Comparative transcriptomics reveals an extracellular worm argonaute as an ancestral regulator of LTR retrotransposons

Isaac Martínez-Ugalde[1,2], Kyriaki Neophytou[3], Yenetzi Villagrana-Pacheco[1], Adriana Orrego Durañona[4], Lewis Stevens[5], Xiaochen Du[3], Rowan Bancroft[1], Jessica L Hall[1], Amy B Pedersen[1], Mark Blaxter[5], Amy H Buck[3], Cei Abreu-Goodger[1]*

1. Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

2. Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, ON.

3. Institute of Immunology & Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

4. Institute of Cell Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh, UK

5. Tree of Life, Wellcome Sanger Institute, Hinxton, UK

*Correspondence should be addressed to Cei Abreu-Goodger (cei.abreu@ed.ac.uk)

Abstract

Safeguarding the genome from non-self elements is essential for reproduction, development, and ageing. One of the major threats to genome integrity is Transposable Elements (TEs), which can be post-transcriptionally silenced through small RNAs (sRNAs) and argonaute proteins. Recent work suggests TE-derived sRNAs may also act as virulence factors in host–pathogen interactions. During infection, the intestinal parasite *Heligmosomoides bakeri* secretes a single argonaute protein (exWAGO) and a wide variety of TE-derived sRNAs. Although exWAGO is highly expressed, conserved, and secreted by parasitic nematodes, its function and sRNA guide preference remain unclear. Using comparative transcriptomics of the sRNAs bound to exWAGO within parasites of rodents, livestock and humans, and its orthologs in *C. elegans*, we found that exWAGO is capable of loading sRNAs produced from all classes of TEs in addition to some protein-coding and non-coding transcripts. However, our results suggest that the ancestral endogenous function of exWAGO was likely linked to LTR retrotransposon regulation. To understand how this relates to potential extracellular functions of exWAGO we also examined the sRNAs bound to exWAGO secreted by *H. bakeri* in both vesicular and non-vesicular forms. Extracellular exWAGO preferentially loads sRNA guides derived from non-autonomous and fragmented LTRs, suggesting the existence of adaptable reservoirs of regulatory sRNAs with potential roles in cross-species RNA communication. Together, our results show that exWAGO is part of an evolutionarily conserved pathway for LTR retrotransposon regulation, while preferentially utilising degenerated elements as sources of secreted sRNAs.

Key-words: Argonaute, Parasite, Retrotransposons, RNA interference, RNA communication

Introduction

Although typically neutral (Arkhipova, 2018), transposable elements (TE) can have deleterious effects on their host genomes, such as gene disruption or ectopic recombination. Disregulated TE activity can also induce genome expansion, potentially leading to genomic instability. To offset this, host genomes have adapted different molecular mechanisms, including non-coding RNA pathways, to regulate TE activity (Van Lopik et al., 2023; Wolf et al., 2020; W. Zhou et al., 2020). The PIWI-interacting RNA (piRNA) pathway, for instance, is conserved across metazoans (Calcino et al., 2018) and plays a key role in TE silencing through sequence-specific targeting. In *Caenorhabditis elegans,* the PIWI argonaute (PRG-1) and its bound piRNAs trigger TE-derived silencing, inducing the amplification of antisense secondary small interfering RNAs (known as 22G siRNAs) by RNA-dependent RNA polymerases, which are loaded by worm-specific argonautes (WAGOs) (reviewed by Fischer 2024). These WAGO-22G complexes are ultimately responsible for TE silencing. Many clades of nematodes, however, have independently lost essential components of the PIWI pathway, including either the PIWI argonaute protein or the capacity to produce piRNAs (Beltran et al., 2019; Sarkies et al., 2015). The clade V of nematodes (divided into Strongylida and Rhabditina) have mostly maintained PIWI argonaute and the capacity to produce piRNAs. Strongylida parasites, however, show oversized genomes with large TE content, compared to the free-living Rhabditina (International Helminth Genomes Consortium, 2019). Despite the presence of the PIWI pathway in these nematodes, little is known about how TEs are regulated beyond *C. elegans*.

The gastrointestinal parasite of mice *Heligmosomoides bakeri,* secretes immunosuppressive extracellular vesicles (EVs) during infection, which carry TE-derived 22G siRNAs and the extracellular worm argonaute (exWAGO) (Buck et al., 2014; Chow et al., 2019). *H. bakeri*, is closely related to human infective nematodes (including the human hookworms *Necator americanus*, and *Ancylostoma* spp.), which collectively affect close to 25% of the human population (Stevens et al., 2023; World Health Organization, 2020). Given the limited anthelmintic options and the increase in resistance to them, novel therapeutic options are required. We recently showed that *H. bakeri* secretes exWAGO, both inside EVs and in a vesicle-free form, and that immunisation against exWAGO confers partial protection against *H. bakeri* infection (Neophytou et al. 2025). Both exWAGO forms associate with TE-derived small RNAs and can be internalized by host cells, suggesting a potential role in RNA-based cross-species communication (Neophytou et al., 2025). Furthermore, exWAGO is conserved among clade V nematodes and

69 highly expressed relative to other argonaute proteins (Chow et al., 2019). Biochemical and
70 proteomic analyses show its orthologs can be secreted by different species, including *C. elegans*,
71 human-infective *Ancylostoma ceylanicum*, ruminant-infective *Teladorsagia circumcincta* and
72 *Trichostrongylus colubriformis*, and rat-infective *Nippostrongylus brasiliensis* (Chow et al., 2019;
73 Neophytou et al., 2025; Nikonorova et al., 2022; Rooney et al., 2022; Uzoechi et al., 2023). While
74 the role of exWAGO in the extracellular environment is potentially linked to extracellular
75 communication including parasite-host communication, its endogenous function and the features
76 that drive its preference for loading specific sRNA guides remain uncharacterized.
77
78 Here, we use comparative genomics and transcriptomics of exWAGO-bound sRNAs in clade V
79 nematodes to investigate how exWAGO guide preference evolved, and to probe how this
80 underlies both genome defence and extracellular RNA export. Our findings suggest that the
81 ancestral role of exWAGO was to regulate LTR retrotransposons. We also identify specific
82 properties of secreted LTR-derived sRNAs in *H. bakeri,* highlighting how non-autonomous LTR
83 fragments are a preferential source for potential cross-species RNA communication functions in
84 parasitic nematodes.

Results

Improving transposable element annotation in Strongylida genomes

We previously observed that exWAGO prefers transposon-associated siRNAs (Chow et al., 2019; Neophytou et al., 2025). Yet to properly understand transposon regulation, we need high-quality annotations of genomic repeats. To get an evolutionary perspective on transposon regulation, we first improve the repeat annotations for the *H. bakeri*, *H. polygyrus*, *N. brasiliensis*, *T. circumcincta* and *A. ceylanicum* genomes, using *de novo* prediction followed by automatic and manual curation (see Methods: Transposable element annotation).

To assess if our curated repeat libraries improved the quality of genome annotations, we compared them against available repeat libraries from WormBase ParaSite and previous annotation projects (Chow et al., 2019; Howe et al., 2017). Given the lack of a previous repeat library, *H. polygyrus* was excluded from this comparison. We first compared the number and length of non-redundant consensus sequences of the most common TE categories: DNA transposons (DNA), long interspersed nuclear elements (LINEs), long terminal repeat retrotransposons (LTRs), short interspersed nuclear elements (SINEs) and Unknown repeats (Unknown). Our curated libraries in general contain a similar or increased number of consensus sequences for DNA, LINE and LTR elements, but a decrease in Unknown repeats, suggesting the latter have been incorporated into named classes (Fig. 1A). Our curated models also show a significant increase in length regardless of the category (one-tailed Wilcoxon rank-sum test, $p < 0.01$) (Fig. 1A). Finally, regardless of TE classification, curated models encode a larger number of relevant proteins (e.g. transposase, reverse transcriptase) (Supplementary Table S1). Our curated models are thus not only longer but will also enhance functional analyses of TEs in these organisms.

We further assessed the degree of similarity between uncurated and curated libraries using the benchmarking method provided by RepeatModeler2 (Flynn et al., 2020). Briefly, this benchmark uses a tested library to annotate a reference library, to then classify models based on their relative sequence similarity, coverage and divergence into perfect, good, present, and missed models.
For *H. bakeri*, when comparing the uncurated against the curated library, we found 60 families with perfect agreement, 278 good, 509 present and 876 missing. This means that 876/1726 (51%) of the models in the curated library are missing in the uncurated library. We found similar

117 proportions of missing models in the uncurated libraries for the rest of the species (mean= 39.8%,
118 SD= 7.6%) (Supplementary Table S2). The number of missing families could reflect highly
119 fragmented/divergent or genuinely absent families in the uncurated sets relative to the curated
120 ones. Together, these results highlight important differences in the coverage and identity level
121 between curated and uncurated libraries.
122
123 Our newly annotated genomes vary substantially in repeat content, consistent with their size
124 differences, from 38% of the 257 Mb genome of *N. brasiliensis* to 65% of the 649 Mb genome of
125 *H. bakeri* (Fig. 1B). Notably, the TE span in Strongylida is up to four times greater than the entire
126 genome of *C. elegans* (100 Mb), underscoring substantial TE expansion in parasitic lineages.
127
128 Importantly, using our curated libraries, we substantially reduce the percentage of bases
129 annotated as Unknown repeats in all the analysed species by 17-22% (Supplementary Table S3).
130 For instance, in *H. bakeri,* Unknown repeats cover 34.2% of the genome (222 Mb) when using
131 the uncurated library, while only 22.2% (144 Mb) when using the curated library (Fig. 1C). The
132 reduction in bases annotated as Unknown is accompanied by an increase in all the other
133 categories. For example, DNA transposons increase from 9.5% (61 Mb) to 14.7% (96 Mb), LINEs
134 from 14% (91 Mb) to 18.5% (120 Mb) and LTRs from 1.7% (11 Mb) to 4.4% (28 Mb). Although
135 we were able to decrease the span of Unknown repeats in the four species we compared
136 (Supplementary Table S3), it is important to note that Unknown are still the most abundant
137 category of repeats (Fig. 1C), highlighting the value of further curation and characterization efforts.
138
139 The improvement in the classification of repeats is particularly meaningful since we previously
140 highlighted Unknown repeats as relevant contributors of exWAGO guides in *H. bakeri* (referred
141 to as Novel repeats in Chow et al., 2019). By refining repeat annotations, we not only increase
142 the accuracy of TE classification but also provide the foundation to improve our understanding of
143 exWAGO guides.

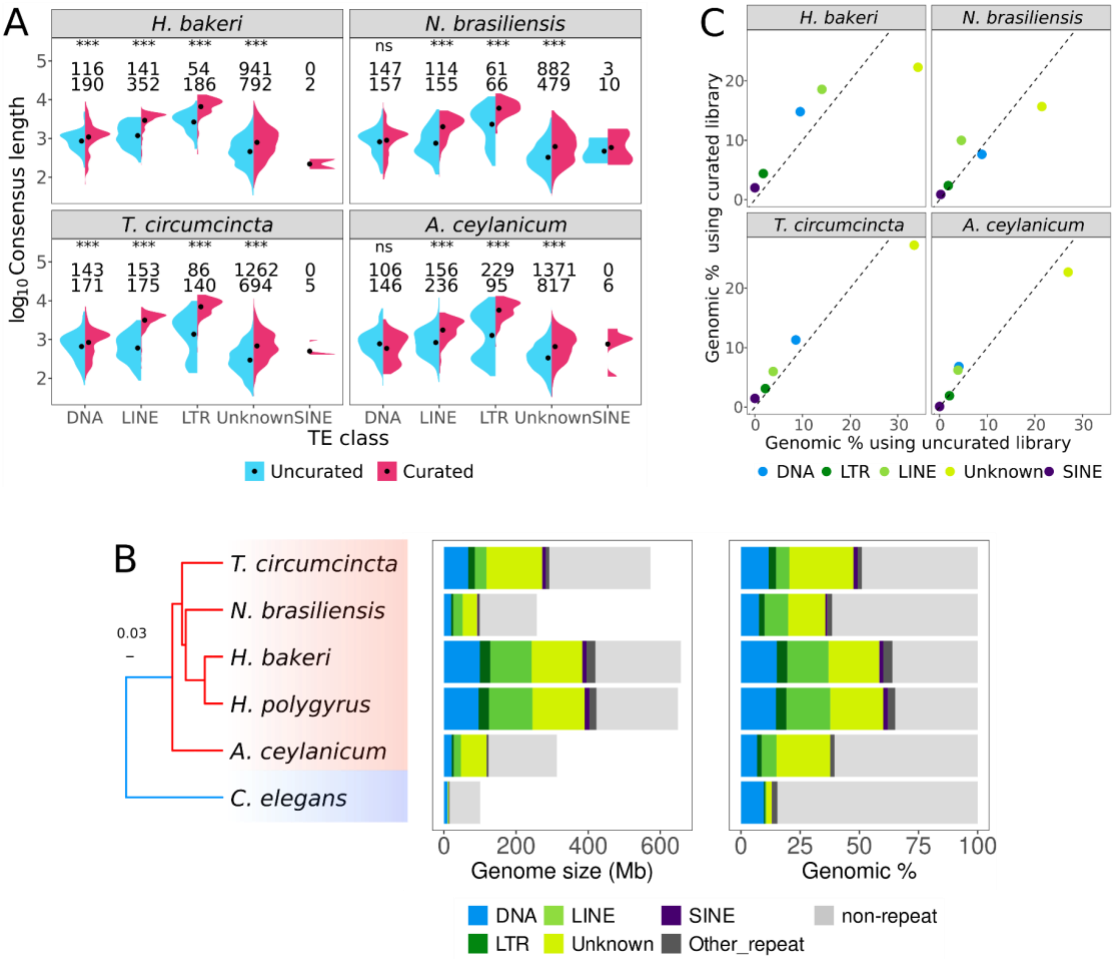


Figure 1. Improved repeat annotations for Strongylida genomes. Comparison of the number and length of uncurated and curated consensus models, by TE category (A). A one-tailed Wilcoxon Rank Sum test was used for comparisons (ns not-significant; * p-value < 0.05). The first/second row of numbers indicates the number of non-redundant uncurated/curated models. Phylogenetic relationship and repeat content in the analysed species based on a maximum likelihood phylogeny inferred with 443 BUSCO genes (parasitic species are highlighted in red and *C. elegans* in blue) (B). The repeat span is shown in megabases (Mb) and as a percentage of genome span (Genomic %). Comparison of the percentage of annotated bases per TE category between curated and uncurated libraries (C). The dotted line denotes x=y.

