scientific reports

Check for updates

OPEN

Small non-coding RNAs have predicted roles in reproductive biology and transposable element regulation in the parasitic worm *Strongyloides venezuelensis*

Dominika Lastik[1], Asuka Kounosu[2], Mehmet Dayi[3], Akemi Yoshida[4], Akito Fujihira[2], Kieran Reynolds[1], Vicky L. Hunt[1]✉ & Taisei Kikuchi[2]✉

The gastrointestinal parasitic nematode *Strongyloides* spp. has a unique life cycle that alternates between a parasitic generation that reproduces through mitotic parthenogenesis and a dioecious free-living sexually reproducing generation. Adult females from these two generations are genetically identical, making them an informative model to identify molecular differences between parasitic and free-living lifestyles and understand different reproductive strategies. We investigated the expression of small RNAs (sRNAs) that are either enriched for a 5′ monophosphate modification (5′pN) or are 5′ modification-independent, across five life cycle stages of the rodent parasite *Strongyloides venezuelensis*. We identified miRNAs and small-interfering RNAs expressed by *S. venezuelensis* that are predicted to target and regulate the expression of protein-coding genes and transposable elements (TEs). Three previously unreported classes of sRNA were identified: (1) 25Gs with a putative role in reproduction in adult females, (2) tRNA-derived 24–28 nt sRNAs (tsRNAs) which are predicted to target TEs in free-living females, and (3) 5′pN-enriched 26–29Cs with 5′ CGAATCC and 3′ TTT motifs expressed in parasitic females. We also confirmed that *S. venezuelensis* expresses the 27G class of sRNAs involved in TE regulation, which was previously identified in the rodent parasite *Strongyloides ratti*. Taken together, these results provide new insights into the role of sRNAs in reproductive biology and parasitism.

Keywords *Strongyloides venezuelensis*, Small RNA, microRNA, Helminth, Nematode, Parasite

Soil-transmitted helminth (STH) infections are recognised as one of the 20 Neglected Tropical Diseases and are among the most common infections worldwide, affecting up to 1.5 billion people[1]. STH infections are transmitted by eggs present in human faeces and mainly affect poverty-ridden communities with poor sanitation. Strongyloidiasis, caused by the parasitic nematode *Strongyloides stercoralis*, affects more than 600 million people worldwide. Although often asymptomatic, strongyloidiasis is also associated with gastrointestinal, bronchial and skin complications[2]. *Strongyloides venezuelensis* is a natural parasite of rats and is an established laboratory analogue of *S. stercoralis*[3].

Strongyloides species have a unique life cycle which includes a parasitic generation inside the host and a free-living generation in the soil[4] (Fig. 1). The parasitic cycle is initiated by the third stage infective larvae (iL3s), which penetrate the host's skin and migrate to the host's intestine where they inhabit as adult parasitic females. Parasitic females produce eggs by mitotic parthenogenesis, which then pass through with faeces and develop into male or female larvae. While male larvae develop into rhabditiform adult males, female larvae can either develop into rhabditiform adult females which will mate with males, and their offspring develop into iL3, or they can develop directly into iL3s[5] (Fig. 1). The alternating parasitic and free-living generations have genetically

[1]Department of Life Sciences, University of Bath, Bath BA26EA, UK. [2]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8562, Japan. [3]Forestry Vocational School, Duzce University, 81620 Duzce, Turkey. [4]Genomics and Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, Miyazaki 889-1692, Japan. ✉email: v.l.hunt@bath.ac.uk; taisei.kikuchi@edu.k.u-tokyo.ac.jp

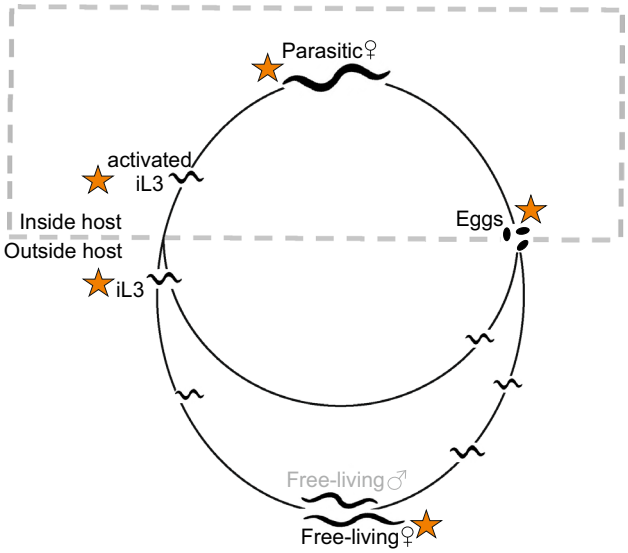

Fig. 1. The *Strongyloides* life cycle stages used for small RNA analysis. Small RNA was sequenced from five life cycle stages of *S. venezuelensis* (yellow star): free-living female adults, free-living infective larvae (iL3) searching for a host, parasitic 'activated' infective larvae (activated iL3) in the host environment, adult parasitic females from the host's intestine, and their eggs. Unlike other *Strongyloides* species, free-living males were not found in *S. venezuelensis*.

identical parasitic and free-living adult females, which provides an ideal model to investigate the role of small RNAs (sRNAs) in parasitism and the different reproductive strategies employed in the *Strongyloides* life cycle.

sRNAs, typically 18–50 nucleotides in length[6,7], play an important role in transcriptional and posttranscriptional gene regulation[8,9]. sRNAs can be divided into three main classes: microRNA (miRNA), short-interfering RNA (siRNA) and piwi-interacting RNA (piRNA)[10,11]. A defining characteristic of nematode sRNAs is the presence of 2'-O-methylation and a 3' hydroxyl and 5' phosphate group[12] with either 5'-monophosphate or 5'-triphosphate modification[13]. miRNAs often have conserved sequence identity across distantly related species[14], whereas siRNAs and piRNAs are more diverse, even between closely related species, but may have conserved functions or pathways[15]. sRNAs can be further classified by their length and first base; for example e.g., 27G is 27 nucleotides long and starts with guanine[16]. Each class of sRNAs belongs to a specific pathway in *C. elegans*[17] and can contribute to posttranscriptional silencing of mRNAs and transposable elements (TEs)[18]. In parasitic nematodes, sRNAs have been reported as key mediators of host-parasite interactions, often through selective association with Argonaute proteins that determine their function and target specificity. Specific sRNAs are encapsulated within extracellular vesicles, which are then secreted and can be absorbed by host cells, potentially influencing host gene expression or immune responses. These secreted sRNAs, often associated with parasite-specific Argonaute proteins, constitute a sophisticated mechanism for molecular communication and immunomodulation at the host–parasite interface[19,20].

sRNAs expressed in *S. venezuelensis* have not been previously reported. To understand the role of sRNA pathways in parasitism and their roles in gene regulation, we analysed small RNAseq data for five life cycle stages of *S. venezuelensis* (Fig. 1). Two library preparation methods were used for all samples, a 5' modification-independent library (RppH-treated library), which captures sRNAs with 5' mono- and poly-phosphate groups including secondary siRNAs with a 5' triphosphate modification and 5' capped sRNAs, and a 5' monophosphate-enriched library (5'pN-enriched library) which is enriched for 5'-monophosphate sRNAs including miRNAs and primary siRNAs. We investigated sRNA diversity, predicted the target transcripts, and identified sRNAs that were differentially expressed across the life cycle. Our results have identified previously undiscovered classes of sRNA with putative roles in reproduction and TE regulation.

Results

sRNAs are differentially expressed across five *S. venezuelensis* life cycle stages

We investigated the expression levels of sRNAs across five *S. venezuelensis* life cycle stages: (i) parasitic females from the host small intestine, (ii) free-living females from faecal culture (iii) activated 'parasitic' iL3s, representing iL3s inside a host, (iv) free-living iL3s in the host-seeking stage and (v) eggs collected from parasitic females. The expression of sRNAs at these stages was examined using two types of libraries: (i) those enriched for sRNAs with a 5' monophosphate (5'pN-enriched libraries), and (ii) those treated with RppH to enhance the cloning efficiency of 5' polyphosphorylated and 5' capped sRNAs (RppH-treated libraries), with each type conducted in triplicate.

The sRNA reads were classified as those originating from either miRNAs or putative siRNAs (p-siRNA). All five stages of the life cycle expressed miRNAs of 21–23 nt in length with a propensity for 5' uracil. The sRNA expression profiles for the 5'pN-enriched and RppH-treated libraries for activated iL3s, free-living iL3s,

and eggs had similar proportions of sRNA classifications (Fig. 2A–C), suggesting that these life cycle stages predominantly express sRNAs with a 5' monophosphate moiety. p-siRNAs of varying lengths and first bases were expressed in activated iL3s and eggs. Of the 345 p-siRNA sequences expressed in activated iL3s and eggs, 44% and 50% originated from genes, and 18% and 26% originated from intergenic regions, respectively (SI Table X1) (Fig. 2B,C). In contrast, the sRNAs expression profiles were distinctly different between the two libraries for the parasitic and free-living females (Fig. 2), where miRNAs dominated the libraries and p-siRNAs were expressed at higher levels in the RppH-treated library.

We investigated the unique sRNA sequences expressed in both the 5'pN-enriched and RppH-treated libraries. Activated iL3s, free-living iL3s and eggs had a similar number of unique sRNA sequences in both libraries (Table 1), while parasitic and free-living females expressed 3.9 times as many unique sRNA sequences in the RppH-treated library, suggesting a greater diversity of sRNAs with a 5' polyphosphate or capped modification. Differential expression analysis identified distinct subgroups of sRNA that were specifically upregulated for each of the five life cycle stages (SI Table X2, SI Fig. 1 and 2). To investigate sRNAs associated with adult female stages, sRNAs that were upregulated in the parasitic and free-living females relative to either activated iL3s, free-living iL3s, or eggs (SI Fig. 2), were categorised into sRNA families according to their sequence length, first 5' nucleotide, and sRNA origin sequence in both the 5'pN-enriched and RppH-treated library (Fig. 3A,B). Based on these features, distinct families of sRNA were observed in the RppH-treated libraries including: (i) 27Gs upregulated in both parasitic and free-living females, compared to iL3 stages and eggs; (ii) 25Gs upregulated in free-living females compared to all four other life cycle stages; (iii) 26-29Cs upregulated in parasitic females compared to all four other life cycle stages; and in the 5'pN-enriched library: (iv) 22-23U miRNAs upregulated in free-living females, parasitic females and free-living iL3s compared to activated iL3s and eggs.

