 sRNA RIP reads aligning to introns and CDS were calculated. Ptiwi08 sRNA RIP reads that aligned to CDS and introns were compared by calculating one RPKM value for all CDS features and one value for all introns by summarizing the total feature lengths of all CDS or intron features and the number of reads mapping to them respectively.

mRNAseq sequence alignment

mRNA sequences were aligned to the *Paramecium tetraurelia* strain 51 mac genome v1.0. Gene features were taken from the gff3 annotation file v2.0.[14]

Differential expression

Differential expression was calculated using R (v3.6.1) and script Rsubread_Limma_EdgeR.VCM.R. Read counts per feature were calculated with the featureCounts function from the R library Rsubread (v2.0.1) and differential expression was calculated by an empirical Bayes procedure using edgeR (3.28.1) and LIMMA (v3.42.2).[39,40] EdgeR was used to create a DGElist object from the featureCounts object and to remove features with low expression. LIMMA was used to fit a simple linear model contrasting EV and KD samples with lmFit while the empirical bayes procedure used the eBayes function. Genes were differentially expressed if the adjusted p value was less than 0.05. GO terms for each gene were downloaded from BioMart.[14]

Heatmaps were generated with the coolmap function from Limma and clustering by "de pattern", and heatmap input values were calculated using the cpm function from edgeR as log2(CPM) of the raw counts after filtering out low expression clusters with edgeR's filterByExpr function. The library size for each sample was manually set as the total number of alignments in each sample using the lib.size argument in the DGElist function, and specifying keep.lib.sizes = TRUE when filtering. This was done to allow CPM to normalize by the number of reads mapped, and not by the number of reads aligning to the 29 clusters.