155 ***Improving the detection of exWAGO-bound sRNAs***

156 To interrogate the sRNA guides bound to exWAGO in different Strongylida parasites, we used
157 sRNA sequencing libraries generated following immunoprecipitation of exWAGO from adult
158 worms. ExWAGO was pulled-down using rat polyclonal anti-exWAGO antibodies raised against
159 recombinant *H. bakeri* exWAGO (for *H. bakeri*, *H. polygyrus*, *N. brasiliensis* and *A. ceylanicum*)
160 or recombinant *T. circumcincta* exWAGO (for *T. circumcincta*) (Neophytou et al., 2025). Western
161 blot analyses show that immunoprecipitations (IP) were successful, but pull-down efficiency
162 varied for the different species (Neophytou et al., 2025). We generated sRNA libraries from the
163 eluate of the IP and total worm lysates using RNA 5' polyphosphatase treatment to allow capture
164 of 5' triphosphorylated sRNAs (Neophytou et al., 2025). To be able to compare against the binding
165 preference of a free-living rhabditid species, we used published IP data for the exWAGO orthologs
166 in *C. elegans*: SAGO-1, SAGO-2 and PPW-1 (Seroussi et al., 2023).
167
168 Sequenced reads from IP experiments represent a subset of the total sRNA population in adult
169 worms (Adult total), showing similar length distribution and genomic origin (Fig. 2A/B). To
170 determine which sRNAs are enriched in exWAGO, we quantified the sRNAs mapping to each
171 annotated genomic region for exWAGO-IP and libraries from adult worm lysates. We then
172 performed a differential expression analysis, comparing IP to Adult total (see Methods). Regions
173 with significantly more sRNAs in the IP, which we call IP-enriched, should represent regions that
174 produce siRNAs specifically bound by exWAGO, while regions with similar amounts of sRNAs in
175 the IP and the input include unbound RNA fragments and those that are non-specifically bound.
176 Initially, we used edgeR with default Trimmed Mean of M-values normalisation (TMM). With this
177 approach, we found 612, 1,591, 978, 1,006, and 3,488 IP-enriched regions for *H. bakeri*, *H.*
178 *polygyrus*, *N. brasiliensis*, *T. circumcincta*, and *A. ceylanicum*, respectively. For *C. elegans* we
179 found 4,216, 13,363 and 1,413 IP-enriched regions for SAGO-1, SAGO-2 and PPW-1,
180 respectively. However, the $\log_2$ Fold-Change (logFC) values show a bimodal distribution, the
181 higher mode likely containing exWAGO-bound sRNA and the lower mode likely representing
182 unbound sRNA (Fig. 2C and Fig. S1A). The default normalisation (TMM) is based on the
183 hypothesis that most regions do not change expression, and in this case centres the higher mode
184 (the likely exWAGO-bound sRNAs) on logFC=zero, since it contains more regions than the lower
185 mode. This suggests that "most regions" are actually producing exWAGO-bound sRNAs, with the
186 unbound regions representing a relative minority. Hence, the default normalisation method (TMM)

187 cancels this biological effect, thereby hindering the detection of true argonaute-associated
188 sRNAs.
189
190 To address this problem, we turned to a “housekeeping” normalisation approach (Bullard et al.,
191 2010). This relies on estimating normalisation factors using a set of housekeeping genes (those
192 expected not to change in expression) so that the logFC values of this subset of genes are centred
193 on zero. In our case, we had to choose sRNAs that are expected not to be bound by argonaute
194 proteins. Thus, we used the sRNAs mapping to the sense-strand of rRNA genes to estimate
195 normalisation factors (see Methods: Differential expression and enrichment analyses of exWAGO
196 and adult total across clade V nematodes). We assume that these rRNA fragments are likely
197 degradation products, and we expect most of them not to be specifically bound by argonaute
198 proteins (Seroussi et al., 2023; Wynant et al., 2017; Zagoskin et al., 2022). To allow for the
199 possibility that some rRNA fragments might be specifically bound and to reduce the effect of
200 outliers, we applied the TMM principle to the rRNA counts to estimate more stable normalisation
201 factors.
202
203 After applying the new normalisation, rRNA genes are consistently centred around logFC=0 (Fig.
204 2D), and we detect 148,185, 96,605, 140,238, 189,676 and 133,022 exWAGO-bound regions as
205 being significantly enriched for *H. bakeri*, *H. polygyrus*, *N. brasiliensis*, *T. circumcincta,* and *A.*
206 *ceylanicum*, respectively. These results suggest that in Strongylida, exWAGO binds sRNAs from
207 a large number of genomic regions. Importantly, the regions that we determine not to be bound
208 by exWAGO show a shift towards miRNA and other ncRNA after housekeeping normalisation
209 (Fig S2). Nonetheless, it is important to notice that even after using rRNA to estimate
210 normalisation factors, some of the rRNA genes do fall within the exWAGO-bound regions, which
211 could represent real rRNA-derived guides or reflect technical issues due to higher background
212 and/or differing IP efficiency among samples.
213
214 Interestingly, when applying the same normalisation to the *C. elegans* orthologs of exWAGO, we
215 do not see as substantial a change in enriched regions for SAGO-1 (4,216 to 3,943) or SAGO-2
216 (13,363 to 14,482), but for PPW-1 the number of enriched regions increases from 1,413 to 42,236,
217 similar to the increase for exWAGO. We further analyzed other *C. elegans* argonaute (AGO) IP
218 data (Seroussi et al., 2023), and we observed a large increase in the number of enriched regions
219 for a few WAGOs, including WAGO-1 and PPW-2 (58,357 and 38,286, respectively), but not for
220 other AGOs such as PRG-1 or ERGO-1 (Fig. S3 and S4). These results align with the expected

221 behaviour of TMM, which centers the expression values where the majority of regions are
222 distributed (Bullard et al., 2010). Therefore, in AGOs that mostly bind to specific types of sRNA,
223 like ALG-1 or PRG-1, which bind miRNAs or piRNAs, respectively, most of the regions will be
224 unbound. In contrast, AGOs which bind sRNAs from a broader number of genomic regions, as
225 often occurs for siRNAs, we expect an increase in enriched regions when applying rRNA-based
226 normalisation.
227
228 Finally, to check that our new analyses were not biased towards an increase in regions while
229 sacrificing the quality of these, we quantified the sRNA abundance of other ncRNA species. We
230 found that most regions annotated as ncRNA family species, such as tRNA, miRNA, snRNA, or
231 snoRNA, remain unbound by exWAGO (Supplementary Table S4). On average, 69.8% of tRNAs
232 remain unbound (SD = 24.4%), 86.1% of miRNAs (SD= 12.2%), 88.1% of snRNA (SD = 2.6%),
233 and 93.5% of snoRNAs (SD = 7.9%). These data indicate that our rRNA-based normalisation
234 results maintain biologically meaningful patterns, since WAGOs in general are not expected to
235 bind these ncRNAs (Seroussi et al., 2023).
236
237 Taken together, our results indicate that exWAGO can bind sRNAs from a wide variety of genomic
238 regions. This is consistent with what is observed for other WAGO proteins in *C. elegans* (Seroussi
239 et al., 2023).

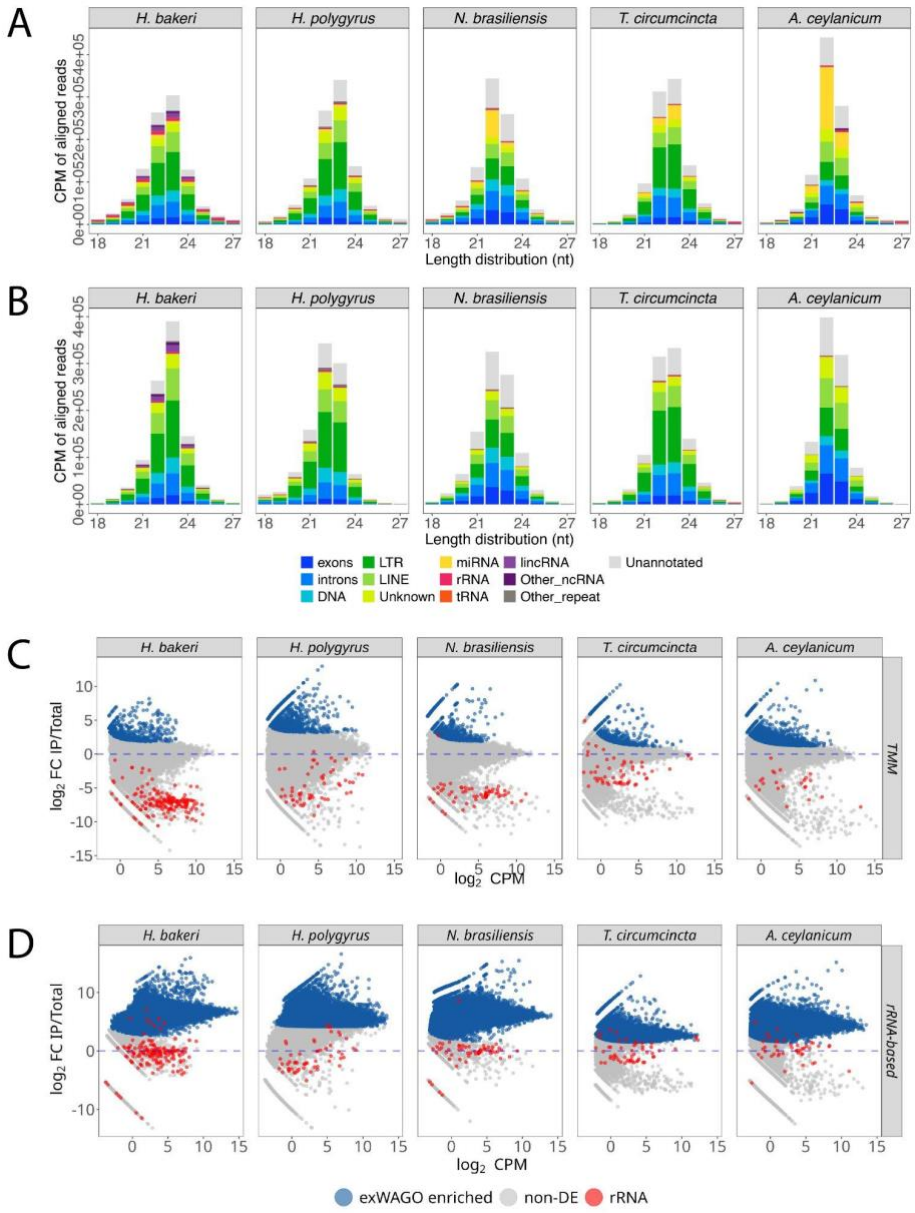


Figure 2. Analysis of exWAGO guides. Length distribution of aligned reads by annotation category for Adult total (A) or exWAGO IP (B) libraries in Strongylida parasites. The x-axis shows the length of aligned sRNAs (18-27 nt), while the y-axis shows average counts per million (CPM) across replicate libraries. MA plots (logCPM vs logFC) after differential expression analysis of exWAGO IP against total libraries using TMM (C) or housekeeping (D) normalisation. Each dot represents a non-overlapping region in each genome. Genomic regions identified as significantly higher in exWAGO IP vs Adult total (FDR < 0.01, see Methods) and interpreted as exWAGO-bound are shown in blue, while non-deferentially expressed regions (non-DE) interpreted as exWAGO-unbound are shown in grey. Regions annotated as rRNA are indicated in red. Dashed lines indicate logFC = zero in each contrast.

251 ***ExWAGO as a global regulator of transposable elements***

252 Since exWAGO-bound sRNAs map to a large number of regions in each genome, we next asked
253 if exWAGO shows specificity for certain types of annotation. For both *Heligmosomoides* species
254 and *T. circumcincta*, 62-71% of the exWAGO-bound sRNA counts are derived from repetitive
255 elements, while for *A. ceylanicum* and *N. brasiliensis,* this value is just under 48% (Fig. 3A). In *C.*
256 *elegans,* this number is further reduced, where out of the three orthologs, only PPW-1 shows
257 more than 14% IP-enriched sRNAs coming from repeats. The opposite trend is observed for
258 protein-coding regions, where the *C. elegans* orthologs show 38-68% of sRNA counts, while in
259 *Strongylida* only 3-12% of their exWAGO-bound sRNAs are derived from protein-coding regions.
260 We further categorised sRNAs based on whether they mapped to the sense or antisense strand
261 of each annotation. Antisense strand-derived sRNAs that map to exons, DNA transposons, LTR
262 retrotransposons, LINEs and lincRNAs, are consistently more abundant than those derived from
263 the sense strand (Fig. S5). This is consistent with WAGO guides being produced by RNA-
264 dependent RNA polymerases RdRps using a primary RNA as a template (Pak & Fire, 2007;
265 Sarkies et al., 2015).
266
267 Notably, in Strongylida parasites, 98-100% of the TE families with detectable sRNAs in the Adult
268 libraries have at least one copy enriched in exWAGO IP. This includes DNA transposons, LTRs,
269 LINEs, SINEs and Unknown repeats, suggesting that exWAGO could regulate most of the
270 expressed TEs in these parasites. Similarly, the *C. elegans* ortholog PPW-1 binds sRNAs from
271 99% of the TE families with detectable sRNA production. Even though SAGO-1 and SAGO-2
272 have a shared evolutionary history with PPW-1, they only bind sRNAs for a reduced set of TE
273 families with detectable sRNAs (15.7% and 16.2%, respectively), suggesting subfunctionalization
274 of these WAGOs, possibly due to differences in tissue localization. Notably, whilst SAGO-1 and
275 SAGO-2 are mainly expressed at the apical membrane of the intestine, PPW-1 shows enrichment
276 for both germline and the apical membrane of the intestine (Seroussi et al., 2023).
277
278 To gain more insight into the diversity of exWAGO guides, we expanded our analysis to other
279 WAGOs in *C. elegans* (Seroussi et al., 2023), where we calculated Shannon entropy and
280 Simpson's diversity indices using a comparable set of 43 genomic features, including exons,
281 introns, DNA transposons, LTRs, LINEs, and different ncRNA families (Supplementary Table S5).
282 We normalized counts using average counts per million (CPM) to account for differences in
283 sequencing depth. In this context, Shannon entropy reflects the evenness in sRNA guide

284 production across different annotation categories, while Simpson's diversity highlights the
285 dominance of specific categories. Highly specialized AGOs such as ALG-1 or PRG-1 show low
286 Shannon entropy (0.004 and 0.3, respectively) and low Simpson diversity (0.001 and 0.11) (see
287 Fig. S6), highlighting low diversity and dominance of specific categories. This is expected since
288 ALG-1 has a high selectivity for miRNAs, and PRG-1 for piRNAs. However, Shannon entropy in
289 *C. elegans* WAGOs ranges from 0.3 to 2.02 (mean=1.3, SD=0.55), whilst Simpson's diversity
290 ranges from 0.09 to 0.79 (mean=0.1, SD=0.79).
291
292 Notably, exWAGO in Strongylida exhibit a significantly higher diversity in comparison to *C.*
293 *elegans*' AGOs (Wilcoxon one-tailed rank-sum test, p-value < 0.001) in both Shannon
294 (mean=2.22, SD=0.11) and Simpson indices (mean=0.84, SD=0.02). This comparison, however,
295 could be biased by the number of regions producing enriched sRNA in these AGO proteins, which
296 is directly related to genome size. Therefore, to account for these differences, we randomly sub-
297 sample the data for each genome to match the smallest number of enriched regions (WAGO-10,
298 2,029 genomic regions), but only considering WAGOs, due to the high selectivity of other AGOs
299 like ALGs or PRG-1. For this, we randomly sampled 100 times each set of enriched regions (Fig.
300 3B). Even using this reduced set, average Shannon and Simpson indices confirm that exWAGO
301 in Strongylida is significantly more diverse than *C. elegans* WAGOs (Wilcoxon one-tailed rank-
302 sum test, p-value < 0.001) (Fig. 3C and D).
303
304 Together, these results suggest that the ancestral exWAGO protein was a versatile argonaute,
305 capable of loading sRNAs derived from a wide variety of sources, including most transposon
306 families present in the ancestral genome as well as a subset of protein-coding and non-coding
307 transcripts. This hypothesis is supported by the conserved capacity of PPW-1 to load TE-derived
308 sRNAs, despite the comparatively reduced diversity (likely due to specialisation) of its paralogs
309 SAGO-1 and SAGO-2. The strong relationship between exWAGO and TE-derived sRNAs,
310 coupled with its diversity in Strongylida, raises the possibility that this AGO protein has been
311 critical for regulating genome stability in species with high transposable element (TE) activity,
312 ultimately shaping genome architecture over a broad evolutionary timescale.

