Non-Templated Oligo-Adenylation of Small RNAs in the Parasite Entamoeba: Potential Roles for Small RNA Control in Single Celled Eukaryotic Pathogen

Hanbang Zhang  
Stanford University

Gretchen M Ehrenkaufer  
Stanford University

Neil Hall  
Earlham Institute

Upinder Singh (✉ usingh@stanford.edu)  
Stanford University  https://orcid.org/0000-0003-0630-0306

Research article

Keywords: RNAi, small RNA, small RNA sequencing, Argonaute, small RNA oligo-adenylation, Parasite

DOI: https://doi.org/10.21203/rs.3.rs-41465/v1

License: ☛ ✰ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: The RNA interference (RNAi) pathway is a gene regulation mechanism that utilizes small RNA (sRNA) and Argonaute (Ago) proteins to silence target genes. Our previous work identified a functional RNAi pathway in the protozoan parasite *Entamoeba histolytica*, including abundant 27nt antisense sRNA populations derived from the secondary RNAi pathway which associate with *EhAgo2-2* protein. However, there is lack of understanding about sRNAs that are bound to two other *EhAgos* (*EhAgo2-1* and 2-3), and the mechanism of sRNA regulation itself is unclear in this parasite.

Results: In the present study, we sequenced sRNA libraries from both total RNAs and *EhAgo* bound RNAs. We identified a new population of 31nt sRNAs that results from the addition of a non-templated 3-4 adenosine nucleotides at the 3´-end of the 27nt sRNA populations, indicating a non-templated RNA-tailing event in the parasite. We found that both sRNA populations (27nt and 31nt) are unchanged during the development of *E. invadens*. However, we detected an alteration in their relative abundance for the targeted gene in parasites transfected with a trigger-gene silencing construct, indicating that non-templated RNA-tailing is likely a pathway for sRNA turnover when the targeted gene is unable to be silenced in this parasite. In sequencing the sRNAs associating with the three *EhAgo* proteins, we observed that despite distinct cellular localization, all three *EhAgo* sRNA libraries contain 27nt sRNAs with 5´-polyphosphate (5´-polyP) structure and a largely overlapping sRNA repertoire, mainly targeting retrotransposons and a subset of ~226 genes that are endogenously silenced. Furthermore, our data show that 31nt sRNA populations paritally associate with wildtype *EhAgo2-2* but not with its mutant protein (*EhAgo2-2 C-terminal deletion*), indicating an intact RISC is essential for the sRNA modification process.

Conclusion: High-throughput sequencing of sRNA in *Entamoeba* has identified a new population of sRNA with non-templated adenylation modification, which is the first such observation amongst single cell protozoan parasites. Our sRNA sequencing libraries provide the first comprehensive sRNA dataset for all three *Entamoeba Ago* proteins, which can serve as a useful database for the amoeba community.

Background

Thought to be evolved from an ancient anti-viral defense mechanism, RNA interference (RNAi) and its gene regulation pathways are conserved in most eukaryotic organisms [1–3]. RNAi can be triggered by double-stranded RNA (dsRNA), which is processed by Dicer into short interfering RNA (siRNA) duplexes. One strand of the siRNA duplex is then loaded into Ago to form the RNA induced silencing complex (RISC) which then leads to the inactivation of target mRNAs through mechanisms of posttranscriptional or transcriptional gene silencing (PTGS/TGS) [4]. In addition to the above mentioned simplified RNAi pathway, an amplified gene silencing mechanism that involves the activity of RNA-dependent RNA polymerases (RdRPs) has been identified in plants, nematodes and fungi [5, 6]. RdRP participation is triggered by initial primary sRNA pool and its associated RISC. RdRP can either use the cleaved mRNA as template to synthesize dsRNA as a new substrate for Dicer (in plants and fungi), or to synthesize
secondary 5′-triphosphate 22G sRNA de novo (in nematodes). Primary siRNAs in *C. elegans* are derived from initial trigger dsRNA (exogenously) or from antisense 26G siRNAs (endogenously) [5]. In plants, microRNAs (miRNAs) cleave the RNA target which then triggers RdRP activity for dsRNA synthesis; subsequently they are processed into phasi/tasiRNAs [7]. In yeast, initial sRNAs are Dicer-independent primal small RNAs (priRNAs), which bind to Ago and target nascent chromatin-bound transcripts for processing into siRNAs [8].

There are different classes of sRNA including siRNAs, miRNAs, and Piwi-interacting RNAs (piRNAs) [9]. Recently, studies have shown that sRNAs are modified for different cellular functions [10, 11]. Uridylation of siRNAs and piRNAs has been observed in many systems including *Chlamydomonas* [12], *C. elegans* [13], *Arabidopsis* [14], *Drosophila* [15] and mammalian cells [16, 17] indicating a conserved mechanism for the regulation of sRNA turnover among these systems. For miRNAs, mono-adenylation of mature miRNAs is reported to have a stabilizing effect, and miRNA precursor is often found to be modified by uridylation [15].

The protozoan parasite *Entamoeba histolytica* causes amebiasis, a major health concern in developing countries [18, 19]. The parasite has two life stages: a dormant, environmentally resistant cyst form and a proliferative trophozoite form, which is capable of causing invasive disease. Our previous work has identified a functional RNAi pathway in this parasite [20–23]. We found that *E. histolytica* has abundant 27nt sRNAs with 5′-polyP termini, a feature that is seen in the secondary siRNAs in *C. elegans* and other nematode parasites [24, 25]. There are three *EhAgo* proteins: *EhAgo2-1* (EHL_186850), *EhAgo2-2* (EHL_125650), and *EhAgo2-3* (EHL_177170), that have distinct subcellular locations including the nucleus (*EhAgo2-2*), perinuclear ring (*EhAgo2-1, EhAgo2-3*), and cytosol (*EhAgo2-3*) [26]. We have previously demonstrated that PAZ domains are essential for sRNA binding for all three *Ehagos*, and that sRNA binding affects cellular localization of *EhAgo2-1* and *EhAgo2-3* but not *EhAgo2-2* [26]. To better understand the RNAi mechanism(s) in this non-model system, we now want to ask three questions (i) do the three *EhAgos* bind different subpopulations of sRNAs? (ii) what are the natural functions of the RNAi pathway in this parasite? (iii) are sRNAs themselves regulated in *Entamoeba?*

In this report, we have performed high throughput sRNA sequencing for size-fractionated total RNAs and three *EhAgo*-bound sRNAs and demonstrated that two dominant sRNA populations are present in the total RNA fraction: one at size 27nt, the other at size 31nt, with later containing non-templated 3–4 adenosines at the 3′-end, indicating an adenylation modification event of the sRNAs in the protist. We further expanded our sRNA sequencing effort for 31nt populations in the related reptile parasite *Entamoeba invadens*, a model system for *Entamoeba* development, and found that, like the 27nt population [22], they are not changed during development. Using an RNAi-trigger gene silencing approach, we show that relative abundance of the two sRNA populations is reversed when a targeted gene is unable to be silenced, indicating that sRNA modification with oligo-As is likely a pathway for sRNA turnover, potentially leading to degradation. All three *EhAgo* immunoprecipitation (IP) RNA libraries contain antisense 27nt sRNAs with 5′-polyP modifications, and these sRNAs significantly overlap, mainly targeting retrotransposons and subset of ~226 genes that are silenced in this organism, suggesting that
one of the natural functions of sRNAs is to maintain genome stability in this parasite. We further show that sRNAs reads that are in association with *EhAgo2-2* contain some portion of adenylated 31nt sRNAs, but adenylated sRNAs are not bound to the other two *EhAgo* proteins or when the *EhAgo2-2* has a C-terminal deletion. Overall, our study provides the first comprehensive dataset for sRNAs bound to the three *EhAgo* proteins, which can serve as a useful database for the *Entamoeba* community. The finding of sRNAs with adenylation revealed an additional layer of sRNA regulation control and functional diversity of sRNAs in this single celled deep-branching eukaryotic pathogen.