Free-living females express 25G sRNAs associated with reproduction

Free-living females express 25 and 26 nucleotide long sequences with a propensity for a 5' guanine (referred to as 25G and 26G sRNAs; SI Table X1) which together make up 29–30% of their expressed sRNAs, and 79–86% originate from intergenic regions of the genome. The 25Gs and 26Gs were specifically expressed in the RppH-treated, but not the 5'pN-enriched library, indicating that they have a polyphosphate or capped 5' modification (Fig. 3A,B). No Dicer-processing signature was identified in the 25G or 26G sequences (SI Fig. 3B), indicating these sRNAs were likely not generated via the canonical siRNA biogenesis pathway, suggesting an alternative processing mechanism or independent transcriptional origin. In free-living females, 25G and 26G sRNAs were significantly upregulated compared to all four other life cycle stages investigated (FDR ≤ 0.01, EdgeR Fishers Exact test). The 25-26Gs were also upregulated in parasitic females compared with free-living iL3s, activated iL3s and eggs. This suggests that 25-26Gs are particularly associated with the adult female life cycle stages.

A greater number of 25G (n = 4,470) and 26G (n = 4,236) sequences were significantly upregulated in free-living females than 25Gs (n = 454) and 26Gs (n = 935) sequences in parasitic females, when compared to other life cycle stages (FDR ≤ 0.01, EdgeR Fisher's exact test) (Fig. 3A, SI Table X3). A direct comparison of the 25G and 26Gs between free-living and parasitic females identified 641 25Gs and 329 26Gs significantly upregulated in free-living females, compared to only six 25Gs and 26 26Gs upregulated in parasitic females (FDR ≤ 0.01, EdgeR Fishers Exact test). Together, these results suggest that a subgroup of 25-26Gs is specific to, and putatively play important roles in, the free-living female stage of *S. venezuelensis* (Fig. 3A). The 25G and 26G sRNAs are not found in *S. ratti,* and differ from ERGO-1 26Gs in *C. elegans* which are Dicer cleaved (Supplementary Table 1).

Genes targeted by 25G and 26G sequences were predicted based on their sequence complementarity followed by GO Enrichment analysis of the predicted target genes. Of the 970 25-26Gs upregulated in free-living females (*cf.* parasitic females), 555 25Gs and 251 26Gs were predicted to target 359 and 171 protein-coding genes, respectively. These genes were enriched for 133 and 42 BP-GO terms, (SI Table X4) including 10 GO terms related to histone modification and chromosome regulation such as chromosome organisation (GO:0051276) and histone modification (GO:0016570), and 11 GO terms associated with reproduction or embryo development such as reproduction (GO:0000003), embryo development (GO:0009790), sexual reproduction (GO:0019953), gamete generation (GO:0007276) and germ cell development (GO:0007281) (Fig. 4A). A further six enriched BP-GO terms for genes where the 3'UTR is specifically targeted by 25Gs, include GO terms related to sex determination (SI Table X4). The 359 genes targeted by 25Gs upregulated in the free-living female (*cf.* parasitic female) had a range of predicted functions including 66 genes associated with transcription, translation and chromosomes, 10 genes associated with embryogenesis-related proteins such as 'Maternal protein pumilio', 'Armadillo-like proteins' and 'Homeobox proteins' (Fig. 4B, SI Table X4), and 83 genes with unknown functions. Of the 25Gs upregulated in free-living (*cf.* parasitic females), only 10.45% aligned to 26G sequences, indicating that in most cases, the 25Gs are not short versions of the 26G sequences. However, 93 of the 359 genes targeted by 25Gs were also targeted by 26Gs, indicating that there is overlap in the genes that they target. The 93 shared target genes were enriched for 36 diverse GO terms, including DNA integration (GO:0015074) and nucleic acid metabolic processes (GO:0090304) (SI Table X4). A further 4,236 25G sequences were not differentially expressed between parasitic and free-living females; 3,783 of which were predicted to target 1,498 genes enriched for general BP-GO terms such as 'Cellular Processes' (GO:0009987) and reproductive-associated GO terms and proteins (Fig. 4A). Together these results suggest that 25Gs regulate reproduction-associated genes in both parasitic and free-living female generations but have an additional role in free-living females.

To determine if the genes predicted to be targeted by 25Gs were also differentially expressed between life cycle stages, we investigated gene expression in parasitic and free-living females using RNA-seq data. Of the 359 predicted genes putatively targeted by 25Gs upregulated in free-living females (*cf.* parasitic females), 16 genes were significantly upregulated and 53 were downregulated in free-living females compared with parasitic females (EdgeR, FDR < 0.01) (Fig. 4C, SI Table X5). Among the predicted genes targeted by the 1,498 25Gs that

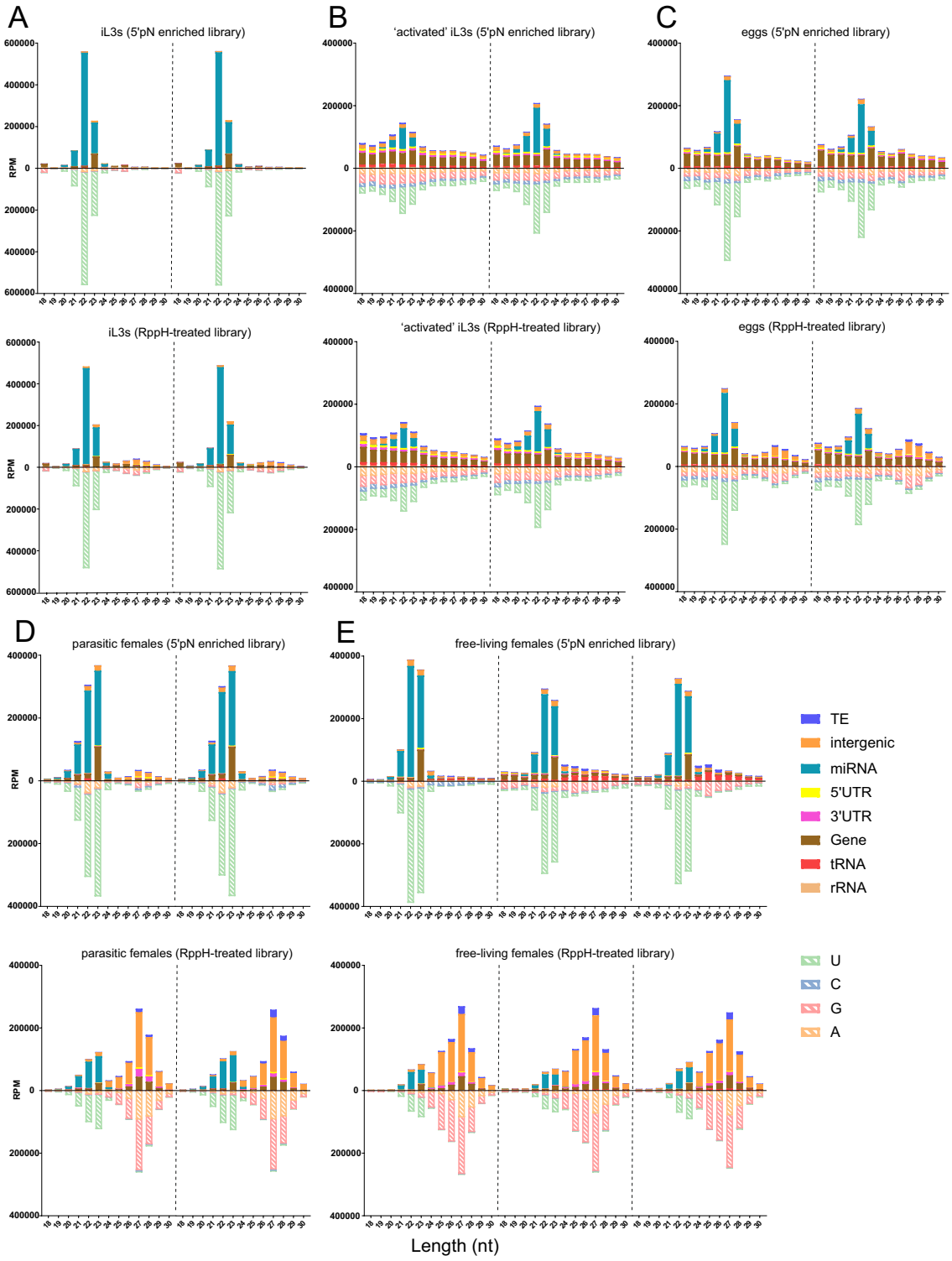

Fig. 2. Expression profiles of sRNAs from RppH-treated and untreated libraries. sRNA expression based on RPM (reads per million) in the RppH-untreated (5'pN-enriched) and treated libraries for two to three replicates of (**A**) free-living iL3s, (**B**) activated iL3s, (**C**) eggs, (**D**) parasitic females and (**E**) free-living females, collected from parasitic females. Classification of the sRNA based on the type of sequence the sRNA originates from in the genome (solid bars) and the first 5' base in the sequence (hashed bars) are highlighted by colour. sRNAs are divided into either miRNA, rRNA-derived sRNA (rRNA), tRNA-derived sRNA (tRNA) or as putative siRNA originating from either protein-coding genes (CDS or intronic regions), 5'- and 3'-UTRs, intergenic regions or transposable elements (TE).

Lifecycle stage	5'pN-enriched library			RppH-treated library		
	Raw reads	Unique sRNA sequences	Unique RPM	Raw reads	Unique sRNA sequences	Unique RPM
Parasitic ♀	936,322	23,462	25,058	1,467,016	134,366	91,591
Free-living ♀	1,899,824	57,419	30,223	4,726,900	563,189	119,146
Free-living iL3	1,018,261	15,928	15,642	510,244	13,356	26,176
activated iL3	1,586,323	56,594	35,676	1,209,485	58,849	48,656
eggs	1,624,578	37,310	22,966	1,219,893	42,851	35,127

Table 1. Number of reads mapped to *the S. venezuelensis* genome and unique sRNA sequences expressed at each lifecycle stage. The values represent the mean number of reads. Unique reads were included only if they were expressed in at least two replicates.