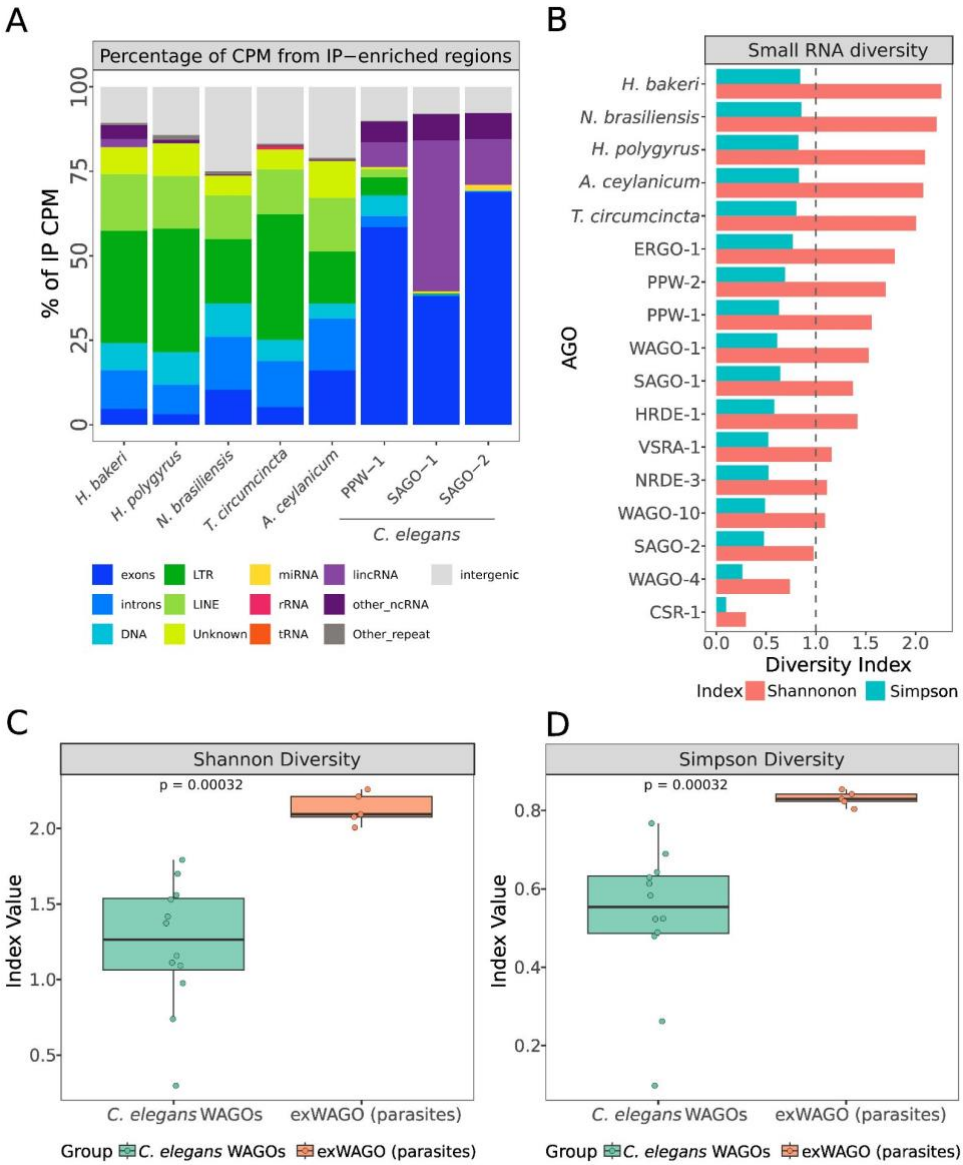


Figure 3. Small RNA diversity in *C. elegans* WAGOs and exWAGO in Strongylida parasites. A comparison of the percentage of Counts Per Million (CPM) of enriched regions per genomic category after differential expression analysis is shown in (A). Small RNA diversity represented by Shanon and Simpson indices in *C. elegans* WAGOs and exWAGO in Strongylida parasites (B). The grey dotted line indicates the maximum value (1) of the Simpson index. Box plots comparing the distribution of Shannon and Simpson indices between *C. elegans* WAGOs and exWAGO in Strongylida parasites (C-D). Panels B–D are based on the average of 100 rarefied random samplings using the number of enriched regions in WAGO-10 (2,029 regions) as the standardized sampling size.

323 ***ExWAGO as an ancestral regulator of LTR retrotransposons***

324 Although exWAGO binds to a highly diverse set of sRNAs in all parasites, this does not rule out
325 a preference for certain sRNA classes over others. To account for the effect of differences in
326 genome size and annotation coverage, we calculated the density of sRNA for each annotation
327 category (see Methods). Although LTRs are not the most abundant repeats (Fig. 4A), antisense
328 LTR-derived sRNAs show the highest density across Strongylida with 6.2-12.6-fold enrichment
329 relative to their genomic span (Fig. 4B). Antisense exon-derived sRNAs are the second highest
330 category, with an average 3.2-fold enrichment in parasites. In contrast, PPW-1 in *C. elegans*
331 shows a preference for antisense lincRNA-derived sRNAs (7.3-fold enrichment), followed by
332 antisense LTR-derived sRNAs (4.6-fold enrichment), with antisense exon-derived sRNAs in third
333 place (3.1-fold enrichment). Interestingly, when antisense pseudogene-derived sRNAs are taken
334 into account (a category we only have for *C. elegans*), they show a similar enrichment as
335 antisense LTR-derived sRNAs (4.8-fold enrichment) (Fig S7). Consistent with previous reports
336 (Seroussi et al., 2023), SAGO-1 and SAGO-2 show a strong preference for antisense lincRNA-
337 derived sRNAs, with a 40.3- and 12.7-fold enrichment, respectively, suggesting the function of
338 these orthologs has diverged from that of PPW-1.
339
340 Certain genomic categories, especially lincRNAs and pseudogenes, are not easy to annotate *de*
341 *novo*, leading them to be underrepresented or absent in genomes other than *C. elegans.*
342 Nevertheless, our results suggest that the common ancestral function of exWAGO is related to
343 LTR retrotransposon regulation, while some protein-coding genes also appear to be regulated in
344 all species. This finding is consistent with previous observations in *H. bakeri*, indicating that the
345 most abundant population of antisense siRNAs at the adult stage are retrotransposon-derived
346 sRNAs and protein-coding-derived sRNAs (Chow et al., 2019). The consistent enrichment of LTR-
347 derived sRNA guides across all species raises the question of whether highly conserved LTRs
348 families, or more recently expanded (lineage-specific families), are the preferred source of
349 exWAGO guides.

350 *Not conservation but structure drives LTR-derived exWAGO guide production*

351 While some of the families involved in exWAGO guide production may be conserved across
352 species, their capacity for transposition and degree of sequence divergence may also influence
353 the levels of produced sRNAs. To assess the influence of evolutionary conservation of TE families
354 on sRNA levels, we first annotated repetitive elements in 21 additional Strongylida and Rhabditina
355 species encoding exWAGO using automatically curated libraries (Fig. S8). As observed in *H.*
356 *bakeri*, *H. polygyrus*, *N. brasiliensis*, *T. circumcincta*, *A. ceylanicum,* and *C. elegans*, LTR
357 retrotransposons are one of the repeat categories with the smallest genomic span across clade
358 V nematodes. We then annotated LTR-encoded proteins and inferred orthology relationships
359 across all identified families using TEsorter (along with HMM models from Rexdb) and
360 OrthoFinder. Using predicted reverse transcriptase (RT) sequences, we defined phylogenetic
361 hierarchical orthogroups (HOGs) and classified LTR families into four categories: highly
362 conserved (present in ≥15 species), conserved (5–14 species), lowly conserved (<5 species), and
363 undetermined (lacking a predicted RT) (Fig. 4C). To quantify sequence divergence, we calculated
364 the relative distance between each LTR copy and its consensus using RepeatMasker. To assess
365 structural completeness, we classified LTRs into three categories based on coding capacity:
366 autonomous (encoding all transposition-related proteins), non-autonomous (missing some
367 domains), and soloLTRs (lacking internal coding regions). Finally, to account for differences in
368 element size due to coding capacity or divergence, we normalized sRNA abundance by copy
369 length and library size using reads per kilobase per million (RPKM).
370
371 Having classified LTR retrotransposons by conservation, structural variation and estimated their
372 divergence, we fitted a generalized linear mixed model. We used conservation (based on HOGs)
373 and structural completeness (based on the presence of transposition-related proteins) as
374 categorical fixed effects, while divergence as a continuous fixed effect. To account for the
375 variation that different families can have due to differences in activity, we used family as a random
376 effect. Although we initially fitted strand as a random effect, diagnostic plots show that this variable
377 introduces systematic bias into the model, suggesting a confounding effect. Since there is a strong
378 bias towards antisense sRNAs among exWAGO guides (Fig. S5), we used the $\log_2$ mean-centred
379 RPKM of antisense LTRs as the response variable.

380

381 After assessing model fit across species—finding no major deviations from normality (Fig. S9)—
382 we observed that, on average, the strongest and most consistent positive effect on sRNA
383 production was associated with soloLTRs (relative to non-autonomous; mean $\beta = 0.42$ and
384 $SE = 0.09$) (Fig. 4D) (Supplementary Table S6). In contrast, autonomous LTR retrotransposons
385 only show a positive but non-significant effect in *A. ceylanicum* ($\beta = 0.35$, $SE = 0.33$, p-value =
386 0.27). Notably, the preference for soloLTR-derived sRNAs remains consistent even when
387 compared separately to either the internal region or the LTR region of full-length LTR
388 retrotransposons (Fig. S10). Meanwhile, high conservation shows a positive effect on sRNA
389 production in some species (compared to lowly-conserved; $\beta = 0.39$, $SE = 0.42$, on average),
390 however, there is greater uncertainty around this estimate, and in general, the effect is not
391 significant. Finally, the effect of divergence seems to be modest ($\beta = -0.003$, $SE = 0.002$, on
392 average) and negatively related to sRNA production, suggesting that less divergent copies (more
393 recent insertions) produce slightly more sRNAs.

394

395 Together, our results suggest that the ancestral function of exWAGO in clade V nematodes is
396 related to LTR retrotransposon regulation. Furthermore, despite being non-autonomous,
397 soloLTRs are relevant for exWAGO guide production, suggesting that retaining these genomic
398 remnants may represent an adaptive advantage either for enhancing TE regulation in *trans* to
399 silence active elements of the same family or potential alternative functions of exWAGO. Notably,
400 our results also suggest that similar to the exWAGO protein, the sources of sRNAs are also
401 ancestrally conserved. These results indicate that exWAGO has been adapted to regulate a set
402 of commonly active LTRs across clade V nematodes.

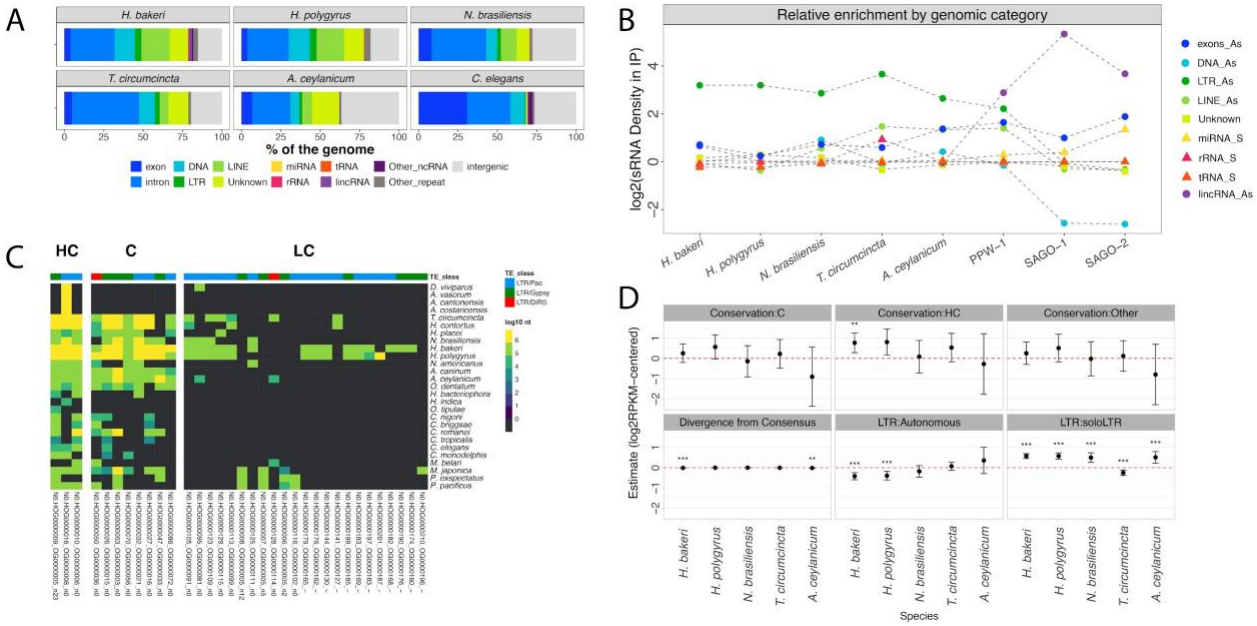


Figure 4. **LTR enrichment in exWAGO guide production**. Comparison of genome annotations in Strongylida parasites and *C. elegans* (A). Comparison of sRNA density of genomic regions enriched in exWAGO guide production per genomic category (B). The X-axis represents each species, while the Y-axis shows density values estimated as the $\log_2$ of the ratio of the percentage of CPM of IP-enriched regions and the percentage of genomic bases annotated for each of the categories. Phylogenetic orthology inference analysis of LTRs based on the reverse transcriptase (RT) protein (C). The top colour bar indicates LTR classification at the superfamily level. The heatmap shows the genomic abundance of the LTR in $\log_{10}$ of bases per HOG. Estimates for the fixed effects from the generalized linear mixed model (D). The red dashed line indicates 0 or a neutral effect (* p-value < 0.05, ** p-value < 0.01 and p-value < 0.001).

415 ***Structural differences among LTRs associated with sRNA secretion pathways***

416 We recently showed that exWAGO can be secreted by *H. bakeri* in two forms (inside and outside
417 of EVs) and that exWAGO is internalized by host cells *in vivo* (Neophytou et al., 2025). Given that
418 EVs containing exWAGO can induce host immunosuppression (Buck et al., 2014), and non-
419 vesicular exWAGO can also enter host cells (Neophytou et al., 2025), we sought to identify the
420 structural features of LTRs that influence the secretion of exWAGO-bound sRNAs. To tackle this,
421 we used vesicular and non-vesicular exWAGO IP libraries from secreted materials.
422
423 We performed differential expression analyses to compare the vesicular or non-vesicular
424 exWAGO libraries against the adult IP libraries (see Methods), including the structural-based
425 classification of LTR retrotransposons. Using this approach, we identified a set of 17,950 and
426 22,548 genomic regions enriched in vesicular and non-vesicular exWAGO, respectively (Fig. 5A
427 and 5B).
428
429 To determine if LTR retrotransposons involved in secreted exWAGO guide production differ from
430 those that produce sRNA guides that remain inside the worm, we fitted binomial generalised linear
431 models using the preference for secreted or retained exWAGO guide production as the response
432 variable in the vesicular or non-vesicular contrast against adult, encoding regions enriched in
433 secreted exWAGO as 1 and the remaining regions as 0. As explanatory variables, we used the
434 previously defined degree of family conservation and structural classification. Moreover, we
435 assessed length (centred on the average $\log_{10}$ length) to account for differences between different
436 types of LTR retrotransposon structures. There were no significant deviations of model
437 assumptions in simulated residuals for both models (Fig. S10 A and B).
438
439 The results of our model, for the vesicular exWAGO form, indicate that the odds of sRNA secretion
440 increase on average 2.24 times when sRNAs are produced from soloLTR (β=0.80, SE=0.08) and
441 1.92 from the LTR region (β=0.66, SE= 0.07), relative to the internal coding region of an LTR,
442 while holding the rest of the variables constant at their base levels ($p < 0.001$) (Fig. 5C and
443 Supplementary Table S7). Moreover, we found that the odds of sRNA secretion from lowly
444 conserved families (β=0.34, SE=0.09) or undetermined conservation (β=0.66, SE=0.07) due to
445 the lack of detectable RT, significantly increase on average 1.42 and 1.95 times, respectively,
446 compared to highly conserved families ($p < 0.01$). Notably, the mean-centred $\log_{10}$ length has a
447 non-significant and negative effect (β=-0.01, SE=0.07, p-value < 0.84), supporting the preference

448 for localised short regions of LTR retrotransposons, such as the LTR region or soloLTRs, as
449 sources for vesicular-exWAGO guide production.
450
451 Our results for the non-vesicular form show that the odds of secretion increase 1.54 times when
452 sRNAs are produced from LTR regions ($\beta = 0.43$, SE = 0.07, $p < 0.001$). Nevertheless, this
453 preference is not from soloLTRs as with vesicles, but from longer albeit incomplete copies (Fig.
454 5C and Supplementary Table S8). Regarding conservation, families annotated as conserved (β
455 $= 0.52$, SE = 0.06) and undetermined ($\beta = 1.52$, SE = 0.08) are 1.68 and 4.56 times more likely
456 ($p < 0.001$ in both cases), compared to highly conserved families, whereas lowly conserved
457 families show no significant association. Unexpectedly, we also found that length is positively
458 associated with non-vesicular secretion ($\beta = 0.62$, SE = 0.06), corresponding to a 1.86 increase
459 in odds for every 10-fold increase in length. The overall results indicate that while the vesicular
460 exWAGO prefers sRNAs from soloLTRs, the non-vesicular form associates sRNAs from longer
461 non-autonomous elements. Consistently, both vesicular and non-vesicular exWAGO are
462 associated with LTR families with undetermined conservation (conservation Other), due to the
463 lack of RT (Fig. 5C). Therefore, the lack of autonomy for retrotransposition increases the odds
464 for serving as a source for secreted exWAGO guides.
465
466 To exemplify regions which produce significantly more secreted exWAGO guides, we highlight a
467 soloLTR on chromosome IV with higher sRNA production (on average 14.44 $\log_2$ CPM) for the
468 vesicular form (Fig. 5D), and a fragmented LTR retrotransposon on chromosome X (on average
469 11.8 $\log_2$ CPM) for the non-vesicular form (Fig. 5E). Both regions show clear differences in LTR
470 structures, which are consistent with the overall preferences we found using binomial generalized
471 linear models, with the vesicular form showing a preference for soloLTR-derived sRNAs, and the
472 non-vesicular form for longer but non-autonomous elements.
473
474 Collectively, our results delineate two selective routes for the source of secreted exWAGO guides:
475 vesicles enriched in lowly conserved, and degenerate soloLTRs, while the non-vesicular form
476 prefers longer elements with intermediate conservation. Notably, in both secreted exWAGO
477 forms, LTR-derived sRNAs are produced from non-autonomous elements, suggesting that once
478 LTR retrotransposons lose their autonomous transposition capacity, they can be co-opted to act
479 as the source of secreted sRNAs.