**Results**

Two small RNA populations (27nt and 31nt) are identified in *Entamoeba*

In order to identify the complete spectrum of sRNA species in *Entamoeba*, including those that may have diverse structures or modifications or may be less abundant we decided to extensively explore the endogenous sRNA species and populations in *E. histolytica* by sequencing total RNA fractions (15-45nt) from wild-type *E. histolytica* trophozoites. We fractionated total RNA into two size fractions (15-30nt and 30-45nt), and recovered RNAs from both were cloned by 5’-P independent cloning method (using Tobacco acid pyrophosphatase (TAP) to convert 5’-polyP into 5’-monoP) (Suppl.Table I). Although similar sRNA libraries were previously reported by us, they were on a small-scale using Sanger sequencing or pyrosequencing approaches or were only at the range of 15-30nt [21, 23]. The goal in this study is to allow the current Illumina deep sequencing platform to provide a full account of *Entamoeba* sRNAs, including potential sRNA species have modifications.

The sRNA size distribution of the two libraries (15-30nt and 30-45nt libraries) cloned by TAP method is shown in Fig. 1A. We observed only one sRNA population (a sharp 27nt peak) for the 15-30nt library, which matched with previous results [23]. However, for the 30-45nt library, we identified two sRNA populations (peaks at 27nt and 31nt). The 27nt peak is likely a carry-over of abundant 27nt population, but the peak at 31nt was unexpected and new to us. We characterized and mapped the sRNA sequences from both libraries using a custom data processing pipeline (Suppl. Figure 1). The unique reads were mapped to tRNA and rDNA sequences using Bowtie [27]; the remaining reads were aligned to the amebic LINEs (Long Interspersed Nuclear Elements), genome and transcriptome. This analysis revealed that most reads in 27nt peak can be mapped to the genome; but that sRNA reads in the 31nt peak do not map to the genome (Table I, Fig. 1A). In order to understand why the 31nt sRNAs could not be mapped to the genome, we plotted the nucleotide frequency at each position for the unmapped reads and identified an oligo-A tail prominent at the 3’-end for the 31nt reads, which causes this 31nt sRNA population to not map to the genome (Fig. 1B). In order to map the 31nt sRNA reads, we clipped sequence reads after 27nt position using a custom Python script, and then re-mapped to the genome. Our analysis revealed that these clipped sequences can now map to genome indicating that the non-templated 3–4 As were added to the existing 27nt sRNAs (Table I). Of note, reads mapped to LINEs make up almost 28% of these 31nt sRNAs compared with < 10% in the 27nt peak (non-modified populations), indicating that retrotransposons may be influenced by non-templated oligo-A sRNAs.
Table I. Genomic categories that are mapped by sRNA reads by size-fractionated total RNA libraries (TAP cloning method).

We followed the same mapping strategy used previously [21, 22]. Raw reads from Illumina sequencing were first sorted by barcodes, then linker sequences were removed and unique reads were generated by the uniq tool. The Bowtie alignments (-v1) were used for mapping to tRNAs, rRNAs and repetitive elements, and genome including ORFs.

Libraries were made using 5′-P independent (TAP) cloning methods: as shown, most reads in 15-30nt library can be mapped to genome (84%), however fewer reads in 30-45nt library can be mapped to genome (50%) and they are mostly 27nt sRNAs. This lead to the discovery of non-templated oligo-A sRNA population. The column “3’ end trimmed” is the remapping of these non-mapped reads in 30-45nt library after trimming off oligo-A at 3’-end. Most reads after trimming (72%) can be mapped to genome. Listed are mapped reads with the calculated percentiles shown in parenthesis.

In order to determine if the sRNA species overlap between the 27nt and 31nt populations, we performed alignment analysis of the trimmed 31nt sRNAs directly against the 27nt sRNA reads and found that most (85%) of the trimmed 31nt reads can be mapped to the 27nt reads, indicating a high overlap between two sRNA populations. Consequently, we found that both sRNA populations target the same set of genes, and the number of unique sRNAs mapped to these genes from both datasets are correlated (Suppl. Figure 2A). However, the abundance of individual sRNAs cloned in each population is not well correlated (Suppl. Figure 2B), indicating abundance of individual sRNAs within the two sRNA populations may be regulated differently within the cell.

We then selected a few probes which were cloned in both 27nt and 31nt populations to confirm by Northern blot analysis. Using total RNA with these probes showed the two expected sizes of sRNAs (Fig. 1C). In addition, we found that both sRNA populations are sensitive to capping enzyme but resistant to Terminator enzyme (Fig. 1D), indicating that both populations have a 5′-polyP structure. Taken together, the sequencing data and Northern blot analyses confirm that E. histolytica contains 27nt as well as 31nt sRNAs, and that the latter are formed by addition of 3 or 4 non-templated As being added to existing 27nt sRNAs.

Both small RNA populations (27nt and 31nt) are unchanged during development of Entamoeba invadens

*Entamoeba invadens* is a reptilian parasite that is used as a model parasite to study amebic development *in vitro* [28, 29]. Previously, we sequenced the 27nt sRNA population in *E. invadens* parasites, and mapped these sRNAs to ~700 genes with low expression levels [22]. We also adapted the trigger gene silencing approach to this parasite, and induced specific gene silencing for the targeted genes [30], indicating that the 27nt sRNA populations and silencing mechanism is conserved in these two amoebic species. We and others have shown that many genes are developmentally regulated during encystation and excystation [31, 32], however genes with antisense sRNAs appear to be not developmentally regulated as sequencing and mapping of the 27nt population at four developmental time-points showed identical gene
composition [22]. We first sought to check whether the 31nt population is also present in *E. invadens*. Total RNA samples from (trophozoite, 72hr encysted parasites, and parasites after 8hr excystation) were radioactively labeled and run on a denaturing 15% polyacrylamide gel. Two sRNA bands can be easily detected at 27nt and 31nt sizes (Suppl. Figure 3A), indicating that *E. invadens* has both sRNA populations. To sequence the 31nt sRNA population, we size-fractioned 30-45nt RNA and made sRNA libraries using the TAP method for all three samples. Similar to the observation in *E. histolytica*, the size distribution and mapping features of these libraries all showed 27nt and 31nt peaks, and 31nt peak reads cannot be directly mapped to the genome (Fig. 1E). Nucleotide compositions of the 31nt population clearly show an oligo-A tail (Suppl. Figure 3B). Genome mapping of these three libraries as well as mapping of their tail-clipped sequences are shown in Suppl.Table II. Thus, we conclude that *E. invadens* also contains an sRNA population with non-templated A-tail. Using a similar approach as outlined previously [22], we analyzed the genes that mapped by sRNA from 31nt populations among trophozoite, 72hr encystation, and 8hr excystation libraries; the overlap is significant as shown in Suppl. Figure 3C, indicating that the development process does not affect these genes, matching previous results with the sRNA from 27nt population. In summary, the endogenous genes with antisense sRNAs are “locked” for silencing during development, which is reflected in both 27nt and 31nt populations.