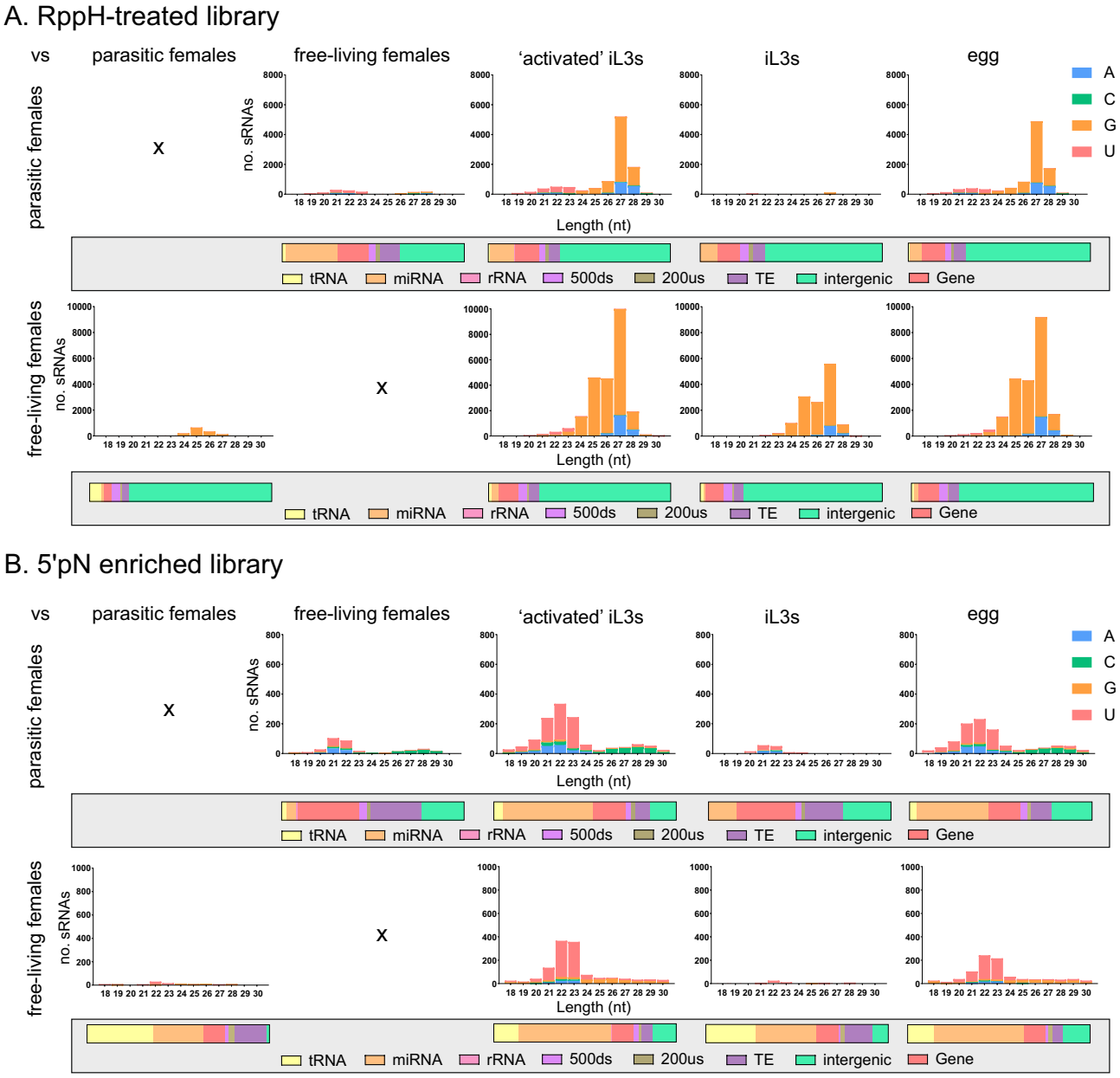

Fig. 3. sRNA upregulated in parasitic and free-living adults. sRNAs differentially expressed in free-living and parasitic females (FDR ≤ 0.01, edgeR Fishers Exact test when compared to 'activated' iL3s, free-living iL3s and eggs) in (**A**) RppH-treated library and (**B**) 5'pN-enriched library. Starting sequence base shown in key and proportion of origin (derived from tRNA, miRNA, rRNA, 500ds (3'UTR), 200us (5'UTR), TEs, intergenic regions or genes) of the expressed sRNAs shown below graphs.

Fig. 4

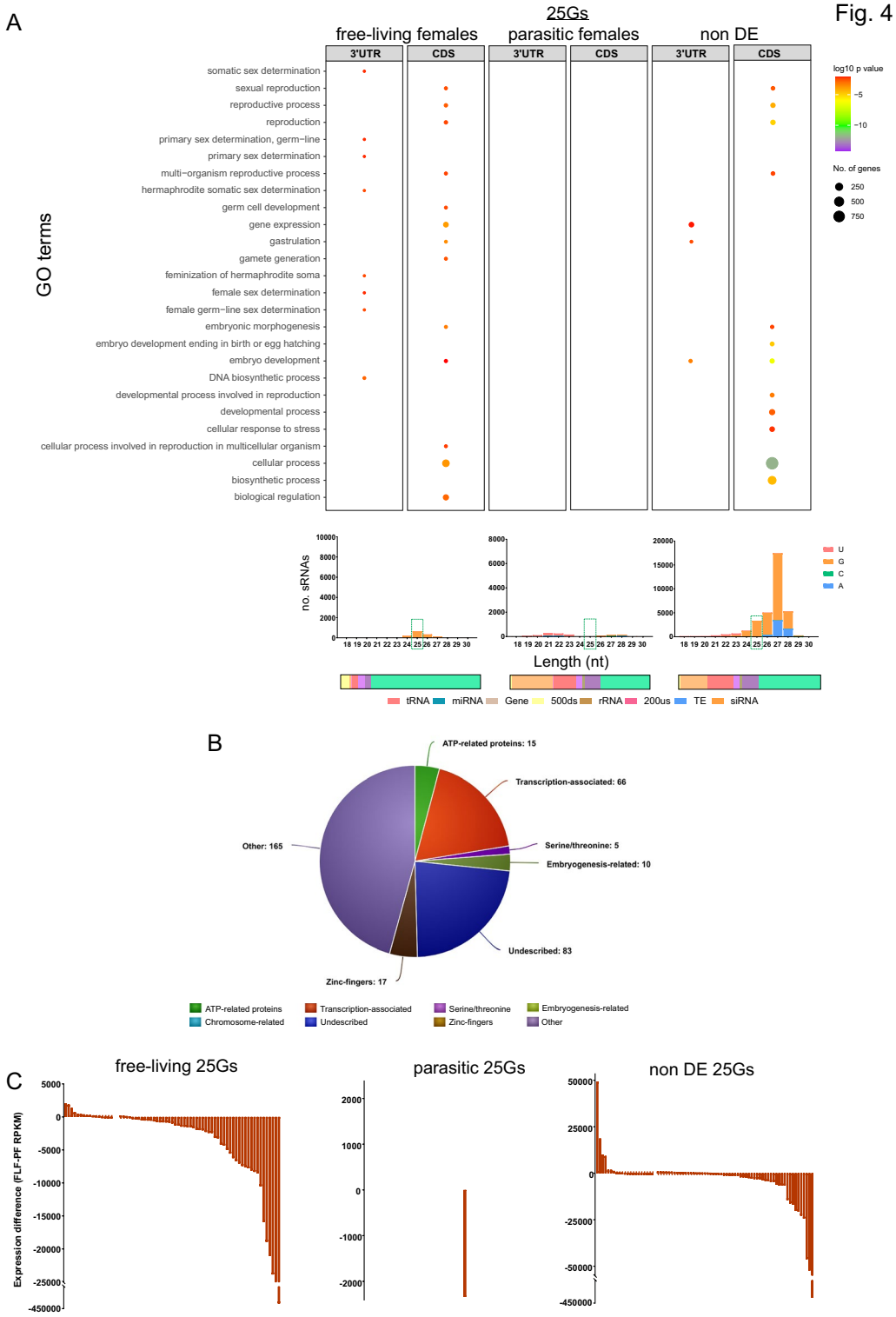

Fig. 4. Gene Ontology (GO) enrichment analysis of free-living 25G gene targets. (**A**) Enriched GO terms associated with the predicted gene targets (the coding sequence (CDS) and 3'UTRs of protein coding genes) of 25G sRNAs upregulated in free-living females, parasitic females and non-differentially expressed sRNAs.. Enriched GO terms were visualised using REVIGO[67]. Corresponding upregulated sRNAs first base, origin and nucleotide profile shown below. (**B**) the proteins associated with the free-living upregulated 25Gs. (**C**) Difference in expression level (RPKM: FLF minus PF) of genes putatively targeted by free-living upregulated, parasitic upregulated and non-DE 25Gs with a logFC of ≤ 1.5. RPKM = Reads per kilobase per million.

were not differentially expressed (DE) between parasitic and free-living females, 86 genes were upregulated and 161 genes were downregulated in free-living females.