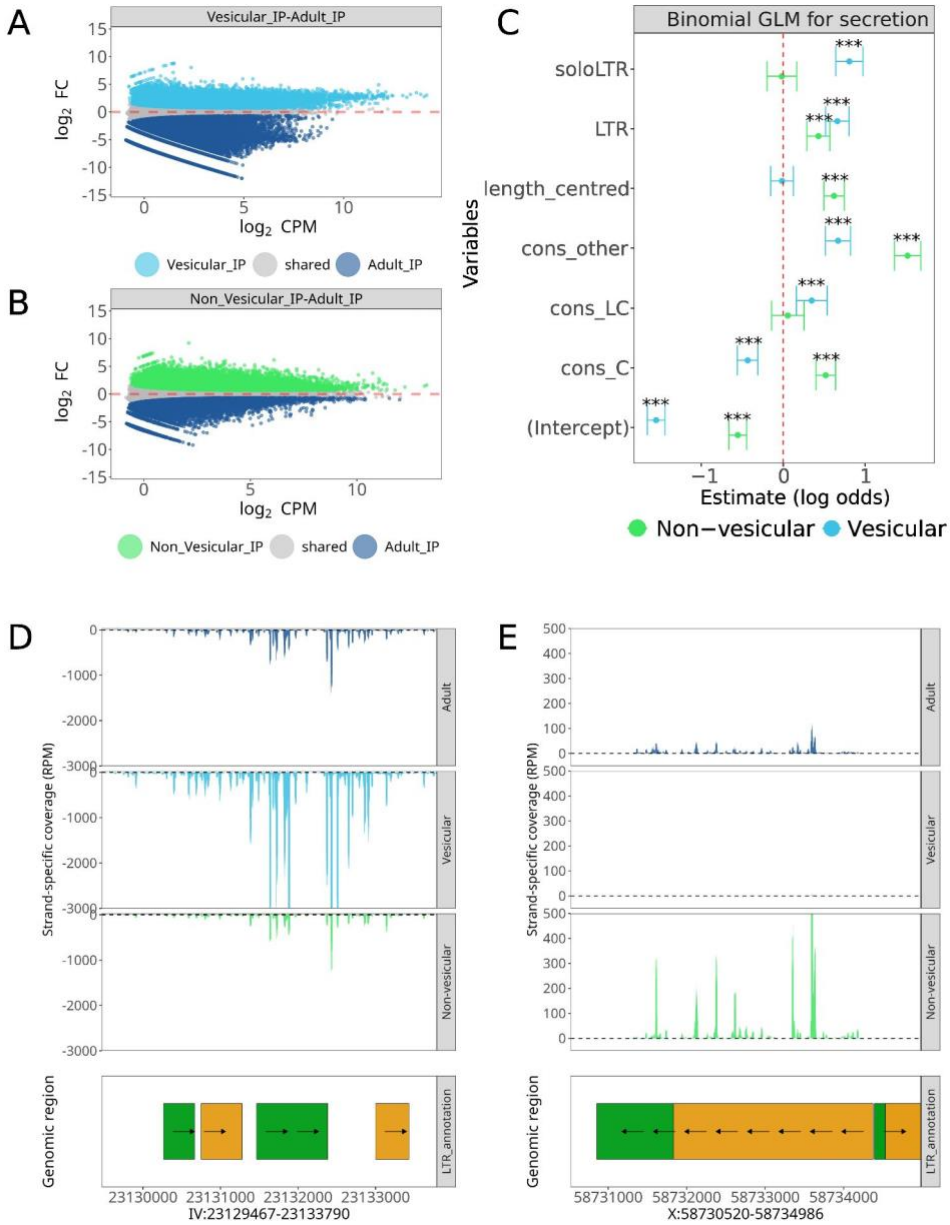


Figure 5. Vesicular and non-vesicular exWAGO guides derive from different types of non-autonomous LTR retrotransposons. MA plots (logCPM vs logFC) after differential expression analysis (FDR < 0.01) of vesicular exWAGO IP against exWAGO IP in adult worms (A), and non-vesicular exWAGO IP against Adult IP (B). The red dotted line indicates $\log_2$FC= zero. Estimated coefficients for binomial generalized linear models testing exWAGO LTR-derived small RNA secretion in vesicular and non-vesicular forms (C). Examples of vesicular (D) and non-vesicular (E) enriched regions showing sRNA production in Adult exWAGO IP (dark blue), vesicular exWAGO IP (light blue) and non-vesicular exWAGO IP (green). Genomic region tracks show LTR annotation (green for the LTR region, yellow for the internal coding region).

490 # Discussion

491 To mitigate the potential deleterious effects of TE activity, organisms have adapted molecular
492 components, including argonaute proteins and their sRNA guides to recognize and silence TEs.
493 Most of what we know about TE regulation in nematodes stems from their detailed annotation
494 and functional characterization in *C. elegans*; however, much less is known about these elements
495 and their regulation in parasitic nematodes.
496
497 Here, using manually curated TE annotations in multiple Strongylida parasites, we characterize
498 the sRNA preferences of the extracellular worm argonaute exWAGO in an evolutionary context.
499 Our improved TE annotations, combined with refined differential expression analysis, revealed
500 exWAGO as a highly versatile AGO, but with an ancestral function linked to LTR retrotransposon
501 regulation. Although exWAGO guides in adult Strongylida parasites preferentially originate from
502 non-autonomous soloLTRs, the conservation of their corresponding TE family does not appear to
503 influence the levels of sRNAs produced.
504
505 Interestingly, exWAGO preference for soloLTR-derived sRNAs extends to extracellular vesicles
506 (EVs) in *H. bakeri*. However, secreted non-vesicular exWAGO shows a particular preference for
507 sRNAs derived from longer and less degenerate elements. The structural differences between
508 LTRs related to the two secreted exWAGO forms shape the composition of sRNAs delivered to
509 the parasite's host. The recurrent lack of detectable reverse transcriptase –and, therefore,
510 autonomy for retrotransposition–of these elements indicates a consistent pattern of prior decay
511 followed by co-option as a source of secreted sRNAs.

512 ### *Addressing compositional bias in Argonaute IPs: rRNA-guided normalisation*

513 In nematodes, worm-specific argonautes (WAGOs) load sRNAs with specific properties, typically
514 ~22 nt in length starting with a 5' triphosphorylated Guanosine (22G sRNAs). Immunoprecipitation
515 and sequencing of bound sRNAs for AGOs that associate sRNAs from a broad set of loci, such
516 as CSR-1, WAGO-4 in *C. elegans* or exWAGO in Strongylida parasites presents certain
517 challenges. Standard normalisation methods such as TMM rely on the assumption that most loci
518 are not differentially expressed across samples (Robinson & Oshlack, 2010). This assumption
519 holds in experimental designs where only a subset of genes is expected to change. However, it
520 has been previously shown that when most of the genes in a sample are differentially expressed,

521 this assumption is invalid (Bullard et al., 2010; Y. Zhou et al., 2017), for example, in RNA-seq
522 experiments comparing tissues with clear differences in transcriptional activity.
523
524 Here, by estimating normalisation factors based on the adjusted expression of rRNA fragments,
525 we were able to improve the detection of genomic regions enriched in exWAGO guide production.
526 We used rRNA because, in general, we do not expect WAGOs to associate with this type of sRNA
527 (Seroussi et al., 2023; Zagoskin et al., 2022). Although previous reports in humans and mice have
528 shown that rRNA fragments can be enriched in AGOs (Guan & Grigoriev, 2021), neither of our
529 analyses in *C. elegans* WAGOs nor exWAGO show general patterns of rRNA enrichment in IPs.
530 Other elements could be used for normalisation, for example, miRNAs could be employed for
531 WAGOs that do not bind them. However, using rRNA allows the same approach to be used across
532 all AGOs. This supports the use of rRNA-derived sRNAs as a stable internal reference for
533 normalisation in this context.
534
535 Although this approach reduces the false negative rate introduced by standard TMM, we caution
536 that this strategy is not perfect, since different amounts of rRNA fragments could originate in
537 samples with different levels of degradation or with higher background levels. Careful inspection
538 of the resulting MA plot is still recommended to verify that the effect of normalisation agrees with
539 biological assumptions. In most of our samples, this normalisation worked well, but our analysis
540 in *T. circumcincta* indicates higher levels of rRNA fragments in the IP, potentially causing an
541 underestimation of enriched regions, leading to more conservative results than for other species.
542 In future work, we could consider using controlled spike-ins, in both IP and input samples to
543 evaluate alternative normalisation strategies. Nevertheless, these problems do not undermine the
544 broader applicability of our normalisation strategy, nor the robustness of our biological
545 conclusions, which are supported by consistent patterns of TE enrichment across Strongylida
546 parasites.
547

548 ***The dual role of exWAGO and soloLTR-derived sRNAs in Strongylida parasites***

549 Despite their relatively low genomic span in clade V nematodes, LTR retrotransposons have been
550 linked to key processes such as genome expansion, innate immune responses, and adaptation
551 to environmental stress (Ansaloni et al., 2019; Kanzaki et al., 2018; Ni et al., 2016). In *C. elegans*,
552 the argonaute protein HRDE-1 has been implicated in LTR regulation under heat shock conditions
553 (Ni et al., 2016), and ERGO-1 has also been shown to target LTR elements (Fischer & Ruvkun,

554 2020). Notably, the loss of ERGO-1 in *Caenorhabditis inopinata* correlates with a marked
555 expansion of LTR retrotransposons in that species (Kanzaki et al., 2018). Remote homology
556 searches across predicted proteomes of Strongylida parasites revealed that ERGO-1 is absent
557 (Supplementary Table S9). While we cannot exclude the possibility that other argonautes in
558 Strongylida bind LTR-derived sRNAs, our enrichment analyses demonstrate that exWAGO
559 orthologs—from *C. elegans* PPW-1 to parasitic Strongylida species—consistently associate with
560 LTR-derived small RNAs. This suggests that exWAGO retains an ancestral role in LTR
561 surveillance. But why does exWAGO show this remarkable preference for LTR-derived sRNAs in
562 Strongylida parasites? An increased activity of LTR retrotransposons can potentially lead to
563 genomic instability. In this regard, the preference for soloLTR-derived sRNAs in adult worms may
564 represent an advantage, because by retaining transcriptional activity, they can act as a safe
565 source for sRNAs to potentially regulate full-length elements of the same family. Although not
566 identical, similar mechanisms are known in *Drosophila* species where genomic loci like the
567 flamenco cluster retain soloLTRs that act as the source of piRNAs (Van Lopik et al., 2023), which
568 in turn can modulate active LTR retrotransposons in *trans*. These parallels raise the possibility
569 that nematodes may have independently evolved decayed LTRs as safe templates of small RNA
570 biogenesis for TE regulation.
571
572 In addition, the immunogenic capacity we observed for *H. bakeri* EVs and the non-vesicular
573 exWAGO (Buck et al., 2014; Neophytou et al., 2025), hints at a potential role of LTR-derived
574 sRNAs in modulating host immunity. In this context, the preferential export of sRNAs derived from
575 non-autonomous LTR elements may represent a mechanism by which parasites exploit genomic
576 fossils or degenerate elements to mediate cross-species communication. Notably, although we
577 observed that elements of undetermined conservation are consistently preferred for secretion,
578 one limitation of our analysis is that conservation was defined based solely on the reverse
579 transcriptase protein. Our results are consistent in that both secreted exWAGO forms show a
580 preference for non-autonomous LTR elements. We hypothesise that these genomic fossils
581 provide a stable, low-risk source of guides that can be safely exported without fueling
582 retrotransposition. However, further work to identify host targets and test functional consequences
583 of these secretion pathways will clarify how widespread and adaptive this strategy is for nematode
584 parasitism.
585
586 Intriguingly, beyond parasitic nematodes, fungal pathogens such as *Botrytis cinerea* also rely on
587 LTR-derived sRNAs as a source of extracellular RNA with immunomodulatory potential (Porquier

588 et al., 2021). While *B. cinerea* preferentially uses full-length LTRs as the source of secreted RNAs,
589 nematodes like *H. bakeri* show a preference for non-autonomous LTRs. This contrast suggests
590 that the evolutionary paths and molecular strategies shaping extracellular RNA repertoires may
591 differ across kingdoms, but highlight a potential convergent source of sRNAs for extracellular RNA
592 communication.
593
594 Together, our results support a model in which exWAGO maintains genome integrity through
595 broad TE regulation while simultaneously co-opting fragmented or non-autonomous LTR
596 retrotransposons as platforms for regulatory sRNA production. This dual role positions exWAGO
597 at the interface of genome defence and parasite-host communication and reveals how LTR-
598 derived sequences can be co-opted for extracellular signalling in parasitic nematodes.

599 Methods

600 *Transposable element annotation*

601 We produced *de novo* repeat annotations using earlGrey v.4.0.3 (Baril et al., 2024), with the
602 Nematoda Dfam library (Dfam 3.7), and ten iterations of the "BLAST, Extract, Extend" process.
603 For *H. bakeri* and *H. polygyrus*, we used our previously published curated libraries (Stevens et
604 al., 2023), with some minor updates.
605
606 Redundancy within the predicted libraries was resolved by collapsing models with at least 80 nt
607 length and sharing 80% of nt identity in at least 80% of the model to the centroid model (the
608 longest one), using cd-hit-est (W. Li & Godzik, 2006). To remove non-TE-related proteins, getorf
609 from the EMBOSS package was implemented to obtain ORFs with a minimum size of 300 aa, to
610 subsequently annotate protein domains using PfamScan
611 (http://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/) with the Pfam database (RELEASE 35). Models
612 with hits to non-TE-related proteins were filtered out. Furthermore, TEsorter 1.4.6 was used to
613 identify TE-related protein domains (Zhang et al., 2022). Unknown models, with hits for RT,
614 ENDO, and INT, were further classified as LINE or LTR based on TEsorter results. Moreover, TE-
615 Aid (available at https://github.com/clemgoub/TE-Aid) was used to identify the structural domains
616 of predicted models. We classified them as DNA/MITE models when showing only TIR structures
617 in self-dot plots and less than 1 kb (Goubert et al., 2022).
618

619 To have insights into the evolution of transposable elements in nematodes where exWAGO can
620 be secreted, we manually curated the models of *N. brasiliensis* and *T. circumcincta*. For *A.*
621 *ceylanicum*, we only manually curated LTR retrotransposons, and the models required to
622 annotate 80% of the total repeat content. For this, we retrieved up to 50 instances of each
623 element, extending 1000 nt of each edge, using the make_fasta_from_blast.sh (with the
624 parameters 0 and 1000) and ready_for_MSA.sh (with the parameters 50 and 40) scripts.
625 (https://github.com/annaprotasio/TE_ManAnnot). The resulting sequences were aligned using
626 mafft v7.520, and the alignments were inspected using AliView v1.28. We identified the canonical
627 TGT and CAC motifs at the beginning of LTRs and TAT for DNA transposons to trim divergent or
628 poorly aligned columns (Goubert et al., 2022). For models where the edges were highly
629 degenerate, we used CIAlign (Tumescheit et al., 2022) with the parameters: --insertion_max_size
630 50 --crop_ends --crop_divergent --crop_divergent_min_prop_ident .5 --
631 crop_divergent_min_prop_nongap .6 --crop_divergent_buffer_size 15 --make_consensus.
632 Consensus models from the resulting alignments were built using CIAlign with the
633 make_consensus function.
634
635 Resulting libraries were then used to produce annotations using RepeatMasker with the
636 parameters (-s -cutoff 400 -a). Since RepeatMasker produces fragmented annotations, the
637 RepeatCraft pipeline with the Loose mode (Baril et al., 2024; Wong & Simakov, 2019) from
638 earlGrey was used to collapse fragmented annotations. Overlapping bases between repeats were
639 assigned using the 01.assign_TE_overlaps.R (available at
640 https://github.com/imu93/ms_evo_exwago). Briefly, elements belonging to the same superfamily
641 were collapsed into a single element using the reduce function from the GenomicRanges R
642 package (Lawrence et al., 2013). If overlapping elements belonged to a different family, the family
643 of the collapsed segment was assigned based on the longest element. Moreover, overlapping
644 bases of elements of different super-families were assigned based on a hierarchical classification.
645 These bases were assigned to the element that belongs to the super-family with more bases in
646 the genome.