**The relative abundance of two sRNA populations indicates gene silencing efficiency in trigger gene silencing transfectants**

Small RNA modification events have been documented in several model systems [10, 11], including oligouridylation of siRNAs and miRNAs in plants and algae [33], adenylation of miRNAs in human [34], and recently adenylation of siRNAs in yeast [35]. In these systems, the modification either leads to sRNA degradation (uridylation) or miRNA protection (single adenylation). We sought to explore the possibility that sRNA adenylation is a turnover pathway for sRNA in amoeba, similar to adenylation of siRNAs in yeast. We used cell lines of trigger gene silencing transfectants which we previously developed in the lab [36–38] and probed for sRNAs using gene specific probes. Northern in Fig. 2A show that two bands corresponding to 27nt and 31nt sRNA populations can be detected for each targeted gene. We also observed that the relative abundance of two sRNA populations is indicative as to whether or not the targeted gene is silenced: for the 19T-*EhROM1* cell line, in which the *EhROM1* gene is silenced (Fig. 2B), there are much higher levels of the 27nt population than the 31nt population. In contrast, for 19T-*EhAgo2-2* cell line, the Ago2-2 gene is not silenced (Fig. 2B), there is more 31nt population than 27nt population. In addition, we tested 19T-induced gene silencing line for a non-RNAi pathway gene (19T-*Eh136160* (calreticulin precursor, putative)), we observed the similar phenomenon in which the targeted gene is not silenced but with prominent 31nt sRNA populations detected (Fig. 2B). The control (EHI_164300 and EHI_125400) probes used in the Northern demonstrated that endogenous silenced genes have more 27nt population signal. These results may suggest that parasite cells have a homeostasis mechanism to balance the turnover of the 27nt sRNA population. In the cases where a targeted gene is silenced, 27nt sRNAs can be accumulated and used as silencing effector; however, when a targeted gene is not silenced, most 27nt sRNAs will be converted into modified form with oligo-As, possibly leading to further degradation. As indicated by Argonaute PAZ crystal structure, PAZ domain shows a poor binding ability
to the 3’-adenylated RNAs [39]. We have shown that \textit{EhAgo2-2} binds abundant 27nt sRNA populations and that mutating its PAZ domain abolishes sRNA binding [26]. With modification with oligo-As, 31nt sRNAs presumably are in a process of disassociation with \textit{EhAgo2-2}. Loss of protection from binding to Ago may lead to degradation.

**The three \textit{EhAgo} proteins all bind to 27nt sRNAs**

We recently reported that three \textit{E. histolytica} Ago proteins have distinct subcellular localizations and demonstrated that the PAZ domain of each \textit{EhAgo} controls sRNA binding [26]. To further characterize the sRNA populations that bind to each \textit{EhAgo}, we used Myc-tagged \textit{EhAgo} overexpression lines and anti-Myc IP pulldown to isolate RNAs associated with each Ago (Fig. 3A). For \textit{EhAgo2-2}, a distinct 27nt sRNA population was noted, as has been previously published [23]. For \textit{EhAgo2-1}, the sRNAs were much less abundant and seen as a faint smear at 20-30nt range. For \textit{EhAgo2-3}, two sRNA populations at 27nt and 21nt were observed.

The specificity of Myc IP for each Ago was verified by additional controls. First, IP using control beads (anti-HA) showed no signal at the sRNA range when compared with anti-Myc IP for each \textit{EhAgo} (Suppl. Figure 4A). Second, we used Western blot analysis to demonstrate that each \textit{EhAgo} has specific Myc signal at the expected sizes which is absent in the control IP (Fig. 3B). The same membrane blot was stripped and probed using anti-\textit{EhAgo2-2} antibody; this demonstrated that the \textit{EhAgo2-2} was only identified in the \textit{EhAgo2-2} IP but not in the IP of \textit{EhAgo2-1} or \textit{EhAgo2-3}, indicating that each IP is specific without cross-contamination with \textit{EhAgo2-2} (Fig. 3B). Of note, \textit{EhAgo2-2} is the only protein that is abundant enough to be detected in wildtype cell lysates by Western blot analysis, the other two \textit{EhAgos} are at low level of expression which can only be detected from overexpressing cell lines. Hence, we could not easily test the other two Ago proteins for cross-contamination [26]. Given that \textit{EhAgo2-2} is the most abundant Ago protein in \textit{Entamoeba} and has the most abundant population of associated sRNAs, the ability to exclude its potential co-IP in \textit{EhAgo2-1} and \textit{EhAgo2-3} pull-down was important. Previously, we showed that the three \textit{EhAgos} have distinct subcellular locations and that mutations in the PAZ domains can abolish sRNA binding specifically for each Ago [26]. Finally, we demonstrated that the sRNA population bound to each Ago is not affected by various high salt concentrations used in the IP wash (Suppl. Figure 4B), which indicates that each Ago firmly binds the associated sRNA population. Based on these data, we concluded that the sRNA profile identified in Fig. 3A with each \textit{EhAgo} IP is specific to the given \textit{EhAgo} protein being studied.

**The sRNAs bound to the three \textit{EhAgo} proteins have 5’-polyP structure**

We have previously shown that sRNAs bound to \textit{EhAgo2-2} have a 5’-polyP structure [23], a feature similar to the 22G sRNA found in \textit{C. elegans} and \textit{Ascaris} [24, 25]. To determine if sRNAs bound to \textit{EhAgo2-1} and \textit{EhAgo2-3} also have a similar 5’-polyP structure, we performed an RNA capping assay [23, 24]. Using capping treatment, we show that the 27nt sRNAs associated with both \textit{EhAgo2-1} and \textit{EhAgo2-3} shifted in size by one nucleotide; however, the smaller size RNAs (18-24nt) within the same sample were unchanged with the capping assay (Fig. 3C). Overall, the data indicate that 27nt sRNAs that associate with \textit{EhAgo2-1
and EhAgo2-3 have a 5’-polyP structure whereas the lower size sRNAs do not. In order to define the 5’-structure for lower size sRNAs, EhAgo2-3 IP sRNA samples were labeled at the 5’-end using either T4 polynucleotide kinase (PNK) or calf intestinal phosphatase (CIP) + T4 PNK (Suppl. Figure 4C). The signal for PNK labeling can be seen for lower size band but not for upper 27nt band; however, as expected, both upper and lower bands can be seen by CIP + PNK labeling, indicating that the lower size sRNAs likely have 5’-OH structure; thus, these sRNAs may arise from an RNA degradation process.

Our capping assay and 5’-end labeling analysis indicated that the three EhAgos are all loaded with 5’-polyP sRNAs. The sRNA loading to redundant Agos has been seen in other systems. In C. elegans, 5’-polyP 22G sRNAs are loaded into worm-specific WAGOs (18 members) with semi-redundancy [40]. In human cells, all four Ago proteins are loaded with miRNAs and siRNAs in a non-distinguishable manner [41].

Characterization of sRNA populations bound to three EhAgo proteins

For a better understanding of the sRNAs associated with all three EhAgos, we performed high throughput sequencing of sRNA libraries generated from anti-Myc IP RNA samples. We combined three independent biological samples of anti-Myc IP RNAs for each EhAgo line. Using these samples, we made sequencing libraries using two separate enzymatic treatments (TAP and RNA 5’-pyrophosphohydrolase (RppH)) to convert 5’-polyP into 5’-monoP. A total of six sRNA libraries (three with each enzyme treatment) were constructed using a 5’-P independent cloning method (this approach is able to clone both 5’-polyP and 5’-monoP RNA species). All samples are listed with the sequencing depth of total reads and unique reads (Suppl. Table I). The RppH IP libraries were sequenced more deeply than TAP IP libraries, with depth of approximately 2 million reads among three EhAgos, which generated approximately two times more unique reads than TAP IP libraries. We therefore relied on the RppH sequencing libraries for the majority of the analysis presented below; the analysis of TAP sequencing libraries is also provided as Suppl. Table III; it is substantially similar. We also included one important EhAgo2-2 mutant, EhAgo2-2ΔNLS−DR, which lacks a C-terminal putative NLS and the DR-rich motif region. The mutant protein is localized to the cytoplasm, in contrast to the nuclear localization of wild-type protein, and had similar sRNA binding of 27nt sRNA comparable to the wild-type EhAgo2-2 [26].

