An extracellular Argonaute protein mediates export of repeat-associated small RNAs into vesicles in parasitic nematodes

Chow FWN[1*], Koutsovoulos G[2,§*], Ovando-Vázquez C[3,§§*], Laetsch DR[2], Bermúdez-Barrientos JR[3], Claycomb JM[4], Blaxter M[2,5**], Abreu-Goodger C[3**], Buck AH[1,5**]

(1) Institute of Immunology and Infection Research, School of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3JT, UK

(2) Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3JT, UK

3) Unidad de Genómica Avanzada (Langebio), Centro de Investigación y de Estudios Avanzados del IPN, Irapuato, Guanajuato 36824, México

(4) Department of Molecular Genetics, University of Toronto, Toronto, ON M5G 1M1, Canada

(5) Centre for Immunity, Infection and Evolution, School of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3JT, UK

§ Current address: INRA, UMR 1355 Institute Sophia Agrobiotech, 06903 Sophia Antipolis, France

§§ Current address: CONACYT-CNS-IPICYT, San Luis Potosí, SLP 78216, México

*Author's contributed equally

**To whom correspondence should be addressed: a.buck@ed.ac.uk, cei.abreu@cinvestav.mx, Mark.Blaxter@ed.ac.uk

Abstract

Mobile small RNAs are an integral component of the arms race between plants and fungal parasites, and several studies suggest microRNAs could similarly operate between parasitic nematodes and their animal hosts. However, whether and how specific sequences are selected for export by parasites is unknown. Here we use density gradient purification and proteinase K sensitivity analysis to demonstrate that a specific Argonaute protein (exWAGO) is secreted in extracellular vesicles (EVs) released by the gastrointestinal nematode *Heligmosomodies bakeri*, at multiple copies per EV. Phylogenetic and gene expression analyses demonstrate exWAGO is highly conserved and abundantly expressed in related parasites, including the human hookworm and proteomic analyses confirm this is the only Argonaute secreted by rodent parasites. In contrast, exWAGO orthologues in species from the free-living genus *Caenorhabditis* are highly diverged. By re-sequencing and re-annotating the *H. bakeri* genome, and sequencing multiple small RNA libraries, we determined that the most abundant small RNAs released from the nematode parasite are not microRNAs but rather secondary small interfering RNAs (siRNAs) that are produced by RNA-dependent RNA Polymerases. We further identify distinct evolutionary properties of the siRNAs resident in free-living or parasitic nematodes versus those exported in EVs by the parasite and show that the latter are specifically associated with exWAGO. Together this work identifies an Argonaute protein as a mediator of RNA export and suggests rhabditomorph nematode parasites may have co-opted a novel nematode-unique pathway to communicate with their hosts.

48 **Introduction**

49 Small RNA-mediated gene regulatory mechanisms are used by cellular organisms
50 and viruses to enable various aspects of their development, defence strategies and
51 physiology (1). In eukaryotes small RNAs (sRNAs) operate within RNA-Induced
52 Silencing Complexes (RISCs). The engines of these complexes are a diverse
53 family of Argonaute proteins that are guided by the sRNA to target nucleic acids in
54 a sequence-specific manner. The downstream effects of sRNA guide-directed
55 recognition are diverse and depend on the biogenesis of the sRNA guide, the class
56 of the target and Argonaute with which they both associate. One deeply conserved
57 sRNA-dependent mechanism is the post-transcriptional regulation of gene
58 expression by microRNAs (miRNAs), which associate with an Argonaute of the
59 AGO clade. MiRNAs were discovered for their crucial roles in development and are
60 now appreciated to regulate numerous aspects of physiology and signalling. In the
61 last 10 years, studies across a broad range of animal systems have implicated
62 miRNAs in intercellular communication through their transport in extracellular
63 vesicles (EVs) (2). Several reports also suggest mammalian miRNAs can move
64 into other organisms, influencing gene expression and growth of microbes in the
65 gut (3) and malaria parasites in the blood (4). We previously reported that the
66 nematode parasite *Heligmosomoides bakeri* (renamed from *Heligmosomoides*
67 *polygyrus* (5)) releases its own miRNAs within extracellular vesicles (EVs) that are
68 internalized by mouse cells and suppress innate immune responses (6). *H. bakeri*
69 is a natural parasite of mice that serves as an important animal model for the study
70 of immunomodulation by strongylid parasites, which establish chronic infections in
71 their hosts by inducing immune suppression and tolerance. These parasites infect
72 half a billion people and are highly prevalent in livestock. The EVs they release
73 have been shown to be immune suppressive (6, 7), and are targets of protective
74 immunity, suggesting they are important for parasite survival (8, 9).