647 ***Benchmark of curated TE libraries***

648 TE library benchmarking was conducted using the get_family_summary script (available at:
649 https://github.com/jmf422/TE_annotation/blob/master/get_family_summary_paper.sh). This
650 script classifies consensus sequences of a tested library, based on their degree of sequence
651 similarity, coverage and divergence, relative to a curated library. The classification consists of four

652 categories, perfect (>95% of sequence similarity, >95% of coverage and <5% of divergence),
653 good (same as perfect but up to <10% of divergence), present (>80% of coverage), and missed
654 (no hits or <80% of coverage).
655

656 ***ncRNA gene annotations***

657 ncRNA annotation was carried out using several homology-based tools. We used infernal v1.1.4
658 and Rfam v.14.10 covariance models to annotate known ncRNA families (--rfam --cut_ga --
659 nohmmonly). In addition, tRNAscan-SE (-Q -E --score 40) (v1.4) and RNAmmer (-S euk -m
660 tsu,lsu,ssu) (v1.2) were used to annotate tRNAs and rRNAs, respectively. miRNA annotation was
661 conducted using the precursors of miRNAs for nematodes in miRBase (v.22). After obtaining
662 precursors, blastn was used (-evalue 1e-5 -perc_identity 90 -qcov_hsp_perc 90). In addition,
663 miRNA annotation of *N. brasiliensis*, *T. circumcinta* and *A. ceylanicum* was enhanced by using
664 data from regular sRNA-seq libraries (NCBI Bioproject PRJNA1200757) using MirDeep2
665 (Friedländer et al., 2012).
666
667 Meanwhile, piRNAs were annotated only in a subset of species. The annotation of this ncRNA
668 family was conducted using bowtie1 (-f -a --best --strata -v 0 -m 3 -S) and previously identified
669 type 1 piRNAs in *H. bakeri*, *N. brasiliensis* (Beltran et al., 2019). Given the close phylogenetic
670 relationship between *H. bakeri* and *H. polygyrus*, we used *H. bakeri* as a reference to annotate
671 putative conserved piRNAs in *H. polygyrus*. Using a similar approach, we annotated piRNAs in a
672 chromosome-level assembly of *N. brasiliensis*.
673
674 LncRNAs and lincRNAs were annotated using publicly available RNA-seq data in *H. bakeri*, *N.*
675 *brasiliensis*, *T. circumcincta*, and *A. ceylanicum* (*H. bakeri*: PRJNA750155; *N. brasiliensis*:
676 PRJEB16076, PRJEB20824, SRR25993873; *T. circumcincta*: PRJEB7677; *A. ceylanicum*:
677 PRJNA231490). Briefly, we aligned RNA-seq reads to their respective genome using STAR
678 v.2.7.10b using the gtf gene annotation file to define junctions. The resulting bam files were sorted
679 using samtools sort (Li et al., 2009) to build transcriptomes with StringTie v.2.2.1 (Pertea et al.,
680 2015). We evaluated the transcript coding potential using FEELnc_filter, cpc2, lgc, and PfamScan
681 (Finn et al., 2014; Kang et al., 2017; G. Wang et al., 2019; Wucher et al., 2017), to filter out
682 transcripts produced from protein-coding genes. To classify lncRNAs and lincRNAs,
683 FEELnc_classifier was used (Wucher et al., 2017). In addition, lncRNAs and lincRNAs were

684 annotated in *H. polygyrus* using blastn with the predicted lncRNA and lincRNAs in *H. bakeri* as
685 reference (-evalue 1e-10 -perc_identity 90 - qcov_hsp_perc 90).

686

687 ***Segmented annotations***

688 Aiming to understand the contribution of different genomic categories to sRNA production, we
689 built non-overlapping annotations in which each base in the genome has a single annotation. We
690 assigned overlapping regions based on a hierarchical classification using the setdiff function from
691 the GenomicRanges R package (Lawrence et al., 2013). This annotation includes both sense and
692 antisense annotations for each genomic region. Our hierarchy prioritises known sRNA families (in
693 the following order: miRNA, piRNA, yRNA, rRNA, tRNA, snRNA, snoRNA, lincRNA, lncRNA),
694 then TEs (TE families were ranked based on their genomic span), exons, introns, and finally, in
695 the last part of the hierarchy we included unstranded repeats such as MITEs, SINEs, Low
696 complexity repeats, Simple repeats, and Unknown repeats, in that order.
697

698 ***exWAGO immunoprecipitations and RNA isolation from worms***

699 To interrogate the sRNAs associated with exWAGO in different nematode species, including *H.*
700 *bakeri*, *T. circumcincta*, *N. brasiliensis* and *A. ceylanicum* we used previously published sRNA
701 library datasets (NCBI Bioproject PRJNA1200757) that were generated by using the eluate and
702 unbound fraction of exWAGO immunoprecipitation from adult worms. Here we generated sRNA
703 libraries from wild caught, naturally *H. polygyrus*-infected wood mice (*Apodemus sylvaticus*) from
704 woodlands in Midlothian, Scotland (e.g. Penicuik Estate (55°49'N 3°15'W) and Hewan Woods,
705 (55°52'N 3°08'W). To immunoprecipitate exWAGO from these samples, ~1.5 cm of the duodenum
706 was harvested and flash frozen immediately in liquid nitrogen to minimize RNA degradation, and
707 stored in -70ºC until required. Protein G beads (ThermoFisher, 10003D) were washed five times
708 using cold Binding Wash Buffer (PBS, 0.02% Tween-20) and polyclonal rat anti-exWAGO serum
709 antibodies (Neophytou et al., 2025) were conjugated on the prepared beads by incubation (2h,
710 rotating wheel, 4°C). Unconjugated antibody was then removed, and the beads were equilibrated
711 using cold lysis buffer (150 mM NaCl, 10 mM Tris.HCl, 0.5 mM EDTA, 0.5% NP40) containing
712 protease inhibitors (1 tablet per 5 ml) (Roche, 11873580001) three times. Then, the frozen tissue
713 was ground into a fine powder using a pestle and mortar under liquid nitrogen and then lysed in
714 1 ml of cold lysis buffer as above supplemented with 200 U/ml RNase inhibitors (Promega,
715 N2515). Unlysed material was removed by centrifugation (16,100 rcf, 10 min, 4°C), and the

716 supernatant lysate was incubated with the antibody-conjugated beads (45 min, rotating wheel,
717 4°C). The beads were washed with cold Low Salt buffer (50 mM Tris.HCl pH 7.5, 300 mM NaCl,
718 5 mM MgCl2, 0.5% NP40 and 2.5% glycerol), followed by two washes with cold High Salt buffer
719 (50 mM Tris.HCl pH 7.5, 800 mM NaCl, 10 mM MgCl2, 0.5% NP40 and 2.5% glycerol) (5 min,
720 rotating wheel, 4°C). The beads were then washed once with cold Low Salt buffer, followed by a
721 cold PNK buffer wash (50 mM Tris.HCl pH 7.5, 50 mM NaCl, 10 mM MgCl2, 0.5% NP40). For
722 small RNA sequencing analysis, the RNA was eluted directly in Qiazol (700 µl, 5 min, room
723 temperature) (Qiagen) and stored -70°C until required.
724
725 For the adult worm *H. polygyrus* sRNA libraries, RNA was isolated from wild-derived *H. polygyrus*
726 harvested from lab-reared wood mice at day 14 post-challenge. This outbred colony of wood mice
727 was originally wild-caught, but is now maintained in standard laboratory conditions at the
728 University of Edinburgh (see Sweeny et al., 2021 for more details). Worms were washed in PBS
729 and then lysed in 700 µl of Qiazol by bead beating using 5 mm steel beads (Qiagen, 69989) using
730 the Tissue Lyser II (Qiagen) in pre-cooled cartridges for 2 minutes at 30 Hz twice. The RNA was
731 then processed for sequencing as described below.
732

733 ***Small RNA libraries***

734 sRNA libraries were generated as described in Neophytou et al. (2025). Briefly, RNA in Qiazol
735 was spiked with 7 µl of 10 pM RT4 synthetic spike (CTTGCGCAGATAGTCGACACGA). The RNA
736 was then extracted using the miRNA Serum/Plasma kit (Qiagen, 217184) and treated with RNA
737 5' Polyphosphatase (Lucigen, RP8092H) according to the manufacturer's instructions. The
738 reaction was terminated using ethanol precipitation (-70°C, overnight), and the precipitated RNA
739 was resuspended in 2.5 µl of nuclease-free water and 2.5 µl of Buffer 1 (TriLink, L-3206). The
740 sRNA libraries were prepared using the CleanTag Small RNA Library Preparation Kit (TriLink, L-
741 3206) using half reaction volumes. The *H. polygyrus* immunoprecipitation or total adult worm
742 samples were generated using 1:10 or 1:12 dilution of 5' and 3' adapters, respectively. All libraries
743 were generated using 20 amplification cycles, and the profile of the libraries was assessed using
744 the High Sensitivity DNA Bioanalyser chip (Agilent, 5067-4626). Pooled libraries were size-
745 selected (140-180 bp for immunoprecipitation; 140-220 bp for total adult worms) using gel
746 purification to remove adapter dimers. The libraries were sequenced on the Illumina NextSeq
747 2000 platform by the Edinburgh Clinical Research Facility using single-end 100 base pair reads.
748

749 ***Small RNA: processing, alignment and quantification***

750 In addition to the small RNA-seq libraries we produced for *H. polygyrus*, we obtained small RNA-
751 seq libraries from NCBI SRA for *H. bakeri* (NCBI bioprojects: PRJNA481340 and
752 PRJNA1200757)*, N. brasiliensis* (PRJNA1200757), *T. circumcincta* (PRJNA1200757), and *A.*
753 *ceylanicum* (PRJNA1200757), and *C. elegans* (PRJNA860633). Illumina small RNA adapters
754 were removed using reaper (Davis et al., 2013). FastQC and MultiQC were used to evaluate read
755 quality (Ewels et al., 2016; Andrews & Felix, 2012). For *T. circumcincta,* we collapsed technical
756 replicates for exWAGO IP libraries, however, due to the lack of biological replicates, technical
757 replicates for adult total libraries were treated as independent samples. A similar approach was
758 used for *A. ceylanicum*. Using Pullseq v.1.0.2 (available at https://github.com/bcthomas/pullseq),
759 we then retrieved reads between 18 to 27 nucleotides for downstream analyses. ShortStack
760 v.3.8.5 (Johnson et al., 2016) was used for genome indexing and alignment with the following
761 parameters: bowtie-buld --offrate 2; --nohp --mincov 5 --pad 1 --mismatches 1 --mmap u --
762 bowtie_m all --ranmax 'none' --bowtie_cores 32 --sort_mem 248G. After using samtools split (Li
763 et al., 2009) to obtain individual bam files, the featureCounts function from Rsubread (Liao et al.,
764 2019) was used to quantify aligned sRNAs to segmented annotations (see Segmented annotation
765 above) (parameters: allowMultiOverlap= TRUE, strandSpecific= 1, fracOverlap= 0.7,
766 useMetaFeatures= TRUE, largestOverlap= TRUE).

767

768 ***Differential expression and enrichment analyses of exWAGO and adult total across***
769 ***clade V nematodes***

770 We performed individual differential expression analysis comparing exWAGO IP against input
771 libraries (adult total) using edgeR (Robinson, McCarthy, and Smyth, 2010). For *C. elegans* we
772 compared PPW-1, SAGO-1 and SAGO-2 IPs against input samples (data from Seroussi et al.,
773 2023). We removed regions with less than five CPM in conditions with fewer replicates using the
774 filterByExpr function from edgeR v.4.0.16 (Robinson, McCarthy, and Smyth, 2010).

775

776 We then estimated normalisation factors using the TMM normalisation method with the
777 calcNormFactors function. After this, we fitted negative binomial generalised log-linear models
778 using the glmFit function. After visualising diagnostic MA plots and determining that TMM
779 assumptions are broken in these contrasts, new normalisation factors were estimated using the
780 expression of rRNA fragments (those that mapped in sense). To account for compositional bias

781 of rRNA in the different libraries, we used TMM on rRNA genes, reducing the effect of any extreme
782 values for normalisation factor estimation. We then used glmFit and glmTreat to fit negative
783 binomial generalised log-linear models and determine enriched regions. Genomic regions that
784 were not enriched in the IPs were considered as unbound.
785

786 ***Argonaute sRNA guide diversity in Strongylida and C. elegans.***

787 To compare the diversity of sRNA guides between exWAGO orthologs in Strongylida parasites
788 and *C. elegans* argonautes, aiming to account for differences in sequencing depth between IP
789 experiments, we normalized raw counts into counts per million (CPM). We then used the vegan
790 R package to estimate Shannon's and Simpson's diversity indices (Oksanen et al., 2012).
791 Statistical comparison was then carried out using the ggpubr R package (Kassambara, 2025).
792

793 ***Structural LTR annotation, sRNA quantification and differential expression***
794 ***analysis***

795 To classify the structural features of the initial genomic regions annotated as LTR
796 retrotransposons, we first split the curated consensus sequences of LTR retrotransposons into
797 the LTR (directed repeats) and internal region regions (containing CDS of LTR proteins), using
798 selfblast produced with TE-Aid, and an *ad hoc* R script (available at:
799 https://github.com/imu93/ms_evo_exwago). Using the split models, we build a new LTR library.
800 After splitting the models, we selected only one LTR region per consensus model. In models
801 where the nucleotide identity was lower than 80% on their directed repeats or lacked directed
802 repeats, we maintain the model, but without structural information. Using RepeatMasker (-s -cutoff
803 400 -a -nlow) (Smit et al., 2013), we then reannotate all the genomic segments previously
804 identified as LTR retrotransposons. These new annotations were used to quantify the assigned
805 sRNAs using the featureCounts function from the Rsubread package (strandSpecific= 1,
806 fracOverlap= 1, useMetaFeatures= TRUE, largestOverlap= TRUE) (Liao et al., 2019).
807

808 We further performed differential expression analysis using edgeR v.4.0.16 (Robinson, McCarthy,
809 and Smyth, 2010). First, we performed differential expressions analysis between IP against
810 Unbound for both vesicular and non-vesicular exWAGO, and adult IP against adult total, using
811 housekeeping normalisation with rRNA as a reference category. We then fit negative binomial
812 generalised log-linear models with glmFit, and define differentially expressed regions using

813 glmTreat (FDR < 0.05 and log_2 FC = 1 for secreted exWAGO contrasts and FDR < 0.01 and log_2
814 FC = 1 for adult IP against adult total). Finally, using the union of differentially expressed regions,
815 we performed a secreted IP against adult IP contrast. Since this contrast does not involve the
816 comparison of unbalanced conditions, we used TMM for normalisation factor estimation, followed
817 by glmFit and glmTreat (FDR < 0.05 and log_2 FC = $log_2(1.2)$).
818

819 ***Orthology inference of LTR retrotransposons***

820 We performed orthology inference of LTR retrotransposon families among clade V nematodes
821 based on protein identification and phylogenetic reconstruction using Orthofinder v.2.5.4 (Emms
822 and Kelly, 2019). For this, we first annotated LTR proteins on genomic copies using TEsorter
823 (Zhang et al., 2022) with Hidden Markov Models (HMMs) from the metazoan Rex database. We
824 then used an *ad hoc* R script to select TE LTR proteins with at least 80% of model coverage and
825 E-value < 1e-5.
826
827 The orthology of predicted protein domains was interrogated using Orthofinder (Emms and Kelly,
828 2019) to produce phylogenetic hierarchical orthogroups and build phylogenetic trees. For
829 clustering, an inflation value of 6 was used. In addition, a phylogenetic tree based on 443 single-
830 copy ortholog BUSCO genes was used to reconcile gene trees.
831