The overall mapping of Ago bound sRNAs to the Entamoeba genome is in Table II. The percentage of reads that map to tRNA and rRNA are at similar level among three EhAgo libraries (0.5-1% for tRNA; 13–19% for rRNA) and these reads are predominantly in the sense orientation. The tRNA and rRNA reads are inevitably present in almost all sRNA sequencing libraries published to date and they are often considered as partial degradation products from these high abundant RNA species in the cell, with a few exceptions [42, 43]. Of note, the mutant EhAgo2-2ΔNLS−DR IP library had fewer rRNA reads (3.7%) when compared with wildtype EhAgo2-2 IP library (13.3%). This may be due to the localization change of the mutant protein (cytoplasmic rather than nuclear).

Table II. Genomic categories that are mapped by sRNA reads by EhAgo IP libraries.
Genomic mapping strategy is same as in Table I. Listed are mapped reads in each category with the calculated percentiles shown in parentheses. LINE mapped reads are much less in \textit{EhAgo2-2} compared with other two \textit{EhAgo}s and \textit{EhAgo2-2}^{Δ\text{NLS}-\text{DR}}.

The \textit{E. histolytica} genome is highly populated with retrotransposons and repeat elements including LINEs, SINEs (Short Interspersed Nuclear Elements), and EREs (\textit{Entamoeba} Repeat Elements) \cite{44}. There are thousands of copies of \textit{EhLINE}s in the genome, but they are considered “inactive” as genome sequencing has found no single copy of a LINE which has a complete open reading frame (ORF) \cite{45}. Our sequencing datasets for the three \textit{EhAgo} sRNA libraries have few reads that mapped to SINEs and EREs, however there are substantial sRNA reads that mapped to LINEs. Among the three \textit{EhAgo} proteins, \textit{EhAgo2-2} had significantly lower amounts of LINE-derived sRNAs (2.5\% with \textit{EhAgo2-2}; 9.9\% with \textit{EhAgo2-1} and 6.8\% with \textit{EhAgo2-3}) (Table II). For the mutant \textit{EhAgo2-2}^{Δ\text{NLS}-\text{DR}}, we observed a higher percentage of LINEs reads compared to the wild type \textit{EhAgo2-2} (10.7\% vs. 2.5\%).

The largest category (40\%) of reads that mapped to genome belong to ORFs, indicating the second major sources for endogenous sRNAs in \textit{Entamoeba} are derived from gene coding regions. We categorized the genes to which the sRNAs map using both a cutoff (≥ 20 sRNAs map to a gene) and antisense/sense ratio (Antisense (ratio > 2), Mixed (ratio 0.5-2) and Sense (ratio < 2)). As seen in Table III, number of ORFs in the Antisense group is the largest among three categories for all three \textit{EhAgo} proteins and they overlap by a set of 226 ORFs. Both TAP and RppH IP libraries rendered very similar results in terms of sRNA mapped genes, indicating two different sRNA treatments works equally well, and sequencing depth used in this study is sufficient to identify the core ORFs targeted by sRNAs. As seen previously, our data for all three \textit{EhAgo}-bound sRNA libraries further demonstrated that genes with antisense sRNAs have very low expression levels and that the distribution of the antisense reads is biased to the 5’-end of genes (Suppl. Figure 5A and 5B). Lastly, we used the sequences from \textit{EhAgo} IP libraries to determine if \textit{E. histolytica} antisense sRNAs have a “phased” feature. We checked the first 540 bp region of each ORF for mapped sRNA reads under a 27 bp window starting from the ATG. The resulting frequency for each position (1–27) was plotted (Suppl Fig. 6). There is no apparent phased register for antisense sRNA in \textit{Entamoeba}, indicating these sRNAs are not from Dicer processing.

**Table III.** Antisense sRNAs mapped genes overlap among three \textit{EhAgo} IP libraries (TAP and RppH) and size-fractionated total RNA libraries (TAP).

We used a cutoff of ≥ 20 unique sRNAs mapping to a gene. We further divided targeted genes based on the number of AS/S ratio, and formed Antisense (ratio > 2), Mixed (ratio 0.5-2) and Sense (ratio < 2) groups. Listed are number of genes for each group. As shown, both TAP or RppH treatment rendered very similar gene set for three \textit{EhAgo}s. The sequencing depth of RppH IP libraries is 4-fold deeper than TAP IP libraries (Suppl. Table I), indicating that: A: sequencing depth in this study is sufficient to identify the core ORFs targeted by sRNAs; B: both enzymatic treatments are effective. In addition, genes identified by two size-fractionated total RNA libraries that are cloned using 5’-P independent cloning method (TAP) are
listed. In all libraries, the Antisense group consists of the largest number of genes, and these genes significantly overlapped among *EhAgo* IP libraries, as well as total RNA libraries.

On a genome-wide scale, we used the Cuffdiff algorithm [46] to check if there are intragenic regions to which sRNAs from the three *EhAgo* IP libraries map differentially. This is an approach similar to our genome-wide RNA-Seq study for identifying loci with differential mapping of mRNAs [22, 32]. Pair-wise comparisons among the three *EhAgo* libraries identified only a small number of loci with differential mapping of Ago-associated sRNAs. For example, 64 significant differences out of 2,225 loci with mapped sRNAs were identified in the comparison between *EhAgo*2-1 and *EhAgo*2-2, and 51 out of 1,734 loci were identified in the comparison between *EhAgo*2-3 and *EhAgo*2-2. These results again indicate that sRNAs bound to each of the three *EhAgos* have very similar targets throughout the genome. Based on our previous work, this core set of silenced genes remains silent even under stress conditions [22]. We speculate that *E. histolytica* may utilize these *EhAgo* in a redundant manner for silencing of endogenous genes.

**The sRNAs that bind *EhAgo* proteins are 27nt in size and have a 5'-G bias**

For the three Ago-associated sRNA libraries, we determined the size distribution of sRNA cloned from both the TAP and RppH methods and found that they are similar (Fig. 4A and Suppl Fig. 7A), indicating both enzymatic treatments made no difference in converting these 5’-polyP sRNAs for library cloning. The 27nt sRNA peak can be seen in all four *EhAgo* libraries, with a sharp 27nt peak for *EhAgo*2-2 and *EhAgo*2-2\(^\Delta\text{NLS-DR}\). However, smaller size sRNAs are seen in *EhAgo*2-1 and *EhAgo*2-3 libraries by both TAP and RppH methods. This matches with the sRNA profile seen on the sRNA gel (Fig. 3A), where the lower size RNAs have 5’-OH structure and are likely from degradation. In addition, we also checked the size distribution of the total reads (non-unique) for each library (Suppl. Figure 8). There is a prominent peak at 27nt and a very small peak at 21nt in *EhAgo*2-3 IP library, indicating the smaller 21nt RNA band is not cloned efficiently (as expected, likely because of their 5’-OH structure).

We previously observed that the sRNAs bound to *EhAgo*2-2 have a G bias in the 5’-nucleotide position [21, 22]. To check the nucleotide composition of sRNAs in *EhAgo*2-1 and *EhAgo*2-3 IP libraries, we plotted nucleotide frequency at each position for each unique sRNA read and found the 5’-G bias again (Fig. 4B and Suppl. Figure 7B).