The mean expression level of the 359 genes putatively targeted by 25Gs upregulated in free-living females 359 free-living 25G gene targets was 540 RPKM, and the expression levels of these same genes in parasitic females were 12,862 RPKM, suggesting that 25Gs may have a specific role in suppressing the expression of target genes in free-living females. Interestingly, the predicted genes targets of non-DE 25Gs also had higher expression levels in parasitic females (7,787 RPKM) compared to the same gene targets expressed in free-living females (1,125 RPKM), suggesting non-DE 25Gs regulate both adult female life cycle stage genes, with more prominent role in repressing expression of genes in the free-living females (Fig. 4C, SI Table X5).

tRNA-derived 24–28 nt sRNAs upregulated in free-living females target TEs

We categorised small RNAs that mapped precisely to tRNA genes with defined cleavage patterns at anticodon loops as tRNA-derived sRNAs (tsRNAs). tsRNAs have been associated with regulating gene expression at the transcriptional and translational levels[21–23], but little is understood known about their expression or roles in parasitic nematodes. In *S. venezuelensis* we identified tsRNAs that were predominantly 24–28 nt long with no propensity for a particular starting base (Figs. 2 and 5A). tsRNAs were expressed at higher levels in free-living females than at other life cycle stages (Fig. 5A). Within the 5'pN-enriched library, 14–36% of the sRNAs that were upregulated in free-living females (*cf.* other life cycle stages) originated from tRNAs (SI Table X1, SI Fig. 4A), compared to 0–5% of sRNAs upregulated in parasitic females (*cf.* other life cycle stages). The 5' tRNA-derived fragments (5'-tRFs) (based on the prediction of 5'-D-loop[24]) were the most abundant tsRNAs, accounting for 78% of all tsRNAs expressed in *S. venezuelensis* (Fig. 5B). We identified 192 unique tsRNA sequences upregulated in the free-living females (*cf.* other life cycle stages) and 36% of these were predicted to target 15 protein coding genes, 14 of which have no known protein function, and nine tsRNAs target the 3'UTRs of genes enriched for GO terms associated with nuclease activity and ATP synthesis (SI Table X6). In addition, 131 free-living

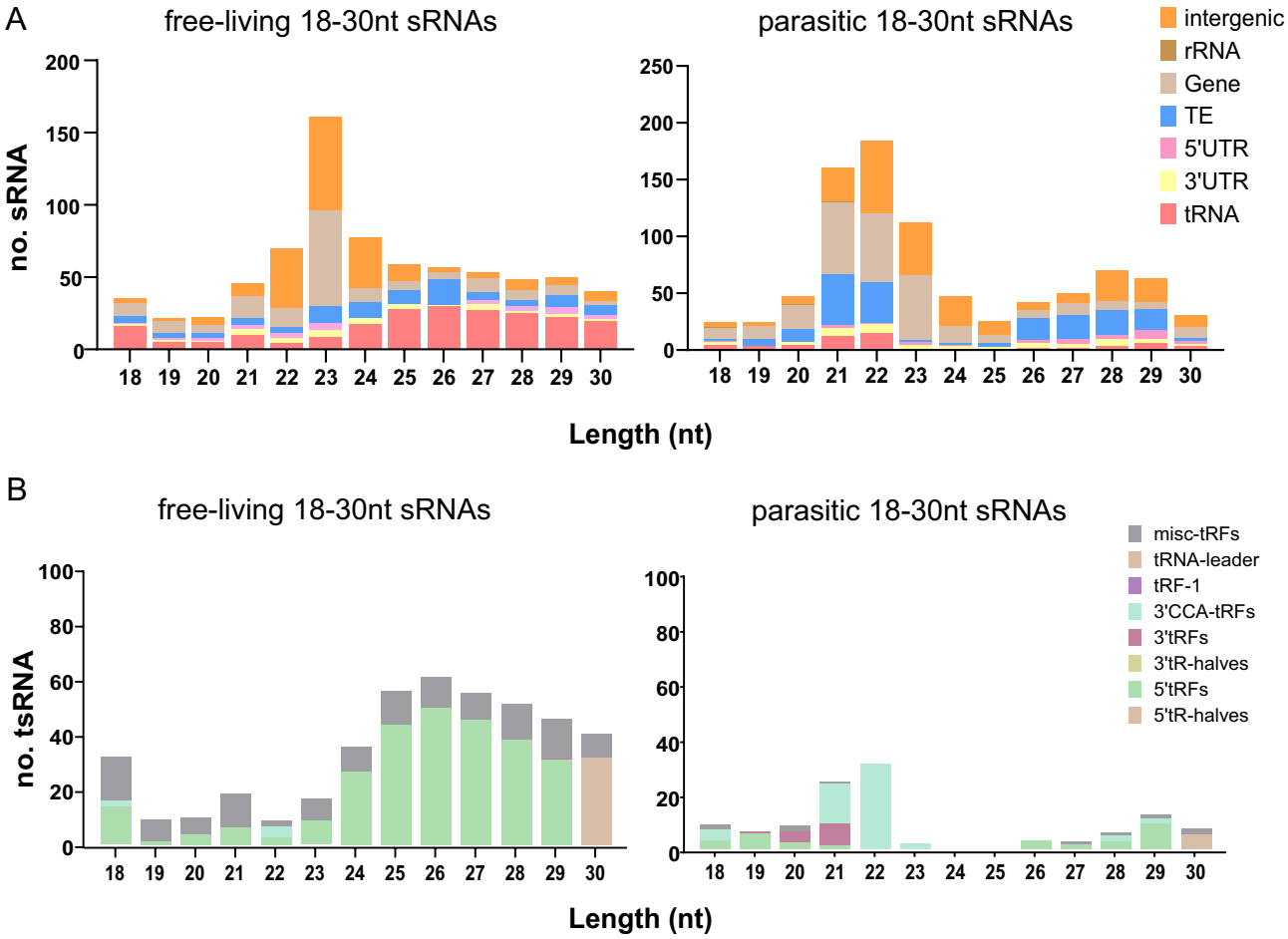

Fig. 5. sRNA expression of free-living and parasitic 18–30 nt long sRNAs with a 5' monophosphate. (**A**) Number of 18–30 nt long sRNAs upregulated in parasitic females and free-living females compared with other stages in 5'pN-enriched libraries. Upregulated sRNAs were classified by the gene locations which those originated from (intergenic regions, TE, genes, 5'UTR, 3'UTR, rRNAs or tRNA). (**B**) Number of tsRNA upregulated in free-living and parasitic females, classified into types of tRNA they originate from.

upregulated tsRNAs targeted 58 TEs. These results suggest that these tsRNAs may play a role in DNA replication and TE activity (SI Table X6).

Parasitic female 5′pN-enriched 26-29Cs have 5′ CGAAUCC and 3′ UUUU motifs

sRNAs that were 26–29 nt long with a propensity for a 5' cytosine (hereafter referred to as 26-29Cs) were highly expressed in parasitic females compared with other life cycle stages in the 5'pN-enriched libraries (FDR ≤ 0.01, EdgeR Fishers Exact test) (Fig. 3B). 26-29Cs did not show evidence of a Dicer-processing signature (SI Fig. 3C). Seventy-five and four 26-29C sequences were upregulated and downregulated, respectively, in parasitic females compared to those in free-living females (SI Table X7). Of the parasitic female upregulated 26-29Cs 60% originated from TEs. The four free-living-upregulated 26-29Cs originated from tRNAs. The parasitic female upregulated 26-29Cs had conserved CGAAUCC and UUU motifs at the 5' and 3' ends, respectively (Fig. 6, SI Fig. 4B), which are not present in the free-living upregulated 26-29Cs. Fifty-five percent of the 26-29Cs were antisense to TE sequences, and 27% targeted five protein-coding genes and 10% targeted 3'UTRs of four genes. All predicted target genes had unknown protein functions.

Parasitic and free-living expressed 27Gs target TEs and genes associated with TE activity

We identified 27G (27 nucleotides long, with a 5' guanine) sRNAs expressed by both parasitic and free-living females, and 73% and 71% of the 27Gs originate from intergenic regions in parasitic and free-living females, respectively. Additionally, 16% of 27Gs originated from genes in both life cycle stages, 7% and 8% from TEs, and 3% from 3'UTR regions.

The 27Gs were either not expressed or expressed at low levels in the eggs and larval stages (Fig. 2). In the RppH-treated library, 27G sRNAs were significantly upregulated in parasitic females and free-living females compared to all other life cycle stages (Fig. 3, FDR ≤ 0.01, EdgeR Fishers Exact test). No evidence of a Dicer-processing signature was identified for 27Gs (SI Fig. 3A). Of the 10,400 unique sRNA sequences upregulated in parasitic females compared to activated iL3s, 5,212 were 27 nt long, of which, 84% had a 5' guanine. This was similar in parasitic females compared to free-living iL3s, and parasitic females compared to eggs, where differentially expressed 27Gs account for 31% and 43% of the total sRNAs expressed in these life cycle stages, respectively. Similarly, in free-living females, 33–35% of the upregulated sRNAs in comparison to free-living iL3s, activated iL3s and eggs were 27Gs. The 27G expression profiles in the adult female stages were similar to each other and only 5% (81 sRNAs) and 7% (105 sRNAs) of parasitic female- and free-living female-upregulated sRNAs were 27Gs when these two life cycle stages were compared (Fig. 7). Together, these results indicate that 27Gs play an important role in both adult female stages of *S. venezuelensis*.

The 27G sRNAs are also expressed in other *Strongyloides spp.* and *Parastrongyloides* parasites (SI Table 1). A similar number of 27G sequences were differentially expressed between *S. venezuelensis* adult female life cycle stages (83 upregulated in parasitic females compared to 106 27Gs upregulated in free-living females) and *S. ratti* (40 upregulated in parasitic females compared to 150 upregulated in free-living females). However, a greater number of 27Gs were non-DE in *S. venezuelensis* (14,771) compared to *S. ratti* (8,693) and only 14 27Gs share a common sequence between the two species indicating that although the families of p-siRNAs are conserved between these closely related species, the sequences themselves are not.

In *S. ratti,* 27Gs are predicted to have a role in regulating TE activity[25,26]. To assess if this is also the case for *S. venezuelensis*, we re-annotated the TE sequences present in the *S. venezuelensis* genome (SI Fig. 5) and predicted the target TE sequences of 27G sRNAs expressed in parasitic and free-living females based on sequence complementarity. Out of the total 15,008 27Gs expressed in *S. venezuelensis*, 12.49% mapped antisense to i.e. are predicted to target, at least onee TE sequence (SI Table X8). More specifically, the parasitic and free-living expressed 27Gs target 85 retrotransposons, 50 DNA transposons and 474 TE sequences with an undefined TE family, ranging from 43 to 13,293 nt in length.