832 ***Generalised linear mixed model for LTR-derived sRNA production***

833 Using expression values, conservation, divergence and structural information from LTR
834 retrotransposons enriched in exWAGO guide production, we fitted a generalized linear mixed
835 model using the glmmTMB R package (McGillycuddy et al., 2025). We estimate log2 reads per
836 kilobase per million (RPKM) values using the rpkm function from edgeR (Robinson, McCarthy,
837 and Smyth, 2010), we then mean-centred the values by subtracting the mean log2 RPKM
838 estimated for antisense-derived sRNAs. As fixed categorical effects, we used conservation
839 encoded as highly conserved (HC), conserved (C), lowly conserved (LC) and undetermined. In
840 addition, we modelled LTR type as autonomous (considering the presence of GAG, protease,
841 endonuclease, reverse transcriptase, endonuclease, and integrase), non-autonomous (lacking at
842 least one protein) and soloLTR (based on the LTR structure, lack of coding potential and
843 surrounding LTR segments). We also used the RepeatMasker estimated relative divergence
844 values per copy as a continuous fixed effect (Smit et al., 2013). To account for the variance each

845 family has, we also modelled family as a categorical random effect. The same model was
846 independently used for each species, and per-species model fit was assessed using diagnostic
847 qq, residual vs fitted, residual density and fixed effect distribution plots, with no major deviations
848 from normality (Fig. S9) .
849

850 ***Binomial generalised linear model for secreted exWAGO guides.***

851 To assess which factors influence exWAGO guide secretion from LTR retrotransposons, we fitted
852 a binomial generalised linear model using the lme4 R package (Bates et al., 2015). We fitted
853 enrichment in exWAGO EV IP and non-enrichment as a dichotomic response variable. As
854 explanatory variables, we fitted conservation (based on previously defined phylogenetic
855 hierarchical orthogroups HOGs), strand (sense or antisense), and structure (soloLTR, LTR region,
856 internal-coding region) as categorical. In addition, to account for differences in length, we fitted
857 $\log_{10}$-transformed mean-centred length as continuous. Model assumptions were inspected using
858 the DHARMa R package (Hartig, 2024), with no deviations observed in the simulated residuals
859 (Fig. S11).
860

861 # Acknowledgments

862 We acknowledge Jennifer Mcintyre, for the prepublication access to the *Teladorsagia*
863 *circumcincta* genome assembly and genome annotation. We acknowledge Clément Goubert for
864 his valuable insight into the transposon manual curation. Computation was performed on the
865 Ashworth Compute Co-operative Cluster (AC3).
866

867 # Funding

868 I.M.U was funded by Consejo Nacional de Ciencia y Tecnologia - Mexico (CONACyT) (896776)
869 and Postgraduate Scholarship—Darwin Trust of Edinburgh. K.N was funded by the Wellcome
870 Trust PhD Programme grant 108905/Z/15/Z and the ERC grant 101002385 awarded to A.H.B.
871 A.O.D. was funded by the Postgraduate Scholarship—Darwin Trust of Edinburgh. JLH was
872 funded from Leverhulme Grant RPG-2019-404.
873

874 # Data Availability

875 New small RNA-seq libraries for exWAGO IPs and total adults, from *H. polygyrus* are available at
876 the NCBI SRA under BioProject PRJNA1327922. Previously published sRNA-Seq data that were
877 reanalysed in this manuscript are available under BioProjects PRJNA1200757 and
878 PRJNA481340. Scripts related to differential expression analysis, genome annotation, and repeat
879 libraries are available at https://github.com/imu93/ms_evo_exwago**.** Curated repeat libraries were
880 also submitted to Dfam.

881 # Literature Cited

882 Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.*
883 [Computer software]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

884 Ansaloni, F., Scarpato, M., Di Schiavi, E., Gustincich, S., & Sanges, R. (2019). Exploratory
885 analysis of transposable elements expression in the *C. elegans* early embryo. *BMC*
886 *Bioinformatics*, *20*(S9), 484. https://doi.org/10.1186/s12859-019-3088-7

887 Arkhipova, I. R. (2018). Neutral Theory, Transposable Elements, and Eukaryotic Genome
888 Evolution. *Molecular Biology and Evolution*, *35*(6), 1332–1337.
889 https://doi.org/10.1093/molbev/msy083

890 Baril, T., Galbraith, J., & Hayward, A. (2024). Earl Grey: A Fully Automated User-Friendly
891 Transposable Element Annotation and Analysis Pipeline. *Molecular Biology and*
892 *Evolution*, *41*(4), msae068. https://doi.org/10.1093/molbev/msae068

893 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects
894 Models Using lme4. *Journal of Statistical Software*, *67*(1).
895 https://doi.org/10.18637/jss.v067.i01

896 Beltran, T., Barroso, C., Birkle, T. Y., Stevens, L., Schwartz, H. T., Sternberg, P. W.,
897 Fradin, H., Gunsalus, K., Piano, F., Sharma, G., Cerrato, C., Ahringer, J., Martínez-Pérez,
898 E., Blaxter, M., & Sarkies, P. (2019). Comparative Epigenomics Reveals that RNA
899 Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA
900 Biogenesis. *Developmental Cell*, *48*(6), 793-810.e6.
901 https://doi.org/10.1016/j.devcel.2018.12.026

902 Buck, A. H., Coakley, G., Simbari, F., McSorley, H. J., Quintana, J. F., Le Bihan, T., Kumar,
903 S., Abreu-Goodger, C., Lear, M., Harcus, Y., Ceroni, A., Babayan, S. A., Blaxter, M., Ivens,
904 A., & Maizels, R. M. (2014). Exosomes secreted by nematode parasites transfer small

905 RNAs to mammalian cells and modulate innate immunity. *Nature Communications*, *5*(1),
906 5488. https://doi.org/10.1038/ncomms6488
907 Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). Evaluation of statistical
908 methods for normalization and differential expression in mRNA-Seq experiments. *BMC*
909 *Bioinformatics*, *11*(1), 94. https://doi.org/10.1186/1471-2105-11-94
910 Calcino, A. D., Fernandez-Valverde, S. L., Taft, R. J., & Degnan, B. M. (2018). Diverse
911 RNA interference strategies in early-branching metazoans. *BMC Evolutionary Biology*,
912 *18*(1), 160. https://doi.org/10.1186/s12862-018-1274-2
913 Chow, F. W.-N., Koutsovoulos, G., Ovando-Vázquez, C., Neophytou, K., Bermúdez-
914 Barrientos, J. R., Laetsch, D. R., Robertson, E., Kumar, S., Claycomb, J. M., Blaxter, M.,
915 Abreu-Goodger, C., & Buck, A. H. (2019). Secretion of an Argonaute protein by a parasitic
916 nematode and the evolution of its siRNA guides. *Nucleic Acids Research*, *47*(7), 3594–
917 3606. https://doi.org/10.1093/nar/gkz142
918 Davis, M. P. A., Van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A. J.
919 (2013). Kraken: A set of tools for quality control and analysis of high-throughput sequence
920 data. *Methods*, *63*(1), 41–49. https://doi.org/10.1016/j.ymeth.2013.06.027
921 Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis
922 results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–
923 3048. https://doi.org/10.1093/bioinformatics/btw354
924 Fischer, S. E. J. (2024). Activity and Silencing of Transposable Elements in C. elegans.
925 *DNA*, *4*(2), 129–140. https://doi.org/10.3390/dna4020007
926 Fischer, S. E. J., & Ruvkun, G. (2020). *Caenorhabditis elegans* ADAR editing and the ERI-
927 6/7/MOV10 RNAi pathway silence endogenous viral elements and LTR retrotransposons.
928 *Proceedings of the National Academy of Sciences*, *117*(11), 5987–5996.
929 https://doi.org/10.1073/pnas.1919028117

930 Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F.
931 (2020). RepeatModeler2 for automated genomic discovery of transposable element
932 families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457.
933 https://doi.org/10.1073/pnas.1921046117
934 Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2
935 accurately identifies known and hundreds of novel microRNA genes in seven animal
936 clades. *Nucleic Acids Research*, *40*(1), 37–52. https://doi.org/10.1093/nar/gkr688
937 Goubert, C., Craig, R. J., Bilat, A. F., Peona, V., Vogan, A. A., & Protasio, A. V. (2022). A
938 beginner's guide to manual curation of transposable elements. *Mobile DNA*, *13*(1), 7.
939 https://doi.org/10.1186/s13100-021-00259-7
940 Guan, L., & Grigoriev, A. (2021). Computational meta-analysis of ribosomal RNA
941 fragments: Potential targets and interaction mechanisms. *Nucleic Acids Research*, *49*(7),
942 4085–4103. https://doi.org/10.1093/nar/gkab190
943 Hartig, F. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed)*
944 *Regression Models*. https://CRAN.R-project.org/package=DHARMa
945 Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., & Berriman, M. (2017). WormBase ParaSite
946 − a comprehensive resource for helminth genomics. *Molecular and Biochemical*
947 *Parasitology*, *215*, 2–10. https://doi.org/10.1016/j.molbiopara.2016.11.005
948 Hu, K., Ni, P., Xu, M., Zou, Y., Chang, J., Gao, X., Li, Y., Ruan, J., Hu, B., & Wang, J.
949 (2024). HiTE: A fast and accurate dynamic boundary adjustment approach for full-length
950 transposable element detection and annotation. *Nature Communications*, *15*(1), 5573.
951 https://doi.org/10.1038/s41467-024-49912-8
952 International Helminth Genomes Consortium. (2019). Comparative genomics of the major
953 parasitic worms. *Nature Genetics*, *51*(1), 163–174. https://doi.org/10.1038/s41588-018-
954 0262-1

Johnson, N. R., Yeoh, J. M., Coruh, C., & Axtell, M. J. (2016). Improved Placement of Multi-mapping Small RNAs. *G3 Genes|Genomes|Genetics*, *6*(7), 2103-2111Ivey and Srivastava. https://doi.org/doi:%252010.1534/g3.116.030452

Kanzaki, N., Tsai, I. J., Tanaka, R., Hunt, V. L., Liu, D., Tsuyama, K., Maeda, Y., Namai, S., Kumagai, R., Tracey, A., Holroyd, N., Doyle, S. R., Woodruff, G. C., Murase, K., Kitazume, H., Chai, C., Akagi, A., Panda, O., Ke, H.-M., … Kikuchi, T. (2018). Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. *Nature Communications*, *9*(1), 3216. https://doi.org/10.1038/s41467-018-05712-5

Kassambara, A. (2025). *ggpubr: "ggplot2" Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

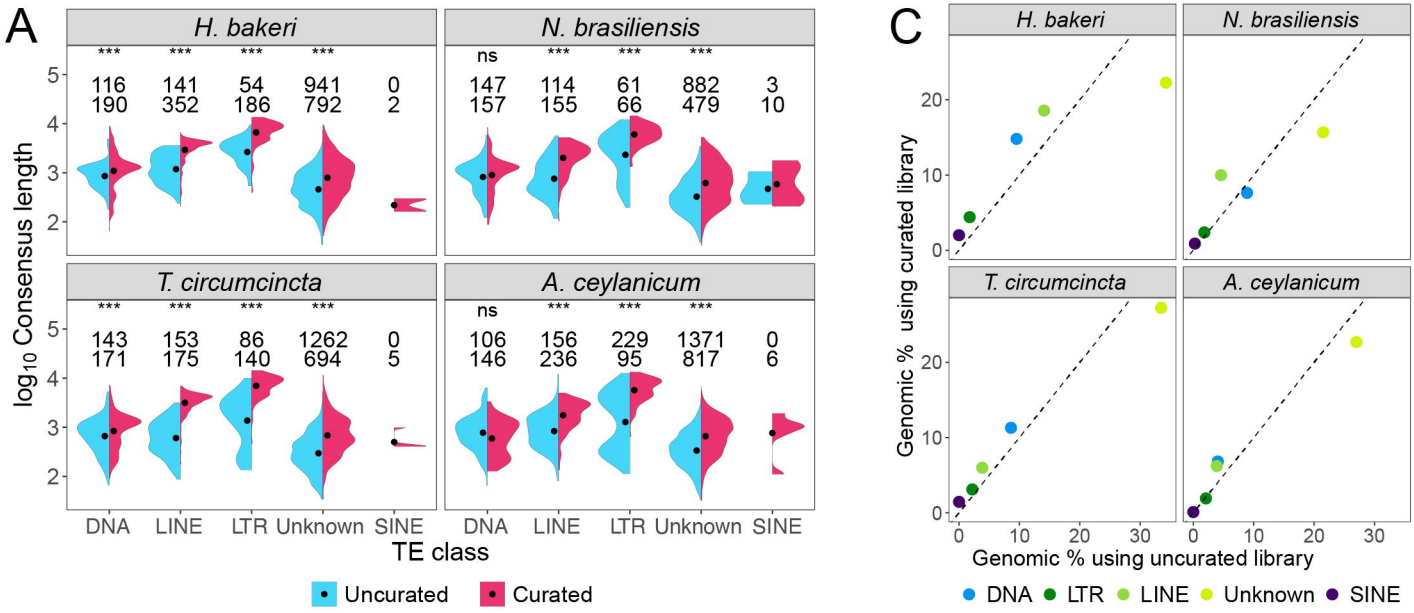
Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, *47*(8), e47–e47. https://doi.org/10.1093/nar/gkz114

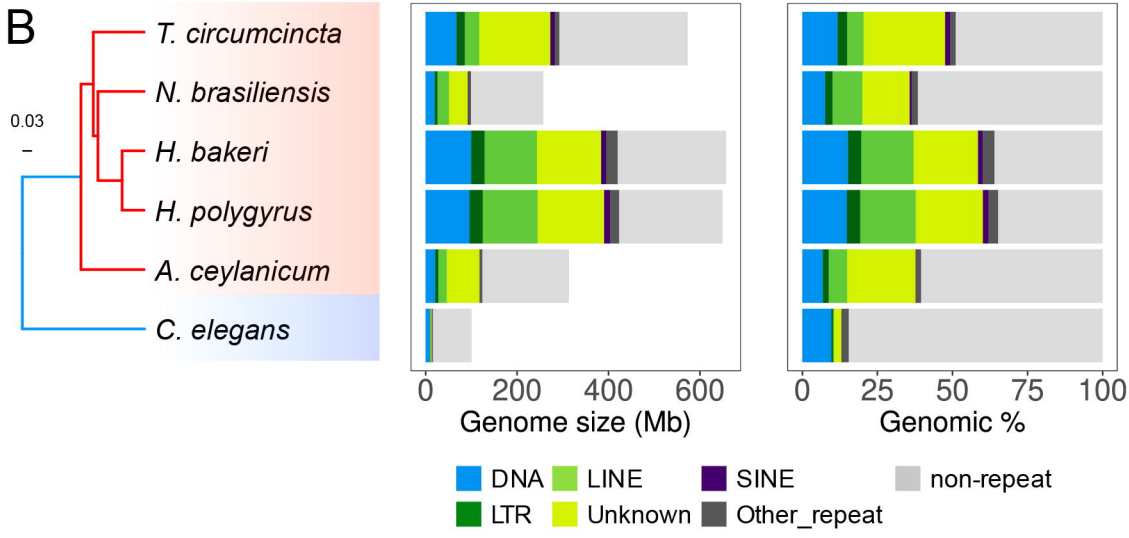
979 McGillycuddy, M., Warton, D. I., Popovic, G., & Bolker, B. M. (2025). Parsimoniously
980 Fitting Large Multivariate Random Effects in glmmTMB. *Journal of Statistical Software*,
981 *112*(1). https://doi.org/10.18637/jss.v112.i01
982 Neophytou, K., Martínez-Ugalde, I., Fenton, T., Robertson, E., Strachan, L. J., Harcus, Y.,
983 Naar, C. M., Wright, D., White, R., Evans, M. J., Bermúdez-Barrientos, J. R., Maizels, R.
984 M., Aroian, R. V., Nisbet, A. J., & Abreu-Goodger, C. Buck, A. H. (2025). A non-vesicular
985 Argonaute protein is transmitted from nematode to mouse and is important for parasite
986 survival. *Biorxiv (Pre Print)*. https://doi.org/DOI:10.1101/2025.04.01.646544
987 Ni, J. Z., Kalinava, N., Chen, E., Huang, A., Trinh, T., & Gu, S. G. (2016). A
988 transgenerational role of the germline nuclear RNAi pathway in repressing heat stress-
989 induced transcriptional activation in C. elegans. *Epigenetics & Chromatin*, *9*(1), 3.
990 https://doi.org/10.1186/s13072-016-0052-x
991 Nikonorova, I. A., Wang, J., Cope, A. L., Tilton, P. E., Power, K. M., Walsh, J. D., Akella,
992 J. S., Krauchunas, A. R., Shah, P., & Barr, M. M. (2022). Isolation, profiling, and tracking
993 of extracellular vesicle cargo in Caenorhabditis elegans. *Current Biology*, *32*(9), 1924-
994 1936.e6. https://doi.org/10.1016/j.cub.2022.03.005
995 Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B.,
996 Simpson, G. L., Sólymos, P., Stevens, M. H. H., & Wagner, H. (2012). *vegan: Community*
997 *Ecology Package*.
998 Pak, J., & Fire, A. (2007). Distinct Populations of Primary and Secondary Effectors During
999 RNAi in *C. elegans*. *Science*, *315*(5809), 241–244.
1000 https://doi.org/10.1126/science.1132839
1001 Porquier, A., Tisserant, C., Salinas, F., Glassl, C., Wange, L., Enard, W., Hauser, A.,
1002 Hahn, M., & Weiberg, A. (2021). Retrotransposons as pathogenicity factors of the plant
1003 pathogenic fungus Botrytis cinerea. *Genome Biology*, *22*(1), 225.
1004 https://doi.org/10.1186/s13059-021-02446-4