As *EhAgo*2-1 and *EhAgo*2-3 have significant numbers of sRNAs in the size range of 18-27nt, we tested if 5’-G bias feature was true for the smaller size sequences in these libraries. We extracted subsets of 23-24nt and 27nt reads from three *EhAgo* libraries and compared their nucleotide composition (Suppl. Figure 9). Both subsets show the 5’-G bias feature indicating that 5’-end of sRNA is likely intact between two sampled 23-24nt and 27nt subsets. We did further mapping analysis of two subsets (23-24nt reads against 27nt reads) using Bowtie and found that the majority of reads in 23-24nt subset align perfectly to reads in 27nt subset (Suppl. Table IV), indicating these smaller reads are likely derived from 27nt sRNA; thus, the 3’ end of the sRNA is more prone to the degradation process than its 5’-end.
In order to determine if the actual sRNA species overlap among the three EhAgo libraries, we performed Bowtie alignment analysis of the EhAgo2-1 and EhAgo2-3 libraries, using EhAgo2-2 dataset as reference. We found that over 70% reads are aligned with reads in the EhAgo2-2 library (Suppl. Table V), indicating the identities of sRNA pool significantly overlap for three EhAgos. We also compared EhAgo2-2△NLS−DR with EhAgo2-2, and the overlap for these two libraries is 75%, indicating that the mutation does not affect sRNA binding to this protein. We had previously noted that the EhAgo2-2△NLS−DR efficiently binds sRNA similar to the wild-type protein [26], and the sequencing now confirms that the EhAgo2-2△NLS−DR mutant does not have an alteration in its associated 27nt sRNA population.

As an added note, the sequencing libraries for TAP EhAgo IP samples were made several years apart from the RppH EhAgo IP samples. Analysis for both libraries rendered very similar results in terms of sRNA mapping features, size distribution profile, 5′-G bias, antisense sRNA mapped genes. Therefore, we think that the highly similar sRNA populations observed in this study among the three EhAgos, despite the different localization, is not due to potential co-IP cross-contamination of EhAgo2-2 but rather is the true reflection on each Ago, as confirmed by the IP controls and Western blot analysis.

Taken together, we concluded that all three EhAgos bind to sRNA populations with significant overlap, mainly targeting retrotransposons and a core set of ~226 genes that are silenced in this organism. These sRNAs have a 5′-polyP structure and are not phased; thus, they are likely from the secondary sRNA pathway that involves RdRP activity. As mentioned earlier, C. elegans has worm-specific WAGOs which are all semi-redundantly loaded with secondary 5′-polyP 22G sRNAs. The C. elegans 22G RNAs are important for germline maintenance as they map to ~50% of the annotated coding gene in the genome which are from both silent and expressed loci [40, 47].

**31nt sRNAs are weakly associated with EhAgo2-2**

Our sequencing data for total RNAs show that E. histolytica has second sRNA size peak due to the 3–4 adenine(s) non-templated additions at the 3′-end of 27nt sRNAs. In order to check if this 31nt population is associated with any specific EhAgo, we analyzed non-mapped sRNA reads for each of the Ago IP libraries. The size distribution of the non-mapped reads showed the 31nt peak only present in EhAgo2-2 but not for the other two EhAgo proteins or EhAgo2-2△NLS−DR (Fig. 4C). Additionally, for EhAgo2-2, the nucleotide distribution analysis for non-mapped reads show 5′-G bias for the first nucleotide, and a string of 3 or 4 As were identified at the 3′-end (Fig. 4D). These results may indicate that EhAgo2-2 is the protein complex site that is involved in adenylation of sRNA and that its C-terminal DR-motif domain or associated proteins may be necessary for the adenylation modification to occur.

The EhAgo2-2 has an unusual DR-rich motif which controls the localization of this protein to the nucleus; two deletion mutants including EhAgo2-2△NLS−DR did not alter sRNA binding but caused protein localization change to the cytoplasm [26]. Our work has suggested that EhAgo2-2 can actively transport sRNAs from cytoplasm to the nucleus, where it can target nascent RNA transcripts and build repressive chromatin marks [26, 48]. One question that remains unanswered is the cellular site for the sRNA
adenylation event in amoeba. In order to address this issue, we performed cell fractionation for nuclear and cytoplasmic lysates based on previously published methods [49]. Two Myc-tagged overexpression lines of EhAgo2-2 and EhAgo2-2△NLS−DR were fractioned side-by-side, anti-Myc IP pulldown experiments were performed, RNAs were isolated for both fractions and their sRNA profile characterized (Suppl. Figure 10). As expected, 27nt sRNAs were significantly depleted in the nuclear fraction for EhAgo2-2△NLS−DR due to the protein localization to the cytoplasm. In contrast, 27nt sRNAs were found at almost equal levels in both nuclear and cytoplasmic fractions for EhAgo2-2. We then sequenced sRNA libraries for EhAgo2-2 nuclear and cytoplasmic IP samples. Sequence mapping revealed that most reads of 27nt sRNAs from both libraries can be mapped to the genome, and that they have similar percentage for every mapped genomic category except that the cytoplasmic IP library has a larger percentage of non-mapped reads (19%) than the nuclear IP library (11.8%) (Suppl. Table VI). The size distribution of the non-mapped reads was analyzed (Fig. 4E) and demonstrated that the 31nt peak is present in EhAgo2-2 cytoplasmic IP but not in the EhAgo2-2 nuclear IP. The A-tail was identified at the 3’-end for cytoplasmic IP by nucleotide distribution analysis (Fig. 4F).

Our sequencing data for both total and Ago-bound RNAs show that E. histolytica has secondary sRNA populations that are modified with non-templated 3–4 adenine(s) at the 3′-end. Northern blot analysis confirmed both populations for both endogenous silenced genes as well as 19-Trigger induced gene silencing. The adenine(s) non-templated sRNAs are found in part to be associated with EhAgo2-2 and the modification processing appears to occur in the cytoplasm. Our finding of sRNA with the adenylation modification is the first report of a sRNA modification in pathogenic protists and adds to the complexity of the sRNA repertoire in non-model organisms.

Summarizing the findings in this study as well as our previous work [20, 23, 26, 36, 48] (Fig. 5), we see that E. histolytica has abundant 5′-polyP 27nt sRNAs that are loaded to three EhAgo proteins in a non-distinguishable manner. The endogenous sRNAs are mainly derived from LINEs and a core set of ~ 226 gene loci with sRNA features consistent with these being RdRP products (they are 5´-G biased, mostly antisense, and are not in-phase). We identified a second sRNA populations at 31nt that are due to the modification of 27nt sRNAs at 3´-end with 3 or 4 As and these modified sRNAs are found in partial association with EhAgo2-2 while the intact RISC is in cytoplasm. The genetic roles corresponding these non-templated sRNAs await further study.

Discussion

The RNAi pathway regulates gene expression and is mediated by Ago proteins and their bound sRNAs. Here, we performed high throughput sRNA sequencing to uncover sRNA species in E. histolytica, focusing on both total sRNA and sRNAs that associate with the three amebic Ago proteins. Our sRNA sequencing uncovered two sRNA populations in the parasite: 27nt sRNA with 5´-polyP termini and a 5´-G bias; a second population of 31nt sRNAs that are the same sequence as the 27nt sRNAs, have the same 5´-structure but are modified at the 3´-end by adenylation for non-templated 3–4 As. We confirmed both populations by Northern blot analysis, and sequencing of 31nt population indicated that they are
unchanged during development for *Entamoeba invadens*. The 31nt population associates with *EhAgo2-2* and can be recovered from a sRNA library generated from cytoplasmic but not nuclear fractions. Despite each *Entamoeba* Ago protein having distinct cellular localization, small RNA sequencing demonstrates that there is a core set of antisense sRNAs that bind to all three Ago proteins and which largely target retrotransposons and a core set of ~226 genes marked for gene silencing.