75 Our previous proteomic analyses identified one worm (nematode)-specific AGO
76 (WAGO) in the excretory-secretory products and EVs of *H. bakeri*. Here we focus
77 on defining and characterizing the extracellular Argonaute (exWAGO) and exported
78 sRNAs. Nematode pathogens in particular may have evolved a suite of novel RNAi
79 functions based on a unique expansion of Argonaute types (i.e. the WAGOs) (10).
80 The majority of our understanding of nematode RNAi pathways is based on the
81 free-living model organism *Caenorhabditis elegans,* which has at least four types of
82 endogenous sRNAs and 25 Argonaute genes (11). In addition to miRNAs and
83 piRNAs, *C. elegans* produces small interfering RNAs (siRNAs) from exogenous or

84 endogenous double-stranded RNAs (dsRNAs). There is also a mechanism for *de*
85 *novo* generation of siRNAs by RNA-dependent RNA polymerases (RdRPs), which
86 are recruited to sRNA-target transcripts to amplify the silencing signal through the
87 generation of secondary siRNAs. The secondary siRNAs dominate the sRNA
88 content of adult *C. elegans* and have also been documented in several parasitic
89 nematode species (12-15). They are distinguished from other sRNAs by the
90 presence of a 5' triphosphate and a preference for a 5' guanine. In *C. elegans*,
91 secondary siRNAs associate with WAGOs and have been shown to be important in
92 self *versus* non-self-recognition in the germline (16-18). siRNAs can also be
93 transmitted from the soma to the germline to mediate heritable responses to
94 infection and nutrient starvation (19). *C. elegans* therefore sets a precedent for
95 involvement of WAGOs in both defence and environmental adaptation, but the
96 functions of most WAGOs and their siRNA guides remain unknown. Since many
97 nematodes are parasites, and parasitism has arisen multiple times independently
98 (20) it is possible that WAGOs could also contribute to this important lifestyle
99 innovation.

100 Here we examine the molecular and evolutionary properties of exWAGO in
101 parasitic and free-living nematodes and demonstrate that exWAGO mediates the
102 selective export of specific siRNAs in EVs. We compare the genomic origin of
103 siRNAs exported in EVs by *H. bakeri* to the resident siRNAs expressed in adults of
104 both *H. bakeri* and *C. elegans*. Our results support a model where the resident
105 sRNAs are dominated by secondary siRNAs, which are used for endogenous gene
106 regulation and control of retrotransposons. In contrast, the parasite preferentially
107 exports secondary siRNAs that are produced from newly evolved repetitive
108 elements in the genome that associate with exWAGO. This adds evolutionary
109 breadth to the handful of reports in mammalian systems suggesting RNA-binding
110 proteins are a mechanism for selective RNA export (21-23) and establishes *H.*
111 *bakeri* as a tractable model for studying extracellular sRNA biology.
112
113 **Results**
114

115 **A nematode-specific extracellular Argonaute is within extracellular vesicles**
116 **released from *H.bakeri* at several copies per EV**

117 We previously identified an Argonaute protein in the excretory-secretory and EV
118 products of *H. bakeri* based on proteomic analyses (6). Several studies in
119 mammalian systems have similarly reported Argonautes associated with EVs, in

120 some cases under specific signalling conditions (24). However, Argonautes have
121 also been reported to be contaminants that co-purify with EVs (21). In order to
122 rigorously determine whether the exWAGO that we have identified exists within
123 EVs, we used ultracentrifugation followed by flotation on a sucrose gradient for
124 purification, quantification by nanoparticle tracking analysis and visualisation by
125 transmission electron microscopy. As shown in Figure 1, the EVs had a density of
126 1.16-1.18 g/cm3 and co-purified with exWAGO. We further subjected the sucrose-
127 purified EV fractions to proteinase K treatment and confirmed that the exWAGO
128 was protected from degradation but became susceptible when the EVs were lysed
129 with detergent (Figure 1D). We analysed a defined number of sucrose-gradient
130 purified EVs by western blot in comparison to recombinant exWAGO and found
131 that exWAGO was present at 3.4 ± 1.1 copies per EV (Figure 1E).