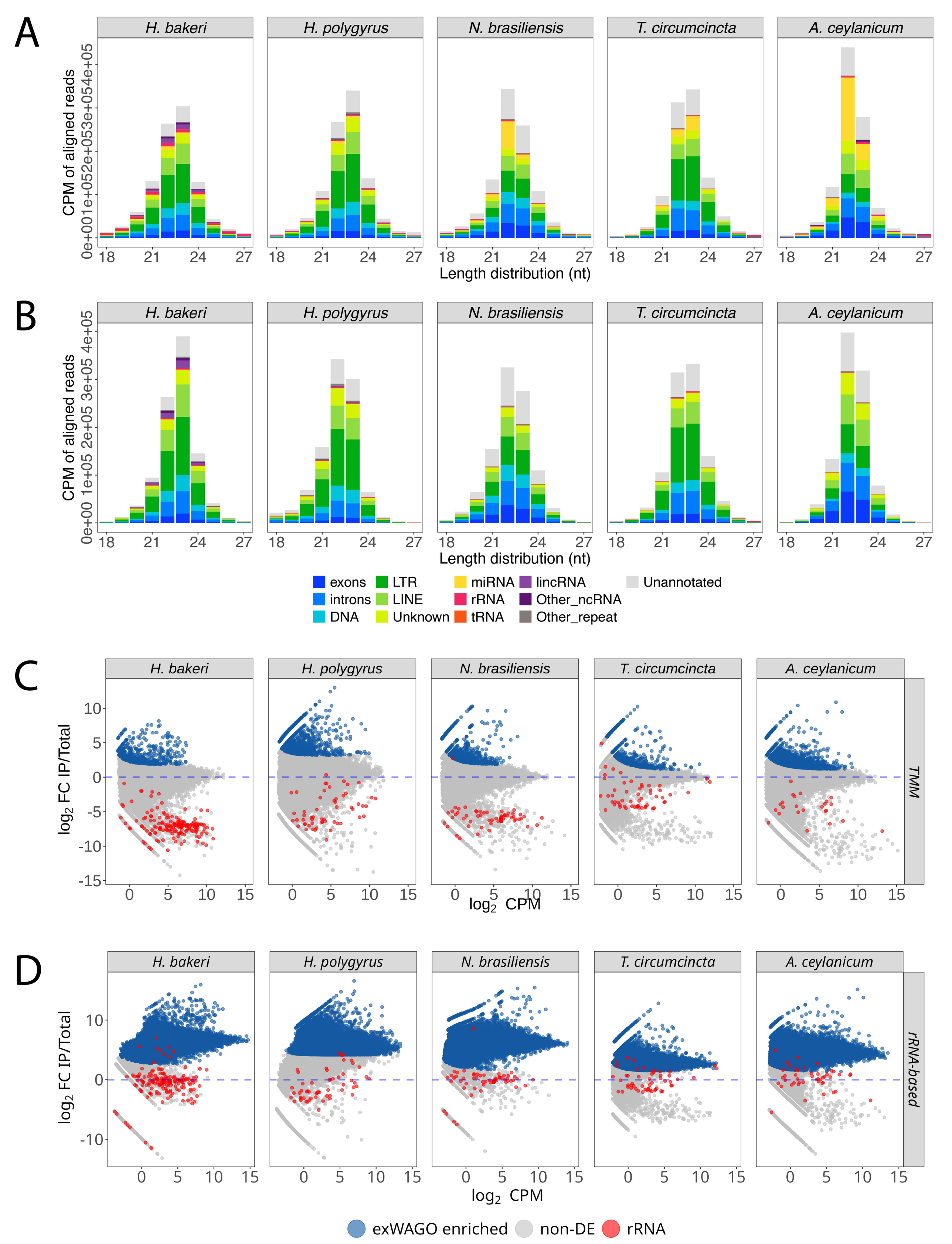
1005 Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential
1006 expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25.
1007 https://doi.org/10.1186/gb-2010-11-3-r25
1008 Rooney, J., Williams, T. L., Northcote, H. M., Frankl, F. E. K., Price, D. R. G., Nisbet, A.
1009 J., Morphew, R. M., & Cantacessi, C. (2022). Excretory-secretory products from the brown
1010 stomach worm, Teladorsagia circumcincta, exert antimicrobial activity in in vitro growth
1011 assays. *Parasites & Vectors*, *15*(1), 354. https://doi.org/10.1186/s13071-022-05443-z
1012 Sarkies, P., Selkirk, M. E., Jones, J. T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B.,
1013 Ardila-Garcia, A., Fast, N. M., Schiffer, P. M., Kraus, C., Taylor, M. J., Koutsovoulos, G.,
1014 Blaxter, M. L., & Miska, E. A. (2015). Ancient and Novel Small RNA Pathways
1015 Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. *PLOS*
1016 *Biology*, *13*(2), e1002061. https://doi.org/10.1371/journal.pbio.1002061
1017 Seroussi, U., Lugowski, A., Wadi, L., Lao, R. X., Willis, A. R., Zhao, W., Sundby, A. E.,
1018 Charlesworth, A. G., Reinke, A. W., & Claycomb, J. M. (2023). A comprehensive survey
1019 of C. elegans argonaute proteins reveals organism-wide gene regulatory networks and
1020 functions. *eLife*, *12*, e83853. https://doi.org/10.7554/eLife.83853
1021 Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker*.
1022 Stevens, L., Martínez-Ugalde, I., King, E., Wagah, M., Absolon, D., Bancroft, R., Gonzalez
1023 De La Rosa, P., Hall, J. L., Kieninger, M., Kloch, A., Pelan, S., Robertson, E., Pedersen,
1024 A. B., Abreu-Goodger, C., Buck, A. H., & Blaxter, M. (2023). Ancient diversity in host-
1025 parasite interaction genes in a model parasitic nematode. *Nature Communications*, *14*(1),
1026 7776. https://doi.org/10.1038/s41467-023-43556-w
1027 Sweeny, A. R., Clerc, M., Pontifes, P. A., Venkatesan, S., Babayan, S. A., & Pedersen, A.
1028 B. (2021). Supplemented nutrition decreases helminth burden and increases drug efficacy
1029 in a natural host–helminth system. *Proceedings of the Royal Society B: Biological*
1030 *Sciences*, *288*(1943), 20202722. https://doi.org/10.1098/rspb.2020.2722

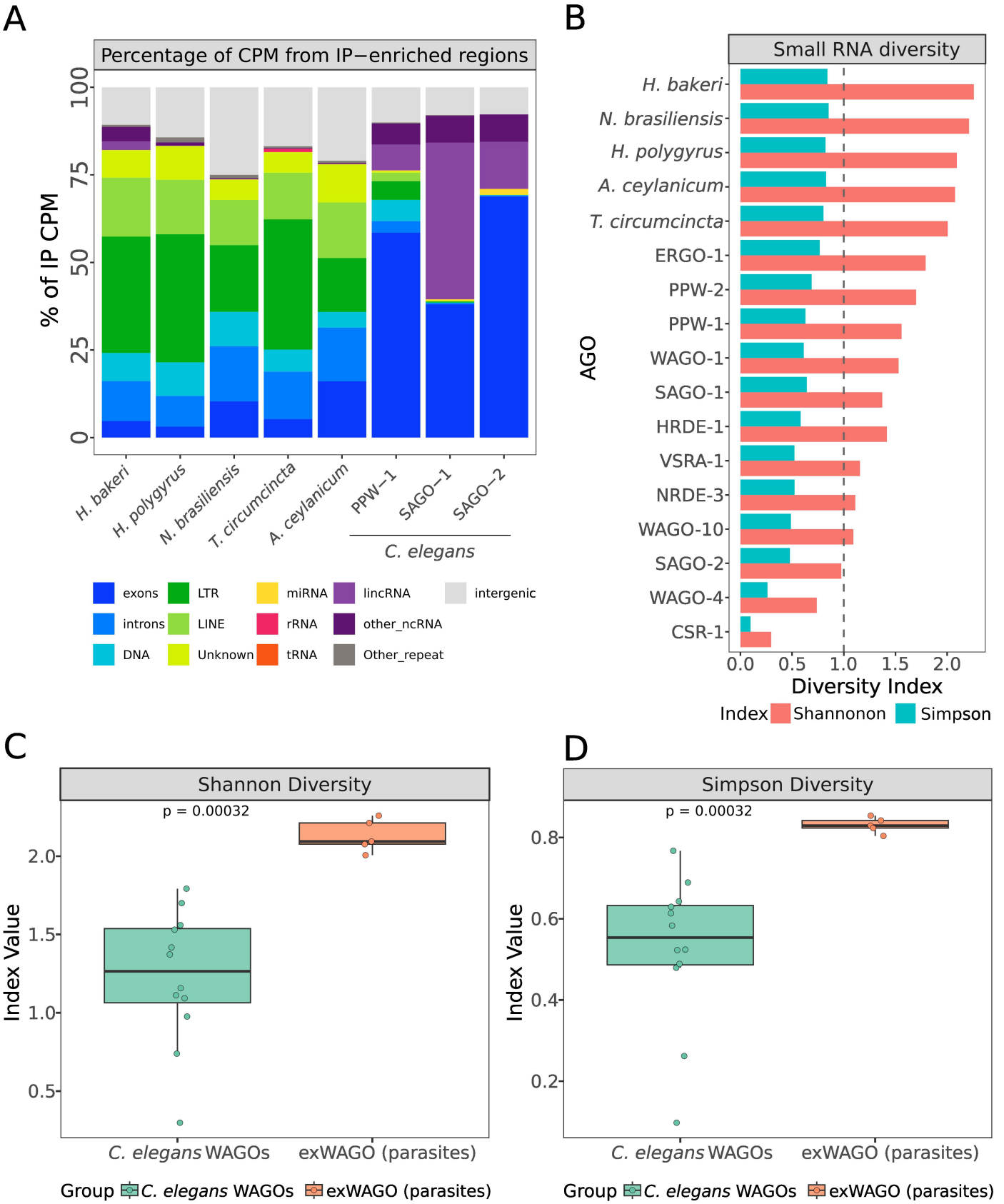
1031 Tumescheit, C., Firth, A. E., & Brown, K. (2022). CIAlign: A highly customisable command
1032 line tool to clean, interpret and visualise multiple sequence alignments. *PeerJ*, *10*, e12983.
1033 https://doi.org/10.7717/peerj.12983
1034 Uzoechi, S. C., Rosa, B. A., Singh, K. S., Choi, Y.-J., Bracken, B. K., Brindley, P. J.,
1035 Townsend, R. R., Sprung, R., Zhan, B., Bottazzi, M.-E., Hawdon, J. M., Wong, Y., Loukas,
1036 A., Djuranovic, S., & Mitreva, M. (2023). Excretory/Secretory Proteome of Females and
1037 Males of the Hookworm Ancylostoma ceylanicum. *Pathogens*, *12*(1), 95.
1038 https://doi.org/10.3390/pathogens12010095
1039 Van Lopik, J., Alizada, A., Trapotsi, M.-A., Hannon, G. J., Bornelöv, S., & Czech Nicholson,
1040 B. (2023). Unistrand piRNA clusters are an evolutionarily conserved mechanism to
1041 suppress endogenous retroviruses across the Drosophila genus. *Nature*
1042 *Communications*, *14*(1), 7337. https://doi.org/10.1038/s41467-023-42787-1
1043 Wolf, G., De Iaco, A., Sun, M.-A., Bruno, M., Tinkham, M., Hoang, D., Mitra, A., Ralls, S.,
1044 Trono, D., & Macfarlan, T. S. (2020). KRAB-zinc finger protein gene expansion in
1045 response to active retrotransposons in the murine lineage. *eLife*, *9*, e56337.
1046 https://doi.org/10.7554/eLife.56337
1047 Wong, W. Y., & Simakov, O. (2019). RepeatCraft: A meta-pipeline for repetitive element
1048 de-fragmentation and annotation. *Bioinformatics*, *35*(6), 1051–1052.
1049 https://doi.org/10.1093/bioinformatics/bty745
1050 World Health Organization. (2020). Global distribution and prevalence of soil transmitted
1051 helminth infections. *World Health Organization Key Fact Sheet*.
1052 Wynant, N., Santos, D., & Vanden Broeck, J. (2017). The evolution of animal Argonautes:
1053 Evidence for the absence of antiviral AGO Argonautes in vertebrates. *Scientific Reports*,
1054 *7*(1), 9230. https://doi.org/10.1038/s41598-017-08043-5

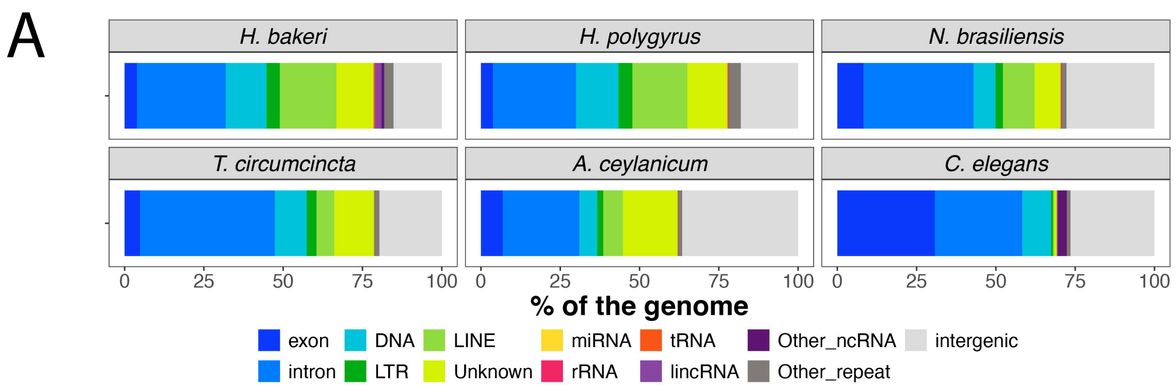
1055 Zagoskin, M. V., Wang, J., Neff, A. T., Veronezi, G. M. B., & Davis, R. E. (2022). Small
1056 RNA pathways in the nematode Ascaris in the absence of piRNAs. *Nature*
1057 *Communications*, *13*(1), 837. https://doi.org/10.1038/s41467-022-28482-7
1058 Zhang, R.-G., Li, G.-Y., Wang, X.-L., Dainat, J., Wang, Z.-X., Ou, S., & Ma, Y. (2022).
1059 TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes.
1060 *Horticulture Research*, *9*, uhac017. https://doi.org/10.1093/hr/uhac017
1061 Zhou, W., Liang, G., Molloy, P. L., & Jones, P. A. (2020). DNA methylation enables
1062 transposable element-driven genome expansion. *Proceedings of the National Academy*
1063 *of Sciences*, *117*(32), 19359–19366. https://doi.org/10.1073/pnas.1921719117
1064 Zhou, Y., Wang, G., Zhang, J., & Li, H. (2017). A Hypothesis Testing Based Method for
1065 Normalization and Differential Expression Analysis of RNA-Seq Data. *PLOS ONE*, *12*(1),
1066 e0169594. https://doi.org/10.1371/journal.pone.0169594