A further 15.67% of *S. venezuelensis* 27Gs target the 3' UTR region and 52.12% of the 27Gs target *S. venezuelensis* protein-coding genes (SI Table X3). These 27Gs were predicted to target 1,310 genes. GO enrichment analysis revealed that these genes were enriched for 46 BP-GO terms, 38 MF-GO terms and 10 CC-GO terms

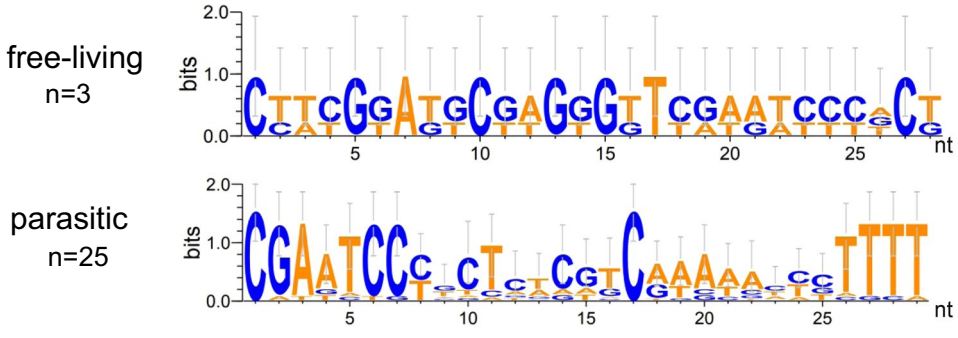

Fig. 6. Nucleotide composition of parasitic and free-living 26–29Cs. The parasitic female upregulated 26–29Cs exhibited conserved CGAAUCC and UUU motifs at their 5' and 3' ends, respectively, which are absent in the free-living upregulated 26-29Cs. Visualised in WebLogo. A complete figure illustrating 26-29Cs can be found in SI Fig. 4.N = number of unique sequences.

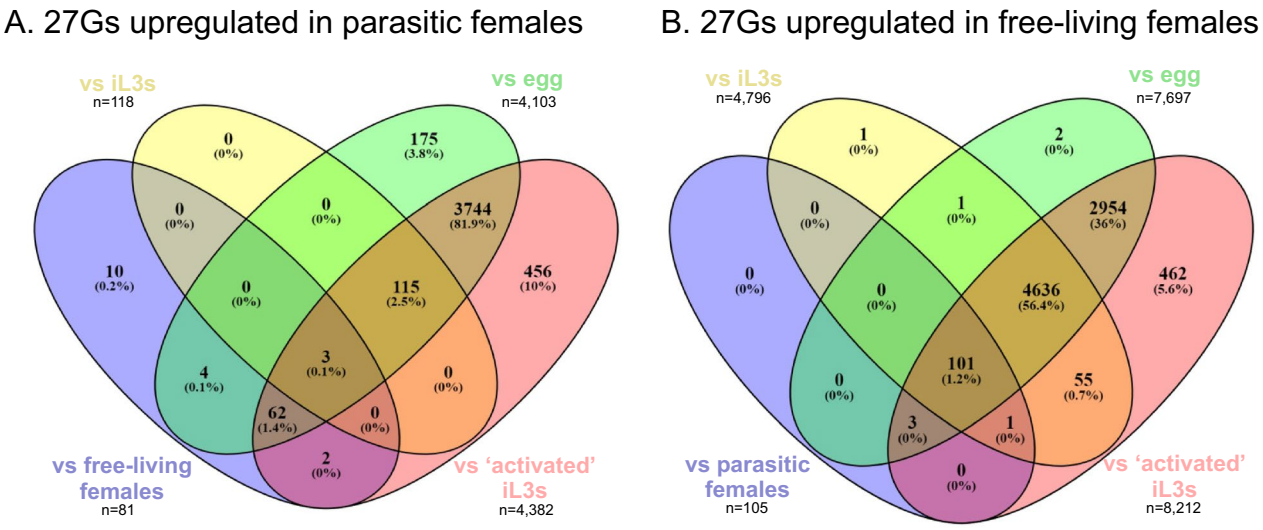

Fig. 7. 27Gs upregulated in parasitic and free-living adults. (**A**) Number of unique 27Gs upregulated in parasitic female against the four other life cycle stages. n = number of unique sequences upregulated in parasitic female against each life cycle stage. (**B**) Number of unique 27Gs upregulated in free-living female against the 4 other life cycle stages. n = number of unique sequences upregulated against each life cycle stage.

including DNA metabolic process (GO:0006259), DNA integration (GO:0015074), DNA biosynthesis process (GO:0071897) and DNA repair (GO:0006281) (Fisher's test, FDR < 0.01) (SI Table X8). Of all the genes predicted to be targeted by 27Gs, 48% code for hypothetical proteins with an undescribed protein function (SI Table X8). Based on the genes that have a predicted protein function, 25% of 27Gs target transcription-associated proteins including 'reverse transcriptase', 'integrase' and 'DNA/RNA helicase'; and 5% directly targeted transposable elements associated proteins such as 'transposase', 'retrotransposable element', 'retrotransposon-like protein', 'putative LTR retrotransposon' and Rretrotransposon-related protein from transposon'. Together, these results suggest that 27Gs target TEs and genes associated with TE activity, and likely play a role in TE regulation in *S. venezuelensis* (SI Table X8).

Putative siRNAs of varying lengths with a 5′ monophosphate are expressed in the activated iL3 and egg stages

In contrast to the sRNA expression profile for the free-living iL3s, which is dominated by miRNAs, in the activated iL3s and eggs, the most highly expressed sRNAs originated from intergenic regions and genes (Fig. 2, SI Table X9 and X10). Parasitic female and free-living female samples also contained eggs. To identify egg-specific sRNAs, the sRNAs expressed in eggs (from the parasitic female stage) were compared to the sRNAs expressed by free-living females and parasitic females, in both RppH-treated and 5'pN-enriched libraries (Fig. 8A,B). This identified 1,831 and 352 sRNAs with a 5' monophosphate upregulated in the eggs, compared to free-living and parasitic females, respectively. The sRNAs upregulated in eggs are predominantly derived from genes (eggs vs parasitic females, 59.75%; eggs vs free-living females 51.6%). No dominant starting base was observed for the sRNAs upregulated in eggs compared with the adult female stages (Fig. 8B). We predicted the targets of the differentially expressed sRNAs based on sequence complementarity. In eggs *cf.* parasitic females, 5.97% of the 352 egg-upregulated sRNAs are predicted to target five genes, none of which have a predicted protein function, 36.4% target 3'UTRs from four genes and 16.8% target TEs. Of the 1,831 sRNAs upregulated in eggs *cf.* free-living females, 4.1% were predicted to target 13 genes including one gene with a predicted function in nematode cuticle collagen and 12 genes with no predicted function, 37.4% target 3'UTRs of seven genes and 19.3% target TEs (SI Table X10).

miRNAs with UUGCGAC seed sequence are expressed across the life cycle and are unique to *Strongyloides*

We predicted 109 miRNA sequences in *S. venezuelensis*. The miRNAs are predominantly 22–23 nt with a propensity for a 5' uracil and are expressed across all life cycle stages (Fig. 2). Of these miRNAs 81, 64 and 75 miRNAs were significantly upregulated compared to at least one other life cycle stage in parasitic females, free-living females and free-living iL3s, respectively (SI Fig. 6A and B, SI Table X11a). The majority of miRNAs were not differentially expressed between the adult and free-living iL3s comparisons, (FDR ≤ 0.01, EdgeR Fishers Exact test; SI Fig. 6A and B) suggesting that these miRNAs have similar roles across these three life cycle stages.

The most common seed sequence among *S. venezuelensis* miRNAs is UUGCGAC which are differentially expressed 36 times across the free-living female, parasitic female, and free-living iL3 lifecycle stages (SI Table X11b, SI Fig. 6C). miRNAs with the UUGCGAC seed sequence are also expressed at high levels in *S. ratti*[26] and can be found in *Panagrellus redivivus* (Table 2), but have not been reported in other nematode species, suggesting this miRNA family is important in the genus *Strongyloides*. GAGAUCA and CACCGGG are the second and