132 **An improved genome assembly and annotation to explore extracellular**
133 **Argonautes and RNAs**

134 In order to determine the full complement of Argonautes and small RNAs in *H.*
135 *bakeri* we first generated a new genome assembly for this nematode based on
136 combining short-read (~100-fold read coverage; Illumina) and long-read (~12-fold
137 coverage; PacBio SMRT) data (Supplementary Methods). The final genome
138 assembly spans 697 Mb, 150 Mb longer than the first (Illumina-only) draft (25).
139 While most sequenced nematode genomes are between 60 and 200 Mb, the
140 strongylids (which include *H. bakeri*) tend to have larger genomes, ranging from
141 170 to 700 Mb (with a mean of ~380 Mb) (25). Our *H. bakeri* assembly is
142 represented by 23,647 contigs (just over half the previous assembly's 44,728
143 contigs), with an N50 of 180 kb (up from 36 kb). Assessment of genome
144 completeness using the Core Eukaryotic Genes Mapping Approach (CEGMA) and
145 Benchmarking Universal Single-Copy Orthologs (BUSCO) suggests ~88% of
146 conserved genes are complete (~8% partial), with 96% of the assembled *H. bakeri*
147 transcriptome mapping to the genome (Supplemental Table 1). Protein-coding
148 genes were predicted with the BRAKER pipeline generating 23,471 protein-coding
149 genes with 25,215 transcripts. Non-coding RNA genes, including rRNA, tRNA and
150 miRNAs, were predicted using Rfam models and family-specific tools (see
151 Supplementary Methods). The expansion of the *H. bakeri* genome compared to
152 closely related clade V parasites is associated with an expanded repeat content.
153 Over half (58.3%) of the *H. bakeri* genome contains some type of repeat element,
154 including LINE elements (12.6% of the genome) and DNA elements (12.8%) (Table
155 1). Of all the repeats, 33.3% were found within genes (mostly in introns, which

156 themselves occupy 33.5% of the genome). Interestingly, 30.6% of the genome was
157 annotated as unclassified repeats, nearly two-thirds of which do not overlap any
158 other kind of annotation.

159 **exWAGO is highly conserved and abundant in rhabditine (Clade V) parasitic**
160 **nematodes and has diverged in *Caenorhabditis***

161 To determine the conservation of exWAGO across Clade V nematodes we
162 clustered proteins from the new *H. bakeri* genome with proteomes predicted from
163 the genomes of a selection of rhabditomorph nematodes (including *C. elegans* and
164 six additional *Caenorhabditis* species, eleven strongyle parasites, the
165 entomopathogen *H. bacteriophora*, the free living *Oscheius tipulae*, and the free-
166 living diplogasteromorph *Pristionchus pacificus*) using OrthoFinder (26). The
167 resulting orthogroups were interrogated with Kinfin (27) to identify orthologues of
168 proteins predicted to be involved in RNAi in *H. bakeri* or known to be implicated in
169 RNAi in *C. elegans.* This revealed that, as expected, nearly all of the machinery for
170 miRNA and piRNA pathways, including highly conserved Argonautes ALG-1/2 and
171 PRG1/2, is conserved across Clade V nematodes (Supplemental Figure 1).
172 Previous studies in *C. elegans* have defined populations of primary siRNAs that are
173 26 nt in length and associated with male or female germline regulation mediated by
174 the Argonautes ALG-3/4 (during spermatogenesis) or ERGO-1 (in oocytes and
175 embryos). Notably ALG-3/4, but not ERGO-1, are conserved in parasites and the
176 general factors associated with 26G RNA biogenesis, also termed the ERI complex
177 (enhanced exogenous RNAi phenotype), are conserved while ERGO-1-specific
178 factors including ERI-6/7/9 and MUT-16 are not (Supplementary Figure 1).