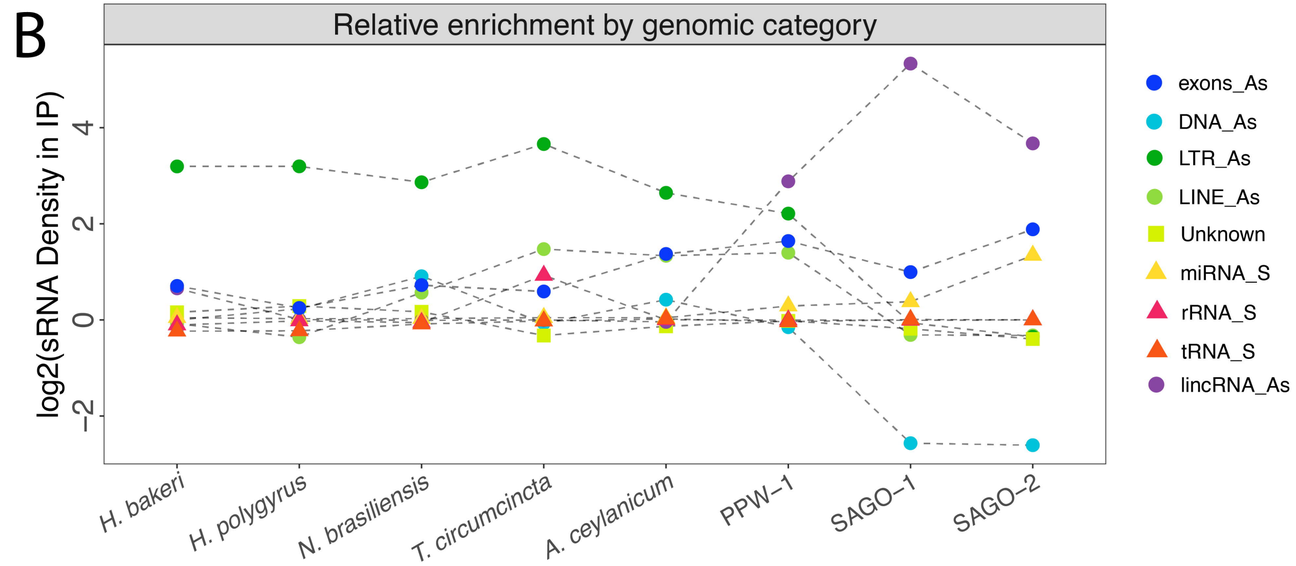

A
H. bakeri
N. brasiliensis
T. circumcincta
A. ceylanicum
log10 Consensus length
TE class
DNA
LINE
LTR
Unknown
SINE
Uncurated
Curated
C
Genomic % using curated library
Genomic % using uncurated library
DNA
LTR
LINE
Unknown
SINE

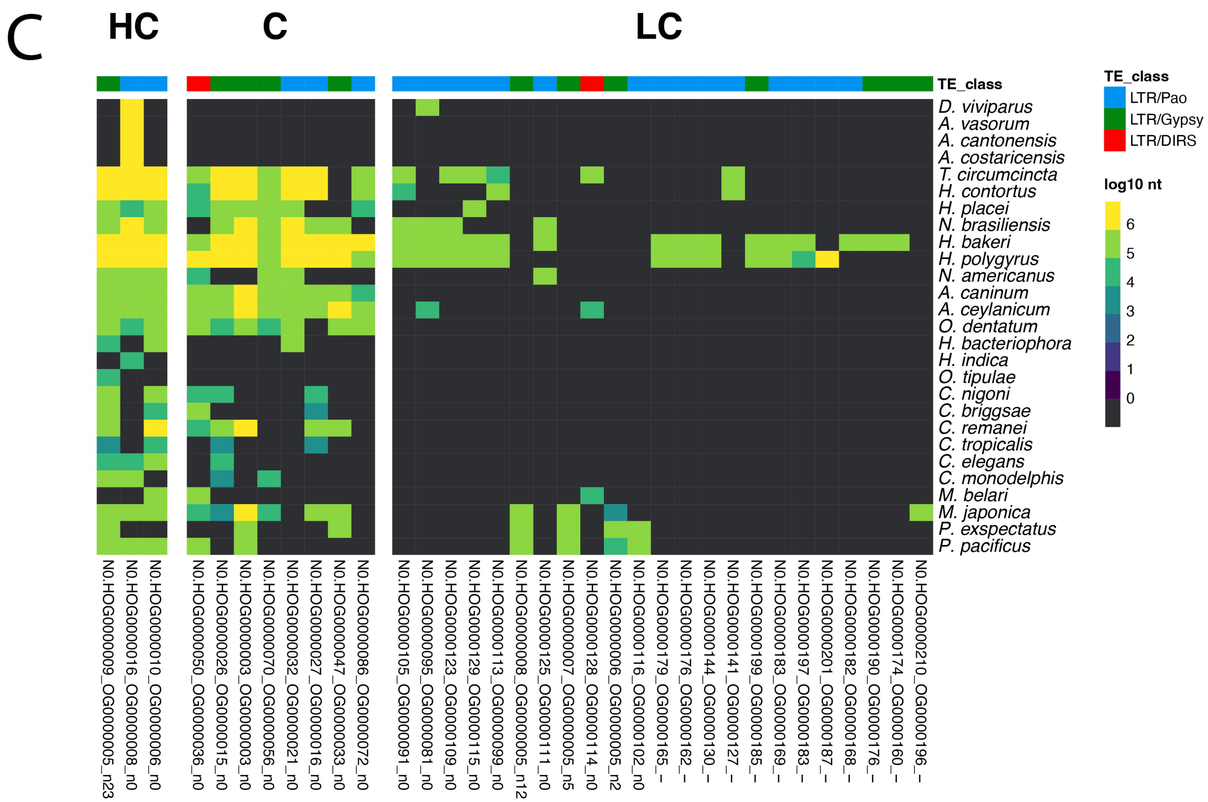

B
T. circumcincta
N. brasiliensis
H. bakeri
H. polygyrus
A. ceylanicum
C. elegans
0.03
Genome size (Mb)
Genomic %
DNA
LTR
LINE
Unknown
SINE
Other_repeat
non-repeat

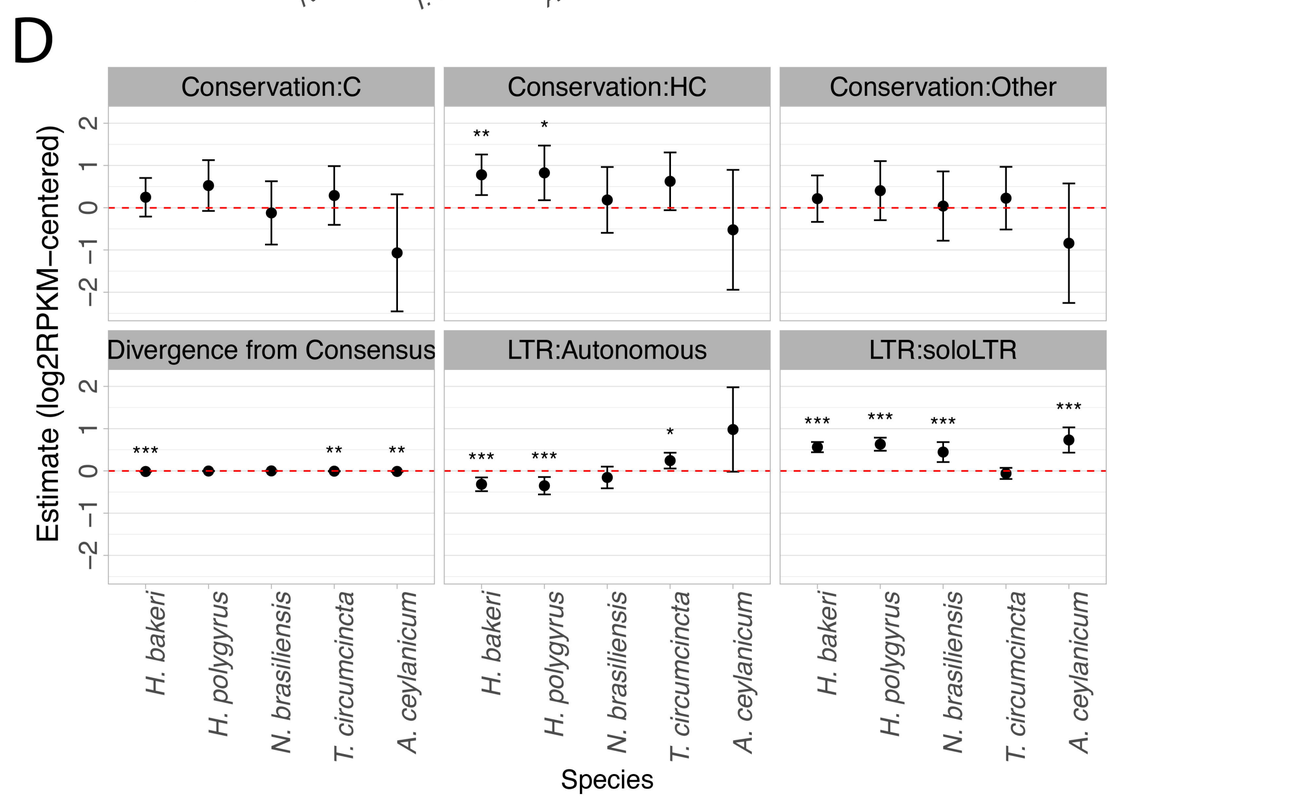

A
B
C
D
H. bakeri
H. polygyrus
N. brasiliensis
T. circumcincta
A. ceylanicum
CPM of aligned reads
Length distribution (nt)
exons
introns
DNA
LTR
LINE
Unknown
miRNA
rRNA
tRNA
lincRNA
Other_ncRNA
Other_repeat
Unannotated
log2 FC IP/Total
log2 CPM
TMM
rRNA-based
exWAGO enriched
non-DE
rRNA

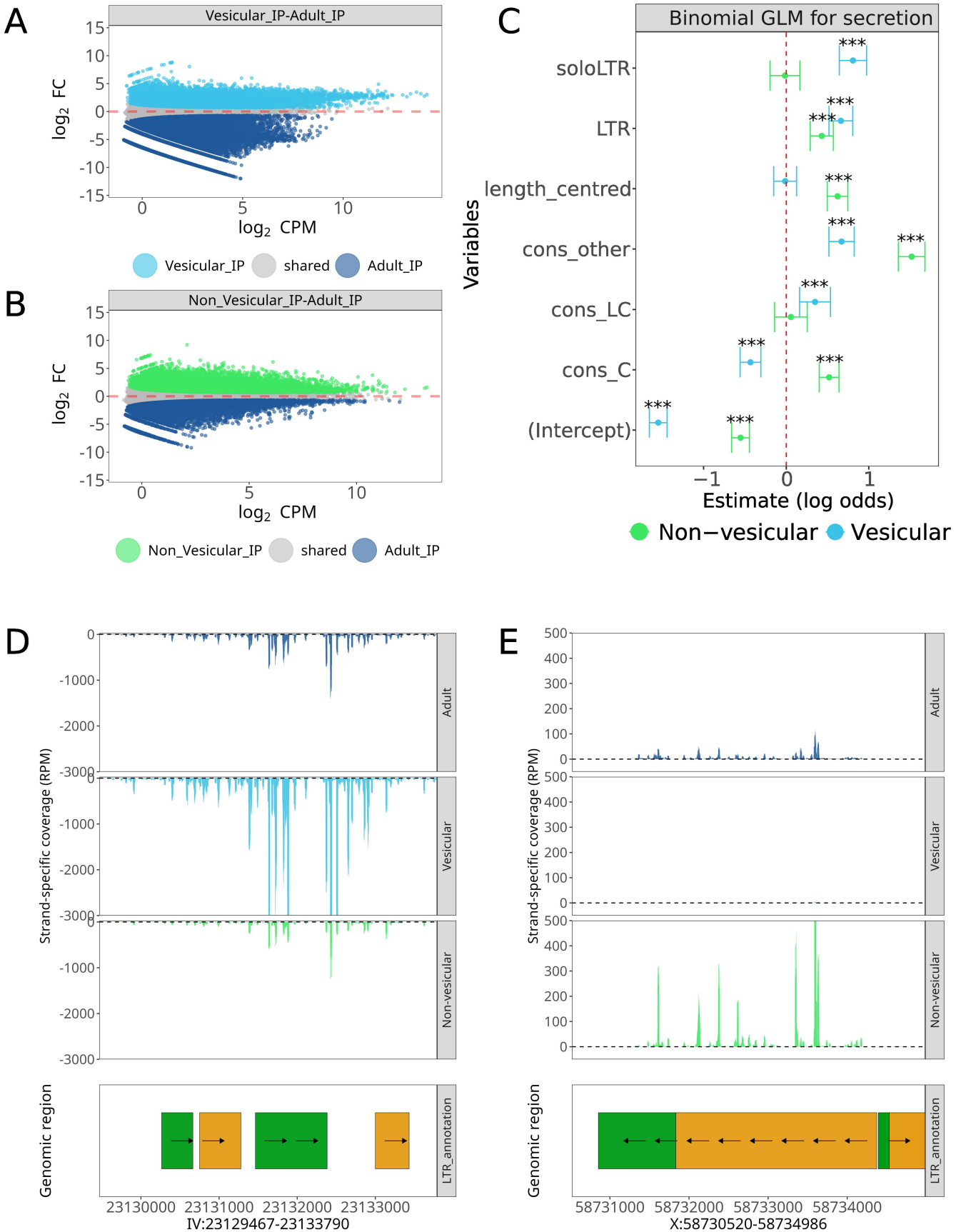

A
Percentage of CPM from IP−enriched regions
% of IP CPM
0
25
50
75
100
H. bakeri
H. polygyrus
N. brasiliensis
T. circumcincta
A. ceylanicum
PPW−1
SAGO−1
SAGO−2
C. elegans
exons
introns
DNA
LTR
LINE
Unknown
miRNA
rRNA
tRNA
lincRNA
other_ncRNA
Other_repeat
intergenic
B
Small RNA diversity
H. bakeri
N. brasiliensis
H. polygyrus
A. ceylanicum
T. circumcincta
ERGO-1
PPW-2
PPW-1
WAGO-1
SAGO-1
HRDE-1
VSRA-1
NRDE-3
WAGO-10
SAGO-2
WAGO-4
CSR-1
AGO
0.0
0.5
1.0
1.5
2.0
Diversity Index
Index
Shannonon
Simpson
C
Shannon Diversity
p = 0.00032
Index Value
0.5
1.0
1.5
2.0
C. elegans WAGOs
exWAGO (parasites)
Group
C. elegans WAGOs
exWAGO (parasites)
D
Simpson Diversity
p = 0.00032
Index Value
0.2
0.4
0.6
0.8
C. elegans WAGOs
exWAGO (parasites)
Group
C. elegans WAGOs
exWAGO (parasites)


A
H. bakeri
H. polygyrus
N. brasiliensis
T. circumcincta
A. ceylanicum
C. elegans
0
25
50
75
100
% of the genome
exon
intron
DNA
LTR
LINE
Unknown
miRNA
rRNA
tRNA
lincRNA
Other_ncRNA
Other_repeat
intergenic


B
Relative enrichment by genomic category
log2(sRNA Density in IP)
−2
0
2
4
H. bakeri
H. polygyrus
N. brasiliensis
T. circumcincta
A. ceylanicum
PPW−1
SAGO−1
SAGO−2
exons_As
DNA_As
LTR_As
LINE_As
Unknown
miRNA_S
rRNA_S
tRNA_S
lincRNA_As


C
HC
C
LC
TE_class
D. viviparus
A. vasorum
A. cantonensis
A. costaricensis
T. circumcincta
H. contortus
H. placei
N. brasiliensis
H. bakeri
H. polygyrus
N. americanus
A. caninum
A. ceylanicum
O. dentatum
H. bacteriophora
H. indica
O. tipulae
C. nigoni
C. briggsae
C. remanei
C. tropicalis
C. elegans
C. monodelphis
M. belari
M. japonica
P. exspectatus
P. pacificus
TE_class
LTR/Pao
LTR/Gypsy
LTR/DIRS
log10 nt
6
5
4
3
2
1
0


D
Conservation:C
Conservation:HC
Conservation:Other
Divergence from Consensus
LTR:Autonomous
LTR:soloLTR
Estimate (log2RPKM−centered)
−2
−1
0
1
2
H. bakeri
H. polygyrus
N. brasiliensis
T. circumcincta
A. ceylanicum
Species


A
Vesicular_IP-Adult_IP
log2 FC
log2 CPM
Vesicular_IP
shared
Adult_IP
B
Non_Vesicular_IP-Adult_IP
log2 FC
log2 CPM
Non_Vesicular_IP
shared
Adult_IP
C
Binomial GLM for secretion
Variables
soloLTR
LTR
length_centred
cons_other
cons_LC
cons_C
(Intercept)
Estimate (log odds)
Non−vesicular
Vesicular
D
Strand-specific coverage (RPM)
Adult
Vesicular
Non-vesicular
Genomic region
LTR_annotation
IV:23129467-23133790
E
Strand-specific coverage (RPM)
Adult
Vesicular
Non-vesicular
Genomic region
LTR_annotation
X:58730520-58734986