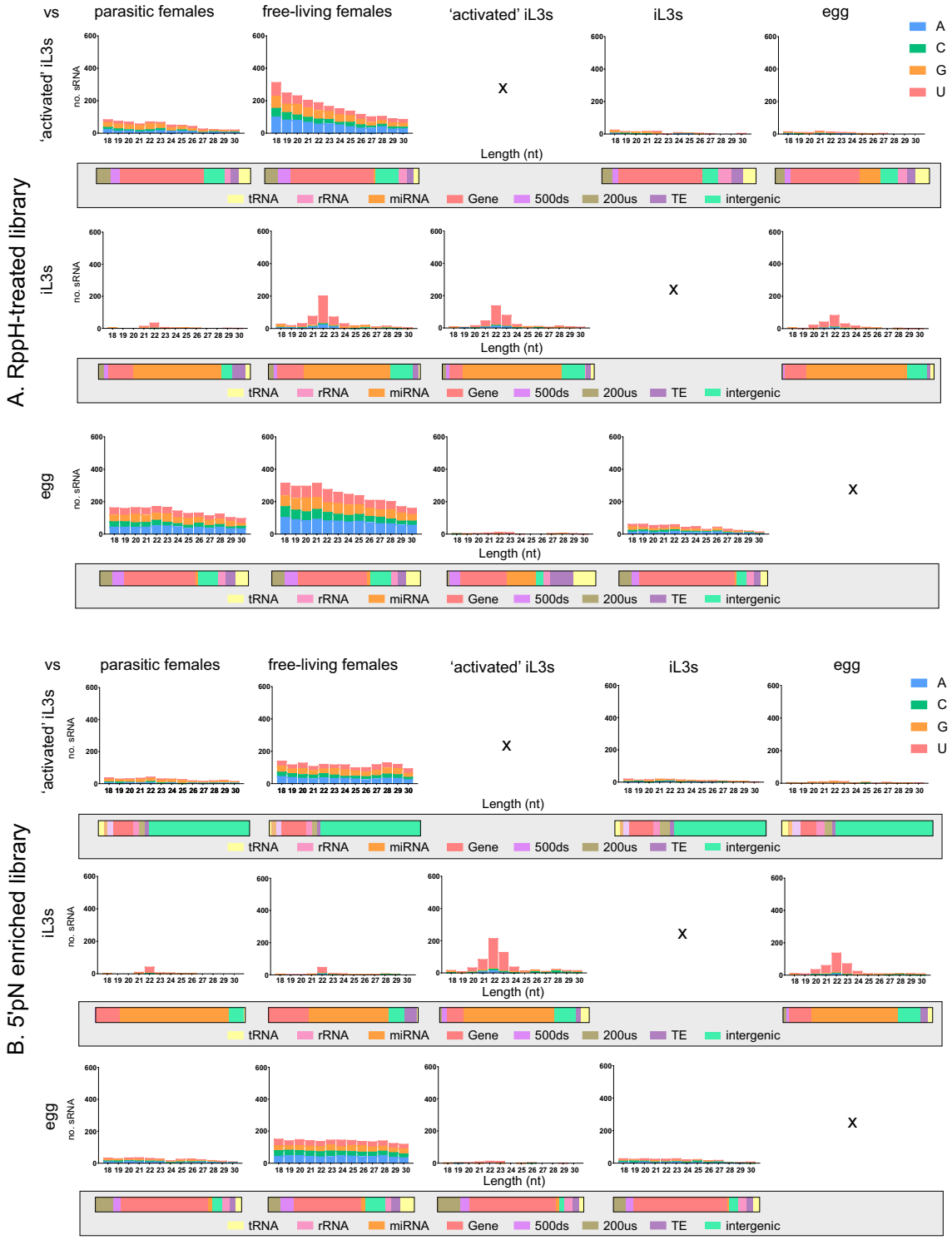

Fig. 8. sRNA upregulated in eggs, activated iL3s and free-living iL3s. sRNAs differentially expressed in activated iL3s, iL3s and eggs (FDR ≤ 0.01, edgeR Fishers Exact test when compared to other life cycle stages) in (**A**) RppH-treated library and (**B**) monophosphate library. Starting sequence base shown in key and proportion of origin (derived from tRNA, miRNA, rRNA, 500ds (3'UTR), 200us (5'UTR), TEs, intergenic regions or genes) of the expressed sRNAs shown below graphs.

	UUGCGAC	GAGAUCA	CACCGGG	Total miRNAs
A. suum	0	3	7	189
B. malayi	0	1	2	166
C. brenneri	0	4	17	152
C. briggsae	0	4	13	163
C. elegans	0	6	7	437
C. remanei	0	5	15	182
H. contortus	0	3	4	194
H. polygyrus	0	5	10	486
P. redivivus	7	2	7	394
P. pacificus	0	1	17	551
S. ratti	10	5	4	208
S. venezuelensis	18	5	3	128

Table 2. Abundance of *S. venezuelensis* miRNAs seed families in other nematodes.

third most common seed sequences in *S. venezuelensis* miRNAs and are found across diverse nematode species, suggesting these seed sequences are important in nematode biology more generally (Table 2). All three of the most common seed sequences are not found in activated iL3s and they either do not play a role in this life cycle stage or were downregulated to regulate expression of their target transcripts.

Discussion

S. venezuelensis express miRNAs and other classes of sRNAs that putatively target protein coding genes and TEs. We have identified three sRNA families that have not previously been reported, including (i) 25Gs with a putative role in reproduction in adult females, (ii) tRNA-derived 24–28 nt sRNAs which putatively regulate TEs in free-living females, (iii) 5'pN-enriched 26-29Cs with 5' CGAAUCC and 3' UUU motifs expressed in parasitic females. We also confirmed *that S. venezuelensis* expresses the 27G class of sRNAs involved in TE regulation as previously identified in *S. ratti*[25,26].

5' modifications in small RNAs play a crucial role in determining their stability, Argonaute protein association, and regulatory function. In this study, we used two libraries; (i) sRNAs enriched for a 5' monophosphate and (ii) sRNAs enriched for a 5' modification-independent library and observed distinct 5′nucleotide preferences among different sRNA classes, suggesting stage-specific regulatory mechanisms. The stage-dependent differences in 5' modifications suggest that small RNA pathways may be differentially regulated in response to environmental cues, host interactions, or developmental transitions. Further experimental validation, such as Argonaute co-immunoprecipitation assays, would help clarify how these modifications influence RNA–protein interactions. Free-living females express 25G sRNAs associated with reproduction. The 25Gs can be divided into two groups: 25Gs that are not differentially expressed between adult free-living females and parasitic females, and 25Gs that are specifically upregulated in the free-living female stage. We predict that both groups have roles in reproduction, developmental processes and embryogenesis. Genes putatively targeted by 25Gs upregulated in free-living females code for proteins such as embryonic leucine zipper, maternal protein pumilio, armadillo-like domains and homeobox proteins, which are associated with embryogenesis and germline development in *Drosophila* and *C. elegans*[27]. The importance of sRNAs has also been shown in embryogenesis of *Arabidopsis* before and after fertilization[28–30], in meiosis I in mice[31] and in DNA elimination, chromosome fragmentation and DNA amplification in the protist ciliates[32]. In *Strongyloides* spp. the free-living adult stage reproduces by sexual reproduction, unlike the parasitic adult stage which reproduces by mitotic parthenogenesis. The subset of 25Gs identified in the free-living female may therefore play a distinct role in the process of sexual reproduction. However, a similar subset of 25Gs have not been associated with free-living females in *S. ratti*[26], which have a comparable life cycle to *S. venezuelensis.* The free-living female associated 25Gs could, therefore, represent differences in fertilization status. The sRNA data for the free-living female and parasitic female includes sRNAs expressed in eggs, in addition to the adult expressed sRNAs. Due to the absence of free-living males in *S. venezuelensis* laboratory populations[3], the eggs in the free-living females were unfertilized, in comparison to the *S. ratti* laboratory life cycle which does produce free-living males. Free-living upregulated 25Gs could be associated with unfertilised eggs, for example to repress expression of reproductive genes associated with the pre-fertilisation stage. In honey bees and fruit flies sRNAs are differential expressed in pre- and post- mated eggs, indicating a role for sRNAs as regulators of post-mating gene expression changes[33,34]. sRNAs regulate embryogenesis and fertilisation in *C. elegans*[35] and further investigation is required to determine how sRNAs are associated with fertilisation status in nematodes.

In *S. venezuelensis,* the targets genes of 25Gs upregulated in free-living females (*cf.* compared to parasitic females) were enriched for GO terms involved in chromatin, chromosome, or mitotic assembly. *S. venezuelensis* and *S. ratti* differ at a chromosome level. *S. ratti,* similar to the human parasite *S. stercoralis*, has two autosomes and a sex chromosome. *S. venezuelensis* is more closely related to the parasite of livestock, *S. papillosus*[36], which has a sex chromosome fused onto one of their two autosomes. This chromosome structure has been proposed for *S. venezuelensis* which also presents with two chromosomes[37–39]. *S. papillosus* undergoes sex-specific chromatin diminution to generate males[38], but this was not observed for *S. venezuelensis*[37]. It is unclear how *S.*

venezuelensis generate free-living males in the wild (attempts to raise free-living males in lab cultures have been unsuccessful)[3]. We hypothesise that 25Gs in *S. venezuelensis* could be involved in chromatin modulation and possibly sex determination at a chromatin level. Previous work on *S. papillosus* sRNAs found that some 25Gs were expressed in mixed sex free-living adult samples[25]. The *Strongyloides* genus provides an interesting system to study reproductive strategies because of its alternative generations of genetically identical adult females undergoing either sexual reproduction or parthenogenesis. More research is needed to further elucidate the roles of sRNAs associated with reproduction and development[40].

We found the parasitic female stage of *S. venezuelensis* expressed 26-29Cs characterised by distinctive 3' and 5' motifs. The 3' UUUU residue, indicative of RNA polymerase III processing in eukaryotic sRNAs, has been identified in some human sRNAs[41]. Notably, the 5' CGAAUCC motif found in these sRNAs appears to be unique to *Strongyloides*, as it has not been documented in other organisms. While the function of these motifs in 26-29Cs remains unclear, they might be involved in protein-RNA interactions[42]. Considering that numerous 26-29Cs target transposable elements, they might contribute to TE regulation during the parasitic stage of *S. venezuelensis*.

In line with previous findings in *S. ratti* and *S. papillosus*, we observed abundant expression of 27G sRNAs with a 5' modification in *S. venezuelensis* which putatively target TE and TE-associated genes[25,26]. Our results further support that 27Gs are a common sRNA expressed by *Strongyloides* species for regulating TE activity. This sRNA class is not observed in *C. elegans*[25] and other nematodes including those belonging to Clade IV, but similar classes of sRNA have been reported in other parasitic species with highly populated TE genomes, including the protist *Entamoeba*[13,43]. The 22Gs, secondary sRNAs reported in *C. elegans* and other nematodes belonging to clades I-III[44], are not a major class of sRNAs in *S. venezuelensis* and *S. ratti*[26], and 27G may have an analogous function in *Strongyloides*.

The 27G sRNAs identified in *S. venezuelensis* and *S. ratti* exhibit a distinct 5' guanine bias, which differs from canonical piRNAs in most model organisms. Typically, piRNAs are 26–31 nucleotides long and exhibit a strong 5' uracil (U) bias, particularly in primary piRNAs. In *C. elegans*, primary piRNAs are 21 nt long and strictly start with 5'-U (21U)[45]. In contrast, secondary piRNAs, produced via the ping-pong amplification cycle, often show an enrichment of adenine (A) at position 10[46], a feature absent in the 27G sRNAs observed here. The absence of a Dicer processing signature in these sRNAs suggests that, like piRNAs, they may arise through a Dicer-independent pathway. However, their distinct 5'-G bias raises the possibility that they represent a unique class of regulatory sRNAs in *S. venezuelensis* and *S. ratti*, potentially linked to chromatin regulation, transposable element suppression, or stage-specific gene regulation. Further experimental characterization, such as PIWI protein association assays, would be required to confirm whether 27G sRNAs are functionally analogous to piRNAs or if they represent a nematode-specific sRNA class with novel regulatory roles.