179 Strikingly, of the thirteen WAGOs in the *C. elegans* gene set, only four had co-
180 clustered orthologues from species other than *Caenorhabditis*. We therefore
181 performed a joint phylogenetic analysis of all orthogroups containing Argonautes
182 (defined by the presence of both PAZ and PIWI domains) (Figure 2A). This
183 identified clades of Argonautes in parasitic species that were sister to
184 *Caenorhabditis*-specific orthogroups. For example, the nuclear WAGOs HRDE-1
185 and NRDE-3 as well as WAGO-10 and WAGO-11 in *Caenorhabditis* are in fact
186 orthologous to parasite-derived Argonautes in orthogroups OG01747 and
187 OG07955 but their relationship has been obscured by differing rates of evolution in
188 the different species groups. Similarly, the phylogenetic analysis shows that
189 exWAGO does in fact co-cluster with a *Caenorhabditis*-only orthogroup that
190 contains *C. elegans* SAGO-1, SAGO-2 and PPW-1. The orthogroup containing

191 exWAGO contains Argonautes from many other parasitic strongyles, *H.*
192 *bacteriophora*, *O. tipulae* and *P. pacificus* as well as Argonautes from
193 *Caenorhabditis* species placed at the base of the genus (*C. monodelphis*, *C.*
194 *castaneus*, and *C. sp.* 38) (Figure 2B). Examination of the intron-exon structure of
195 these Argonautes supports this relationship (Figure 2C). The most basal
196 *Caenorhabditis*, *C. monodelphis*, has a gene structure very similar to that of the
197 other exWAGOs, but gene structure in other *Caenorhabditis* species has evolved
198 rapidly. We suggest that *C. elegans* SAGO-1, SAGO-2 and PPW-1 are co-
199 orthologues of *H. bakeri* exWAGO, and thus the biology of these genes may
200 illuminate the origins and functions of exWAGO in parasites.

201 Using our new annotation of Argonautes we used existing RNAseq data to
202 determine the expression levels of all Argonautes in adult life stages of parasitic
203 versus free-living Clade V nematodes. Strikingly, we found that exWAGOs are
204 generally the most abundantly expressed of all Argonautes (Supplementary Table
205 2), including the sheep parasites *Haemonchus contortus* and *T. circumcinta* and
206 the human hookworm, *Nector americanus* (Figure 3). In contrast, the SAGO-1, 2
207 and PPW orthologs in *C. elegans* adults are not expressed at high levels
208 (Supplementary Table 2). We further identified exWAGO in the excretory-secretory
209 (ES) products of adult *Nippostrongylus brasiliensis* (another rodent parasite) (Table
210 1). No peptides mapping to any other Argonaute proteins were identified in multiple
211 samples, pointing to a unique extracellular role for this particular Argonaute across
212 the parasite species.

213 **Comparative analysis of resident sRNA distribution in *H. bakeri* versus *C.***
214 ***elegans* suggests shared functions in endogenous gene regulation**

215 To determine whether the dominance of exWAGO in the parasites was reflected at
216 the level of sRNA composition we first carried out side-by-side analysis of
217 endogenous sRNAs present in *H. bakeri* and *C. elegans* adult nematodes. sRNA
218 datasets were generated in triplicate, capturing either only 5'-monophosphate
219 RNAs or all RNAs (after treatment with 5' polyphosphatase). As expected, the
220 untreated libraries from whole nematodes were dominated by reads mapping to
221 miRNAs in each genome, having the characteristic first nucleotide preference of U
222 and peak length of 22 nt (Figure 4A,D). In contrast, the whole-nematode libraries
223 treated with 5' polyphosphatase showed a clear enrichment for RNAs with a first
224 base preference of guanine, the majority of which were 22 nt in length in *C.*
225 *elegans* and 23 nt in *H. bakeri* (Figure 4B,E). This signature is characteristic of

226 secondary siRNA products of RdRPs, and suggests secondary siRNAs dominate
227 the resident sRNA populations of both nematodes. The length variation (22 *versus*
228 23 nt) may indicate mechanistic differences between the RdRPs that generate
229 them or the Argonaute proteins that stabilize them.