Our previous limited sRNA sequencing on libraries using either size-selected total RNAs or Ago2-2 IP RNAs [21, 22] revealed sRNA with features including 5′-polyP and a strong 5′-G bias. The sRNA sequencing for *EhAgo2-1* and *EhAgo2-3* presented in this study provides a full account of Ago-bound sRNAs in *E. histolytica*. We identified a similar pool of sRNAs among all three *EhAgos*, indicating that the three *EhAgos* are probably loaded from the same pool of secondary sRNAs (Fig. 5). Differential loading of sRNAs into different Agos has been reported in model systems [50, 51]. The 5′-terminal nucleotide of the sRNA is often used as selective identity for AGO loading, as demonstrated in plants, where AGO1 has preference for 5′-uridine, but AGO2 and AGO4 for 5′-adenosine [52]. In *C. elegans*, primary and secondary sRNAs are generated in a sequential mode and are bound to different type of Ago proteins based on sRNA sizes and 5′-end structures [24, 53, 54]. However, the four human Ago proteins mostly bind an overlapping set of sRNAs [55, 56] indicating redundant functions of the four Agos, although recently some differences in miRNA loading and distribution were reported using deep sequencing methods [41, 57]. Our sequencing results revealed the main targets of the sRNAs bound to the three *EhAgo* proteins are retrotransposons and a core set of silenced genes, indicating that the main functions of *E. histolytica* RNAi are to control genome stability and to silence genes which can be related to strain difference or virulence factors [21].

Ago-bound sRNAs are themselves regulated in the cell. Uridylation of siRNAs has been reported in several model organisms for their degradation; however single adenylation of miRNA has been suggested to stabilize them [10, 11]. In yeast, two RNA-tailing events were reported: for the non-coding RNAs (rRNA, tRNAs and small nuclear and nucleolar RNAs), nuclear TRAMP complex can add oligo-As (4–5 As), which in turn signal the recruitment of exosome for RNA degradation [58]; For siRNAs, it has been reported that addition of 1–2 non-templated nucleotides (adenines or uridines) to the 3′-end leads to sRNA elimination [35]. We have shown that the relative abundance of two sRNA populations is indicative as to whether or not the targeted gene is silenced: there is more 31nt than 27nt populations when a gene is not silenced. This may suggest that parasites have a way to convert 27nt sRNA into 31nt sRNA population, which ultimately leads to their degradation. Further studies to identify the biological conditions and genetic pathways that contribute to this sRNA modification will be important to pursue.

**Conclusion**

In conclusion, our data provide the first comprehensive dataset of the three *E. histolytica* Ago-bound sRNA populations. We identified that endogenous sRNAs arise from retrotransposons and a core set ~226 genes marked for gene silencing. The sRNA adenylation modification event is the first to be
discovered in the parasite. Future studies on genetic roles and identification of the protein complex for adenylation will give more insights on the biology of amebic sRNA regulation.

Methods

Parasite culture, plasmids and cell lines

The wild type (WT) *E. histolytica* trophozoites (HM-1:IMSS) were grown axenically under standard conditions as described previously [23, 59]. Three *EhAgo* wildtype cell lines for overexpressing N-terminal Myc-tagged protein were constructed as in [26]. *EhAgo2-2ΔNLS-DR* is a mutant cell line of *EhAgo2-2* that lacks NLS-DR-rich motif region as reported in [26]. The trigger gene silencing plasmids (19T-**EhROM1** (EHI_197460), 19T-*EhAgo2-2* (EHI_125650)) are from previous studies [37, 38]; new transfecant cell lines were made using these plasmids for this study. For 19T-**Eh136160**, the full-length gene was amplified from genomic DNA and cloned into trigger gene silencing construct using Smal and XhoI sites, as reported in [37]. All parasite transfecant lines were maintained at 6 µg/ml G418. *E. invadens* strain IP-1 was cultured in LYI-S-2 at 25 °C as previously described [32]. We followed previously published methods for encystation and excystation [22].

Cell lysis, cytoplasmic and nuclear enriched fractions and IP

The cell lysates were made using either WT cells or transfected cells. The basic lysis buffer contains 20 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 10% (v/v) glycerol and 50 mM NaCl. We made the complete lysis buffer by adding IGEPAL CA-630 (equivalent of NP-40) at 0.5% (v/v) plus 1 mM NaF, 1 mM DTT, 1 mM PMSF and HALT EDTA free protease inhibitors (100x stock diluted to 2x)(Thermo Scientific) and RNase inhibitor (1 unit/ml). The cells were lysed on ice for 15 min, and centrifuged at max speed for 30 min at bench top centrifuge at 4 °C, and supernatant were saved as crude lysate at -80 °C.

Cytoplasmic and nuclear enriched fractions were isolated based on published methods [49, 60] with some changes: three T25 flasks of *Entamoeba* cells were collected and resuspended in 3 ml buffer A (10 mM Tris-HCl, pH 7.5, 3 mM MgCl₂, 10 mM NaCl) with protease inhibitors 1 µm leupeptin, 1 µm E-64-d, and 1 × HALT protease inhibitor mixture (Thermo Scientific) and incubated on ice for 15 min. 0.5% IGEPAL was then added, mixed briefly and centrifuged for 10 min at 2000 ×*g* at 4 °C, the supernatant was collected and adjusted NaCl to 150 mM and stored at - 80 °C (the cytoplasmic fraction). The pellet was then washed in 1000 µl of buffer A (without IGEPAL), followed by centrifugation for 10 min at 500 ×*g* and 4 °C. The pellet was resuspended in 500 µl of buffer C (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 3 mM MgCl₂, 10% glycerol, and 0.5% IGEPAL) supplemented with the same protease inhibitor mixture and put through a 27G needle through 5 times to break the nuclei. Samples were centrifuged (20 min, 4 °C, max speed). The supernatant was collected and stored at - 80 °C (the nuclear fraction).

The IP protocol is identical to [26], which is an adaptation from [61]. Both anti-Myc and anti-HA beads (Thermo Scientific) were used. Typically, 20µ l packed beads were used for each IP. A total of 250µ l crude lysate was diluted with equal volume of complete lysis buffer giving a protein concentration around (1–
2µ g/ml). The IP mixture was rotated for 2hr at 4 °C. After binding, the beads were washed 6 times at 4 °C (5 min each) using low stringency wash (the basic lysis buffer with 0.1% (v/v) Tween-20 and 0.1% (v/v) NP-40, PMSF and 0.5X HALT EDTA free protease inhibitors). In order to assess the stringency of the protein-RNA interaction for certain experiments, the last three washes (5 min each) had varied salt concentrations with NaCl at low (50 mM), medium (250 mM), and high (500 mM). After the final wash step, beads were centrifuged at max speed for 1 min at bench top centrifuge at 4 °C, and the bead pellet was used for RNA preparation or for protein Western blot.

**RNA isolation, labeling and capping assays**

For IP RNA, 300 ul TRizol (Invitrogen) reagent was added to IP beads, and total RNA was isolated using standard protocol. For total RNA/sRNA enriched RNA, we used the mirVana kit (Thermo Scientific) according to the manufacturer’s protocol. The procedure for capping assays are similar as in [23]. NEB Vaccinia capping system (New England Biolabs) was used for the capping assay. A reaction volume (20µ l) contains 2µ l of IP RNA or 10µ g sRNA-enriched RNA, 1x Capping buffer, 1µ l vaccinia capping enzyme and 1 ul GTP (10 mM) and 1µ l SAM (2 mM). The reaction was incubated for 0.5 hr at 37 °C. The capped RNAs were then extracted with acid phenol:chloroform and labeled at 3’ termini by T4 RNA Ligase (New England Biolabs) using α-[32P]-pCp. For 5’-termini labeling, KinaseMax kit (Thermo Scientific) was used, either PNK or CIP + PNK reaction was set up using γ-[32P]-ATP following protocol in the kit. For the Terminator assay, RNA sample (20µ g) was incubated with Terminator enzyme (30 °C for 1 hour). A “spiked in” control (a pre-labeled radioactive 5’-monoP RNA) was included to the reaction. All RNA samples were resolved on a denaturing 7M urea-containing 15% polyacrylamide gel and radioactive signal was detected using a Phosphor screen and imaged on a Personal Molecular Imager (Bio-Rad).