The most common seed sequences in *S. venezuelensis* miRNA were GAGAUCA, CACCGGG and UUGCGA C. However, none of these three miRNA families were expressed in activated iL3s. GAGAUCA seeds belong to the Bantam family in *Caenorhabditis* and other organisms[47,48]. Bantam is involved in regulating cell proliferation and apoptosis in *Drosophila*[49] and have been associated with a role in locomotion, body size, egg laying and dauer entry in *C. elegans*[50]. The seed sequence CACCGGG belongs to the mir-36 family which is essential in embryogenesis and early larval development in other organisms including *C. elegans*[49]. We found this sequence primarily expressed in eggs, parasitic and free-living adult females, supporting that a similar role is possible in *S. venezuelensis*. miRNAs with the seed region UUGCGAC are likely to be conserved in *Strongyloides,* based on the presence of these miRNAs in *S. venezuelensis* and *S. ratti*[26], one other nematode species have only been reported in nematode species. Given that UUGCGAC miRNAs are both abundant and expressed at relatively high levels suggest that they have an important role in *Strongyloides* biology.

While this study provided a comprehensive analysis of sRNAs across multiple life stages of *S. venezuelensis*, several limitations should be acknowledged. Our findings are based on whole-organism RNA sequencing, which, while informative, does not allow for precise localisation of sRNA expression within specific tissues. Tissue-specific sRNA profiling, obtained through laser capture microdissection or single-cell RNA sequencing, could provide deeper insights into the roles of these sRNAs in distinct developmental and physiological contexts. Furthermore, our study relies on bioinformatics-based gene target prediction, which, while valuable, does not confirm direct regulatory interactions. Functional validation using RNA interference (RNAi), CRISPR-based knockouts, or transgenic reporter assays would strengthen our conclusions on the biological roles of novel sRNA classes, such as 27G and 25G RNAs. Additionally, biochemical characterisation, including PIWI and other Argonaute protein association assays, could clarify whether these sRNAs are mechanistically similar to known pathways in *C. elegans*. Future studies integrating environmental perturbation experiments, such as nutrient deprivation or host immune interactions, may also shed light on how small RNA expression is modulated in response to external stimuli. By combining these approaches, future work can further elucidate the functional significance of sRNAs in helminth development, reproduction, and host adaptation.

Conclusions

This is the first identification of small RNAs expressed in *S. venezuelensis* and the first report of small RNA expression of eggs and activated iL3s in any *Strongyloides* species. Three previously unreported classes of sRNA were identified: (i) 25Gs with a putative role in reproduction in adult females, (ii) tRNA-derived 24–28 nt sRNAs (tsRNAs) which are predicted to target TEs in free-living females, and (iii) 5'pN-enriched 26-29Cs with 5' CGAA TCC and 3' TTT motifs expressed in parasitic females. We also confirmed that *S. venezuelensis* expresses the 27G class of sRNAs involved in TE regulation, previously identified in the other rodent parasite *S. ratti*.

Methods

Sample collection

S. venezuelensis HH1 strain was maintained at the University of Miyazaki, Japan, using male Wistar rats obtained from Japan SLC Inc, by subcutaneous injection of ~ 10,000 third stage infective larvae (iL3) obtained by faecal culture using filter paper as described previously[51]. Isolated iL3 were washed three times with phosphate-buffered saline (PBS) prior to infection. Parasitic females were isolated from the small intestine of rats 7 days post-infection. Free-living females were isolated from faecal cultures. Eggs were collected from parasitic females using cell strainer (mesh size 30 µm and 1 µm). Free-living iL3s were isolated using the filter paper method. These free-living iL3s were activated by incubation at 37°C with 5% CO2 in DMEM (high glucose, Gibco) to mimic the conditions within the host.

Ethics statement

All experiments were conducted in strict accordance with procedures that had been approved by the Animal Experiment Committee of the University of Miyazaki (Miyazaki, Japan) under approval no. 2009–506-6, as specified in the Fundamental Guidelines for Proper Conduct of Animal Experiments and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education Culture Sports Science and Technology, Japan, 2006. The study also adhered to the ARRIVE guidelines for reporting in vivo animal experiments.

RNA extraction and small-RNAseq

Nematode samples were snap-frozen using liquid nitrogen and homogenized using Biomasher II (Nippi-inc). RNA was extracted using TRI reagent (Life Technology) according to the standard protocol. Following RNA extraction, each sample replicate was divided. One part was treated with RNA 5' Pyrophosphohydrolase (RppH) (NEB) at 37 °C for 30 min followed by heat inactivation at 65 °C for 20 min to remove 5' modifications in sRNA for phosphate-independent libraries as described in Suleiman et al.[26], and one part was left untreated. Libraries were prepared using the Qiaseq miRNA Library Kit (Qiagen) and sequenced on Illumina MiSeq sequencer using the standard protocol (Illumina). Two–three biological replicates were used for each life cycle stage. Adaptor sequences were trimmed using UMI_tools (github.com/CGATOxford/UMI-tools).

miRNA identification and analysis

S. venezuelensis miRNAs were predicted using miRDeep2[52] using all nematode hairpin and mature miRNAs downloaded from miRBase version 22.1 as a reference. Seed sequences were identified as the 2-8th nucleotides in the mature miRNA sequence. For comparison with other nematode species, mature miRNA sequences were downloaded from miRBase v 22.1[53].

TE annotation

TEs sequences and families in the *S. venezuelensis* genome were identified by RepeatModeler[54], LTR harvest[55], LTR digest[56], TransposonPSI (transposonpsi.sourceforge.net) and MITE-hunter[57] using default settings. Usearch[58] was then used to cluster repeats using an 80% identity threshold to produce a consensus non-redundant repeat library. RepeatMasker (repeatmasker.org) was then used with the library to mask repeat regions in the *S. venezuelensis* genome.

sRNA classification

Small-RNAseq reads were classified using a custom pipeline (https://github.com/Vicky-Hunt-Lab/HuntLab-smallRNA) and Unitas v1.6.1[24] with the parameter "-species x" and using a custom reference list of sequences including miRNA, rRNA, tRNA, mRNA and TE sequences using "-refseq" parameters. The *S. venezuelensis* rRNAs were predicted using rnammer v1.2 with default parameters[59] and tRNAs were predicted with tRNAscan-SE v2.0.6, using default parameters[60]. Full length transcripts for mRNA and the *S. venezuelensis* genome (PRJEB530.WBPS18.genomic.fa) were downloaded from WormBase ParaSite (version 18).

Length-base distribution

Reads were mapped to the *S. venezuelensis* genome (PRJEB530) using Bowtie2[61] and separated by nucleotide length—min 18 -max 30 using the TBr2_length-filter.pl script from NGS toolbox[62]; first nucleotide base information was obtained from Unitas v1.6.1[24]. Reads were normalised to reads per million (RPM) and bar charts were created in GraphPad Prism, version 7.04.

Differential expression

Mapped reads were collapsed and counted using the TBr2_collapse.pl NGS toolbox script[62]. Differentially expressed sRNAs were identified using EdgeR[63] implemented in R (version 4.0.2) between the five life cycle stages with a minimum count per million of 2 in at least two samples. sRNAs were considered differentially expressed if the FDR adjusted p-value was < 0.01. Reads were separated into predicted putative siRNA and miRNA, using Unitas[24]. Unique sequences represent collapsed single sequences from expression data, present in at least two replicates.

Target prediction for sRNAs

The reverse complement of sRNAs was mapped using Bowtie2[61] (bowtie2 -x index -f -U input.fasta -S output.sam -N 0 –norc)[22] to either (i) 5'UTR (estimated as the 200 nt upstream the mRNA coding sequence), (ii) 3'UTR (estimated as the 500 nt downstream of the mRNA coding sequence), (iii) gene regions which include exons and introns (excluding UTRs), (iv) tRNAs, (v) rRNA and (vi) predicted TE sequences.

Gene ontology (GO) analysis and protein description

GO terms were predicted using the TopGO (v2.40.0) package[64] implemented in R (version 4.0.2). Gene function was predicted using BLAST and InterProScan on OmicsBox with default parameters. GO terms in Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF) were visualised using a custom REVIGO[65] script where log10 *p*-value was plotted against number of significant genes in R (version 4.0.0).

Alignment of shorter sequences into matching longer sequences

A custom script was used (available at https://github.com/Vicky-Hunt-Lab) to collapse overlapping sequences into the longest similar sequence. Sequences were collapsed and inserted into a suffix tree and all of the leaf nodes were located. From each leaf node the tree was traversed back towards the root, recording the number of endpoints which only subtracted one character from another endpoint. The longest sequences with at least one other that subtracted one character were extracted with a record of the number of shorter sequences with one character different were collapsed into them. These were then analysed for origins and targets as described above.

Dicer-signature analysis

To search for a Dicer-processing signature, stepRNA (version 1.0.6)[66] was used on the identified sRNA families using default settings on unique sequences.

Motif identification

WebLogo: a sequence logo generator[67] was used to identify conserved motifs and nucleotide richness.

KEGG pathway analysis

For KEGG Pathway analysis, target genes were extracted using Biomart on WormBase ParaSite (version 18) and input into BlastKOALA[68] for functional characterisation.

RNAseq analysis

Transcriptome expression data from parasitic and free-living stages of *S. venezuelensis* were obtained from Hunt et al., 2018 (RNAseq data obtained from DNA Data Bank of Japan (DDBJ) under BioSample accession number SAMD00096905[69]). Putative target sequences were converted to RPKM and expression difference calculated by FLF-PF and plotted using GraphPad Prism, version 7.04. Genes considered significantly expressed if FDR ≤ 0.01.

Data availability

All sequence data from the genome projects have been deposited at INSDC under the BioProject accession PRJDB13089.