**Northern blot analysis**

Northern protocol was performed as in [23]. sRNA-enriched RNA sample (20 µg or 50 µg) was separated on a denaturing 15% polyacrylamide gel and transferred to a membrane (Amersham™ Hybond™ -N + Membrane, GE Healthcare). Probe DNA (Suppl. Table VII) was 5´-end labeled by PNK reaction using γ-[32P]-ATP. The 32P-labeled DNA probe was then hybridized with membrane in perfectHyb buffer (Sigma) overnight at 37 °C. The membrane was washed using low (2X SSC, 0.1% SDS at RT for 15 min) and medium (1X SSC, 0.1% SDS at 37 °C for 15 min) stringency conditions, and radioactive signal was detected using a Phosphor screen and imaged on a Personal Molecular Imager (Bio-Rad).

**Size fractionation of total RNA**

We used 60 µg sRNA-enriched RNA sample (from WT cells) for the size fractionation experiments. RNA was loaded onto a denaturing 15% polyacrylamide gel, and separation of RNA was monitored by small RNA ladder. The RNA gel was stained with SYBR gold (Thermo fisher) and visualized under UV, the gel sections corresponding to the desired sRNA size range were cut out and minced into small parts for overnight extraction using buffer (20 mM Tris, pH 8.0, 1 mM EDTA, 0.3M NH₄Acetate, 0.05% SDS). Extracted RNAs were then precipitated by isopropanol and resuspended in 20µ l water.
Library construction and sequencing

For *Eh*Ago IP libraries including *Eh*Ago2-2 nuclear and cytoplasmic IP samples, we used 5′-P independent cloning method based on either TAP (Epicentre, this product was discontinued in 2014) or RppH (NEB). The pooled IP RNAs were treated under suggested enzyme conditions at 37 °C for 20 min and extracted by acid phenol:chloroform. For all size-fractionated RNA libraries, 5′-P independent cloning method (TAP) were used.

We followed a similar procedure as in [22]. All RNA libraries were made using the NEBNext multiplex small RNA library prep set for Illumina (NEB#E7300S) following the manufacturer’s protocol. After cDNA generation, we incorporated barcode primers to each library by PCR (10–12 cycles). We quantified samples using Nanodrop and pooled for Illumina sequencing using the MiSeq platform.

Bioinformatics analysis

We used the same data processing pipelines as described previously [21, 22]. Raw reads were first processed to remove barcodes. We then mapped sequences using unique reads (processed using Unix uniq command) to *E. histolytica* tRNA, rDNA, repetitive elements, transcripts and genome, using Bowtie v1.2.2 (bowtie-bio.sourceforge.net) with the parameters: -v1 --all. Amebic sequences were obtained from amoebadb.org. The length profile and nucleotide distribution at each position were determined as previously reported [21], using the R package ShortRead [62]. To identify genes targeted by sRNAs, we used ≥ 20 sRNA reads as cutoff as previously described [22]. For sRNA phasing analysis, we followed the concept used in [63]. Three *Eh*Ago IP libraries were used, the nucleotide position of the start of each antisense sRNA was obtained from Bowtie ORF alignment files. To simplify the question, we checked only the first 540 bp region of each ORF. The frequency of alignment of the first nucleotide of sRNAs to each position was calculated, and a custom R script was used to determine the count of reads at each position within a 27 bp window starting from the ATG. The resulting frequency for each position (1–27) was plotted.

Abbreviations

RNAi
RNA interference
sRNA
small RNA
Ago
Argonaute
dsRNA
double-stranded RNA
siRNA
short interfering RNA
RISC
RNA induced silencing complex
PTGS/TGS
posttranscriptional / transcriptional gene silencing
RdRPs
RNA-dependent RNA polymerases
miRNAs
microRNAs
priRNAs
primal small RNAs
piRNAs
Piwi-interacting RNAs
TAP
Tobacco acid pyrophosphatase
RppH
RNA 5´-pyrophosphohydrolase
5´-polyP
5´-polyphosphate
PNK
polynucleotide kinase
CIP
calf intestinal phosphatase
LINEs
long interspersed nuclear elements
SINEs
short interspersed nuclear elements
EREs
Entamoeba Repeat Elements
ORF
open reading frame
WT
wild type

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable
Availability of data and materials

The datasets of sRNA libraries generated and analysed in the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by funds to U.S. from the National Institutes of Health (R01 AI121084).

Authors' Contributions

The study was conceived and designed by HZ, US. HZ performed experiments, data analysis and drafted the manuscript. GE performed bioinformatics analysis and gave input on the manuscript. NH performed the small RNA sequencing run. US supervised the project and edited the manuscript. All authors reviewed and accepted the final version of the manuscript.

Acknowledgements

We thank all members of the Singh lab for scientific discussions.

Corresponding author

Correspondence to US.

References

1. Agrawal N, Dasaradhi PV, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK. RNA interference: biology, mechanism, and applications. Microbiol Mol Biol Rev. 2003;67(4):657–85.
2. Cerutti H, Casas-Mollano JA. On the origin and functions of RNA-mediated silencing: from protists to man. Current genetics. 2006;50(2):81–99.
3. Obbard DJ, Gordon KH, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci. 2009;364(1513):99–115.
4. Ketting RF. The many faces of RNAi. Dev Cell. 2011;20(2):148–61.
5. Zhang C, Ruvkun G. New insights into siRNA amplification and RNAi. RNA Biol. 2012;9(8):1045–9.
6. Vazquez F, Hohn T. Biogenesis and Biological Activity of Secondary siRNAs in Plants. Scientifica (Cairo). 2013;2013:783253.
7. Schwab R, Voïnnet O. RNA silencing amplification in plants: size matters. Proc Natl Acad Sci USA. 2010;107(34):14945–6.
8. Halic M, Moazed D. Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. Cell. 2010;140(4):504–16.

9. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nature reviews. 2009;10(2):94–108.

10. Kim YK, Heo I, Kim VN. Modifications of small RNAs and their associated proteins. Cell. 2010;143(5):703–9.

11. Song J, Song J, Mo B, Chen X. Uridylation and adenylation of RNAs. Sci China Life Sci. 2015;58(11):1057–66.

12. Ibrahim F, Rymarquis LA, Kim EJ, Becker J, Balassa E, Green PJ, Cerutti H. Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. Proc Natl Acad Sci USA. 2010;107(8):3906–11.

13. Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 mediates the terminal uridylation of let-7 precursor microRNA. Mol Cell. 2008;32(2):276–84.

14. Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, Meyers BC, Mo B, Chen X. The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. Curr Biol. 2012;22(8):689–94.

15. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y, et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome research. 2010;20(10):1398–410.

16. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, Wolf DA, Mizgerd JP. Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nat Cell Biol. 2009;11(9):1157–63.

17. Thornton JE, Du P, Jing L, Sjekloca L, Lin S, Grossi E, Sliz P, Zon LI, Gregory RI: Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). Nucleic acids research 2014, 42(18):11777–11791.

18. Haque R, Mondal D, Kirpatrick BD, Akther S, Farr BM, Sack RB, Petri WA Jr. Epidemiologic and clinical characteristics of acute diarrhea with emphasis on Entamoeba histolytica infections in preschool children in an urban slum of Dhaka, Bangladesh. Am J Trop Med Hyg. 2003;69(4):398–405.

19. Stanley SL. Jr: Amoebiasis Lancet. 2003;361(9362):1025–34.

20. Zhang H, Alramini H, Tran V, Singh U. Nucleus-localized antisense small RNAs with 5’-polyphosphate termini regulate long term transcriptional gene silencing in Entamoeba histolytica G3 strain. J Biol Chem. 2011;286(52):44467–79.

21. Zhang H, Ehrenkaufer GM, Hall N, Singh U. Small RNA pyrosequencing in the protozoan parasite Entamoeba histolytica reveals strain-specific small RNAs that target virulence genes. BMC Genom. 2013;14:53.

22. Zhang H, Ehrenkaufer GM, Manna D, Hall N, Singh U. High Throughput Sequencing of Entamoeba 27nt Small RNA Population Reveals Role in Permanent Gene Silencing But No Effect on Regulating
Gene Expression Changes during Stage Conversion, Oxidative, or Heat Shock Stress. PLoS One. 2015;10(8):e0134481.

23. Zhang H, Ehrenkaufer GM, Pompey JM, Hackney JA, Singh U. Small RNAs with 5’-polyphosphate termini associate with a Piwi-related protein and regulate gene expression in the single-celled eukaryote Entamoeba histolytica. PLoS pathogens. 2008;4(11):e1000219.

24. Pak J, Fire A. Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science. 2007;315(5809):241–4.

25. Wang J, Czech B, Crunk A, Wallace A, Mitreva M, Hannon GJ, Davis RE. Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome research. 2011;21(9):1462–77.

26. Zhang H, Tran V, Manna D, Ehrenkaufer G, Singh U: Functional Characterization of Entamoeba histolytica Argonaute Proteins Reveals a Repetitive DR-Rich Motif Region That Controls Nuclear Localization. mSphere 2019, 4(5).

27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

28. Mitra BN, Pradel G, Frevert U, Eichinger D. Compounds of the upper gastrointestinal tract induce rapid and efficient excystation of Entamoeba invadens. International journal for parasitology. 2010;40(6):751–60.

29. Vazquezdelara-Cisneros LG, Arroyo-Begovich A. Induction of encystation of Entamoeba invadens by removal of glucose from the culture medium. J Parasitol. 1984;70(5):629–33.

30. Suresh S, Ehrenkaufer G, Zhang H, Singh U. Development of RNA Interference Trigger-Mediated Gene Silencing in Entamoeba invadens. Infect Immun. 2016;84(4):964–75.

31. De Cadiz AE, Jeelani G, Nakada-Tsukui K, Caler E, Nozaki T. Transcriptome analysis of encystation in Entamoeba invadens. PLoS One. 2013;8(9):e74840.

32. Ehrenkaufer GM, Weedall GD, Williams D, Lorenzi HA, Caler E, Hall N, Singh U. The genome and transcriptome of the enteric parasite Entamoeba invadens, a model for encystation. Genome Biol. 2013;14(7):R77.

33. Cerutti H, Ibrahim F. Turnover of mature miRNAs and siRNAs in plants and algae. Adv Exp Med Biol. 2010;700:124–39.

34. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S, Baba T, Suzuki T: Selective stabilization of mammalian microRNAs by 3’adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes & development 2009, 23(4):433–438.

35. Pisacane P, Halic M. Tailing and degradation of Argonaute-bound small RNAs protect the genome from uncontrolled RNAi. Nat Commun. 2017;8:15332.

36. Morf L, Pearson RJ, Wang AS, Singh U. Robust gene silencing mediated by antisense small RNAs in the pathogenic protist Entamoeba histolytica. Nucleic acids research. 2013;41(20):9424–37.
37. Pompey JM, Morf L, Singh U. RNAi pathway genes are resistant to small RNA mediated gene silencing in the protozoan parasite Entamoeba histolytica. PLoS One. 2014;9(9):e106477.

38. Khalil MI, Foda BM, Suresh S, Singh U. Technical advances in trigger-induced RNA interference gene silencing in the parasite Entamoeba histolytica. International journal for parasitology. 2016;46(3):205–12.

39. Ma JB, Ye K, Patel DJ. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature. 2004;429(6989):318–22.

40. Gu W, Shirayama M, Conte D Jr, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, et al. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell. 2009;36(2):231–44.

41. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Suzuki H, Hayashizaki Y, Daub CO. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol. 2011;8(1):158–77.

42. Wang Y, Li H, Sun Q, Yao Y. Characterization of Small RNAs Derived from tRNAs, rRNAs and snoRNAs and Their Response to Heat Stress in Wheat Seedlings. PLoS One. 2016;11(3):e0150933.

43. Martinez G. tRNA-derived small RNAs: New players in genome protection against retrotransposons. RNA Biol. 2018;15(2):170–5.

44. Bakre AA, Rawal K, Ramaswamy R, Bhattacharya A, Bhattacharya S. The LINEs and SINEs of Entamoeba histolytica: comparative analysis and genomic distribution. Experimental parasitology. 2005;110(3):207–13.

45. Lorenzi H, Thiagarajan M, Haas B, Wortman J, Hall N, Caler E. Genome wide survey, discovery and evolution of repetitive elements in three Entamoeba species. BMC Genom. 2008;9:595.

46. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-sEq. Nat Biotechnol. 2013;31(1):46–53.

47. Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, et al. The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell. 2009;139(1):123–34.

48. Foda BM, Singh U. Dimethylated H3K27 Is a Repressive Epigenetic Histone Mark in the Protist Entamoeba histolytica and Is Significantly Enriched in Genes Silenced via the RNAi Pathway. J Biol Chem. 2015;290(34):21114–30.

49. Pearson RJ, Morf L, Singh U. Regulation of H2O2 stress-responsive genes through a novel transcription factor in the protozoan pathogen Entamoeba histolytica. J Biol Chem. 2013;288(6):4462–74.

50. Goh E, Okamura K. Hidden sequence specificity in loading of single-stranded RNAs onto Drosophila Argonautes. Nucleic acids research. 2019;47(6):3101–16.

51. Czech B, Hannon GJ. Small RNA sorting: matchmaking for Argonautes. Nature reviews. 2011;12(1):19–31.
52. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, et al. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell. 2008;133(1):116–27.
53. Pak J, Maniar JM, Mello CC, Fire A. Protection from Feed-Forward Amplification in an Amplified RNAi Mechanism. Cell. 2012;151(4):885–99.
54. Billi AC, Fischer SE, Kim JK. Endogenous RNAi pathways in C. elegans. WormBook 2014:1–49.
55. Su H, Trombly MI, Chen J, Wang X. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. Genes Dev. 2009;23(3):304–17.
56. Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci USA. 2008;105(23):7964–9.
57. Liao JY, Ma LM, Guo YH, Zhang YC, Zhou H, Shao P, Chen YQ, Qu LH. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. PLoS One. 2010;5(5):e10563.
58. Schmidt K, Butler JS. Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. Wiley Interdiscip Rev RNA. 2013;4(2):217–31.
59. Diamond LS, Harlow DR, Cunnick CC. A new medium for the axenic cultivation of Entamoeba histolytica and other Entamoeba. Trans R Soc Trop Med Hyg. 1978;72(4):431–2.
60. Gagnon KT, Li L, Janowski BA, Corey DR. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. Nat Protoc. 2014;9(9):2045–60.
61. Lee SR, Collins K. Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. Nat Struct Mol Biol. 2007;14(7):604–10.
62. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics. 2009;25(19):2607–8.
63. Axtell MJ. A method to discover phased siRNA loci. Methods Mol Biol. 2010;592:59–70.
64. Ehrenkaufer GM, Haque R, Hackney JA, Eichinger DJ, Singh U. Identification of developmentally regulated genes in Entamoeba histolytica: insights into mechanisms of stage conversion in a protozoan parasite. Cell Microbiol 2007.