Received: 14 January 2025; Accepted: 9 May 2025
Published online: 01 July 2025

References

1. World Health Organization. *2030 Targets for Soil-Transmitted Helminthiases Control Programmes* (World Health Organization, 2020).
2. Paltridge, M. & Traves, A. The health effects of strongyloidiasis on pregnant women and children: A systematic literature review. *Trop. Med. Infect. Dis.* **3**, 50 (2018).
3. Viney, M. & Kikuchi, T. *Strongyloides ratti* and *S. venezuelensis*—Rodent models of *Strongyloides* infection. *Parasitology* **144**, 285–294 (2017).
4. Viney, M. E. A genetic analysis of reproduction in *Strongyloides ratti*. *Parasitology* **109**, 511–515 (1994).
5. Viney, M. E. The biology of Strongyloides spp. *WormBook* 1–17 (2015) https://doi.org/10.1895/wormbook.1.141.2.
6. Fu, Y., Wu, P.-H., Beane, T., Zamore, P. D. & Weng, Z. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* **19**, 531 (2018).
7. Großhans, H. & Filipowicz, W. The expanding world of small RNAs. *Nature* **451**, 414–416 (2008).
8. Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
9. Steber, H. S., Gallante, C., O'Brien, S., Chiu, P.-L. & Mangone, M. T. The *C. elegans* 3′ UTRome v2 resource for studying mRNA cleavage and polyadenylation, 3′-UTR biology, and miRNA targeting. *Genome Res.* **29**, 2104–2116 (2019).
10. Watson, C. N., Belli, A. & Di Pietro, V. Small non-coding RNAs: New class of biomarkers and potential therapeutic targets in neurodegenerative disease. *Front. Genet.* **10**, 364 (2019).
11. Zhang, C. Novel functions for small RNA molecules. *Curr. Opin. Mol. Ther.* **11**, 641–651 (2009).
12. Hafner, M. et al. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**, 3–12 (2008).
13. Zhang, H., Ehrenkaufer, G. M., Pompey, J. M., Hackney, J. A. & 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 Pathog.* **4**, e1000219 (2008).
14. Kaufman, E. J. & Miska, E. A. The microRNAs of *Caenorhabditis elegans*. *Semin. Cell Dev. Biol.* **21**, 728–737 (2010).
15. Almeida, M. V., Andrade-Navarro, M. A. & Ketting, R. F. Function and evolution of nematode RNAi pathways. *Non-Coding RNA* **5**, 8 (2019).
16. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional MicroRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
17. Hoogstrate, S. W., Volkers, R. J., Sterken, M. G., Kammenga, J. E. & Snoek, L. B. Nematode endogenous small RNA pathways. *Worm* **3**, e28234 (2014).
18. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: The vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258 (2011).
19. Chow, F.W.-N. et al. Secretion of an Argonaute protein by a parasitic nematode and the evolution of its siRNA guides. *Nucleic Acids Res.* **47**, 3594–3606 (2019).

20. White, R. et al. Extracellular vesicles from Heligmosomoides bakeri and Trichuris muris contain distinct microRNA families and small RNAs that could underpin different functions in the host. *SI Helminth Extracell. Vesicles* **50**, 719–729 (2020).
21. Ha, S. G. & Lee, S.-J.V. The role of tRNA-derived small RNAs in aging. *BMB Rep.* **56**, 49–55 (2023).
22. Liu, B. et al. Deciphering the tRNA-derived small RNAs: Origin, development, and future. *Cell Death Dis.* **13**, 24 (2021).
23. Shin, G. et al. Transfer RNA-derived fragments in aging Caenorhabditis elegans originate from abundant homologous gene copies. *Sci. Rep.* **11**, 12304 (2021).
24. Gebert, D., Hewel, C. & Rosenkranz, D. unitas: The universal tool for annotation of small RNAs. *BMC Genomics* **18**, 644 (2017).
25. Holz, A. & Streit, A. Gain and loss of small RNA classes—Characterization of small RNAs in the parasitic nematode family strongyloididae. *Genome Biol. Evol.* **9**, 2826–2843 (2017).
26. Suleiman, M. et al. piRNA-like small RNAs target transposable elements in a Clade IV parasitic nematode. *Sci. Rep.* **12**, 10156 (2022).
27. Kimble, J. & Nüsslein-Volhard, C. The great small organisms of developmental genetics: Caenorhabditis elegans and Drosophila melanogaster. *Dev. Biol.* **485**, 93–122 (2022).
28. Calarco, J. P. et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151**, 194–205 (2012).
29. Parent, J.-S., Cahn, J., Herridge, R. P., Grimanelli, D. & Martienssen, R. A. Small RNAs guide histone methylation in Arabidopsis embryos. *Genes Dev.* **35**, 841–846 (2021).
30. Fei, Y., Nyikó, T. & Molnar, A. Non-perfectly matching small RNAs can induce stable and heritable epigenetic modifications and can be used as molecular markers to trace the origin and fate of silencing RNAs. *Nucleic Acids Res.* **49**, 1900–1913 (2021).
31. Stein, P. et al. Essential role for endogenous siRNAs during meiosis in mouse oocytes. *PLOS Genet.* **11**, e1005013 (2015).
32. Rzeszutek, I., Maurer-Alcalá, X. X. & Nowacki, M. Programmed genome rearrangements in ciliates. *Cell. Mol. Life Sci.* **77**, 4615–4629 (2020).
33. Watson, O. T. et al. Abundant small RNAs in the reproductive tissues and eggs of the honey bee, *Apis mellifera*. *BMC Genomics* **23**, 257 (2022).
34. Fowler, E. K., Bradley, T., Moxon, S. & Chapman, T. Divergence in transcriptional and regulatory responses to mating in male and female fruitflies. *Sci. Rep.* **9**, 16100 (2019).
35. Hall, S. E., Chirn, G.-W., Lau, N. C. & Sengupta, P. RNAi pathways contribute to developmental history-dependent phenotypic plasticity in *C. elegans*. *RNA* **19**, 306–319 (2013).
36. Hunt, V. L. et al. The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nat. Genet.* **48**, 299–307 (2016).
37. Hino, A. et al. Karyotype and reproduction mode of the rodent parasite *Strongyloides venezuelensis*. *Parasitology* **141**, 1736–1745 (2014).
38. Nemetschke, L., Eberhardt, A. G., Hertzberg, H. & Streit, A. Genetics, chromatin diminution, and sex chromosome evolution in the parasitic nematode genus strongyloides. *Curr. Biol.* **20**, 1687–1696 (2010).
39. Kounosu, A. *et al.* Syntenic relationship of chromosomes in *Strongyloides* species and *Rhabditophanes diutinus* based on the chromosome-level genome assemblies. *Philos. Trans. R. Soc. B Biol. Sci.* **379**, 20220446 (2024).
40. Al-Jawabreh, R. *et al.* Advancing *Strongyloides* omics data: bridging the gap with *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B Biol. Sci.* **379**, 20220437 (2024).
41. Perumal, K. & Reddy, R. The 3' end formation in small RNAs. *Gene Expr.* **10**, 59–78 (2002).
42. Park, B. & Han, K. Discovering protein-binding RNA motifs with a generative model of RNA sequences. *Comput. Biol. Chem.* **84**, 107171 (2020).
43. Zhang, H., Ehrenkaufer, G. M., Hall, N. & Singh, U. Identification of oligo-adenylated small RNAs in the parasite Entamoeba and a potential role for small RNA control. *BMC Genomics* **21**, 879 (2020).
44. Sarkies, P. et al. Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLOS Biol.* **13**, e1002061 (2015).
45. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol. Cell* **31**, 67–78 (2008).
46. Gunawardane, L. S. et al. A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in Drosophila. *Science* **315**, 1587–1590 (2007).
47. Fromm, B. et al. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu. Rev. Genet.* **49**, 213–242 (2015).
48. Fromm, B. et al. MirGeneDB 2.0: The metazoan microRNA complement. *Nucleic Acids Res.* **48**, D132–D141 (2020).
49. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in drosophila. *Cell* **113**, 25–36 (2003).
50. Alvarez-Saavedra, E. & Horvitz, H. R. Many families of C. elegans MicroRNAs are not essential for development or viability. *Curr. Biol.* **20**, 367–373 (2010).
51. Maeda, Y. et al. Secretome analysis of Strongyloides venezuelensis parasitic stages reveals that soluble and insoluble proteins are involved in its parasitism. *Parasit. Vectors* **12**, 21 (2019).
52. Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).
53. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: From microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).
54. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457 (2020).
55. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf.* **9**, 18 (2008).
56. Steinbiss, S., Willhoeft, U., Gremme, G. & Kurtz, S. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* **37**, 7002–7013 (2009).
57. Han, Y. & Wessler, S. R. MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199–e199 (2010).
58. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
59. Lagesen, K. et al. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
60. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. in *Gene Prediction* (ed. Kollmar, M.) vol. 1962 1–14 (Springer New York, New York, NY, 2019).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
62. Rosenkranz, D., Han, C.-T., Roovers, E. F., Zischler, H. & Ketting, R. F. Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence. *Genomics Data* **5**, 309–313 (2015).
63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
64. Adrian Alexa, J. R. topGO. Bioconductor https://doi.org/10.18129/B9.BIOC.TOPGO (2017).
65. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

66. Murcott, B. et al. stepRNA: Identification of Dicer cleavage signatures and passenger strand lengths in small RNA sequences. *Front. Bioinf.* **2**, 994871 (2022).
67. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
68. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
69. Hunt, V. L., Hino, A., Yoshida, A. & Kikuchi, T. Comparative transcriptomics gives insights into the evolution of parasitism in Strongyloides nematodes at the genus, subclade and species level. *Sci. Rep.* **8**, 5192 (2018).

Acknowledgements

The authors thank all members of Parasite Systems Biology lab for their critical comments and technical supports. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

Author contributions

V.L.H. and T.K. conceived the study. D.L. V.L.H. and T.K. wrote the manuscript with inputs from others. A.K. and A.Y. prepared biological samples and performed sequencing. D.L., M.D., B.M., A.F., performed informatics analyses and prepared figures. All authors reviewed the manuscript.

Funding

VLH was funded by a Japanese Society for the Promotion of Science Fellowship (PE16024) and a Wellcome Trust Sir Henry Dale Fellowship (211227/Z/18/Z). TK was funded by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 19H03212 and 17KT0013, and JST CREST Grant Number JPMJCR18S7 and JPMJCR23B1. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01968-2.

Correspondence and requests for materials should be addressed to V.L.H. or T.K.