230 By comparing the 5' polyphosphatase-treated and untreated libraries in both
231 species, we identified 137,531 regions in the *H. bakeri* genome that have more
232 mapped reads from the 5' polyphosphatase-treated libraries (we call these regions
233 polyP-enriched clusters) and 6,075 regions with relatively more reads from the
234 untreated libraries (monoP-enriched clusters, see Methods and Supplemental
235 Figure 2). We reasoned that these represent sRNAs with two distinct modes of
236 biogenesis, with the monoP-enriched clusters containing sRNAs cleaved by
237 ribonucleases (such as Dicer) or being degradation products, and the polyP-
238 enriched clusters containing unprocessed products of RdRPs or RNA polymerase
239 III. Consistent with this model, the monoP-enriched clusters contained a higher
240 fraction of miRNA-mapping reads than the untreated libraries, while the polyP-
241 enriched clusters had a much reduced fraction of miRNA reads (Supplemental
242 Figure 3). The same general strategy was applied to the *C. elegans* sRNAs, to
243 compare the polyP-enriched clusters of both nematodes, which represent by far the
244 most abundant type of sRNA in adults. The majority (62.3%) of the reads within the
245 polyP-enriched clusters of *C. elegans*, mapped antisense to messenger RNAs,
246 consistent with roles in regulating endogenous gene expression (Figure 4C). In
247 contrast, 9.9% of the reads from polyP-enriched clusters of *H. bakeri* mapped
248 antisense to mRNAs (Figure 4F). To compare these numbers, we need to take into
249 account the fraction of each genome occupied by mRNAs. The *C. elegans* genome
250 devotes 28.3% of its base pairs to coding exons (Table 1, Supplemental Figure 4),
251 therefore if sRNAs were produced randomly across the genome 14.1% would map
252 antisense to these. Consequently, our observed proportion of antisense mRNA
253 polyP sRNAs represents a 4.4-fold increase over what is expected by chance.
254 Since only 3.4% of the *H. bakeri* genome encodes exons, the polyP sRNAs that
255 map antisense to mRNAs are 5.8-fold more frequent than expected. Following
256 similar logic, both nematodes have a significant overrepresentation of polyP sRNAs
257 mapping antisense to known retrotransposons (7.4-fold increase in *C. elegans*, 2.5-
258 fold increase in *H. bakeri,* Figure 4C,F). Thus, despite the drastic differences in
259 genome content of the two species, there is a conserved pattern of siRNAs that
260 likely reflect common functionality in endogenous gene regulation and genome
261 defence.

262 **Vesicular siRNAs are largely derived from novel-repeat elements and**
263 **associate with exWAGO**

264 Our previous work indicated that EVs secreted from *H. bakeri* adults are associated
265 with a population of sRNAs. We characterized miRNAs and Y RNAs, but only
266 sequenced sRNAs with a 5' monophosphate (6). To generate a more
267 comprehensive characterization of EV sRNAs and examine selectivity, we
268 analysed duplicate sRNA datasets from purified EVs, capturing either only 5'-
269 monophosphate RNAs or all RNAs (after treatment with 5' polyphosphatase). To
270 ensure the sequenced sRNAs were derived from EVs, and not co-purifying or free
271 complexes, EVs were purified by ultracentrifugation and sucrose gradient prior to
272 RNA extraction, library preparation and sequencing. We detected several species
273 of miRNA in EV-derived libraries as expected from our previous work, however the
274 vast majority of EV sRNAs are 23G siRNAs, only detected with 5' polyphosphatase
275 treatment (Figure 4H). To focus on the RdRP products, we selected the polyP-
276 enriched clusters (Supplemental Figure 2). EV-derived, polyP-enriched sRNAs had
277 a 1.9-fold enrichment for siRNAs derived from transposons and a 1.7-fold
278 enrichment for species-specific repeats of the *H. bakeri* genome, compared to the
279 polyP-enriched sRNAs from adults (Figure 4I). In contrast, the siRNAs mapping
280 antisense to protein coding genes and retrotransposons were relatively depleted
281 within the EV libraries. These results suggest selective partitioning of siRNA
282 biotypes into the EVs and identifies recently evolved regions of the genome as a
283 primary source of EV sRNA.

284 To further explore and quantify this selectivity, we calculated, for each polyP-
285 enriched cluster, a measure of entropy-based Information Content (IC) using either
286 adult or EV reads (see Methods). The higher the IC value, the more concentrated
287 the reads are in a few peaks, while the lower the IC value, the more evenly
288 distributed the reads are across the cluster (e.g. Figure 5A, inset). Interestingly, the
289 IC values are consistently higher for reads coming from EVs than from adult
290 libraries (Figure 5A), indicating that the EVs more often contain reads from specific
291 peaks and are not a random sampling of the adult sRNA pool. Figure 5B illustrates
292 a region in the *H. bakeri* genome that produces siRNAs enriched in the EVs. The
293 only annotated elements in this region are repeats, mostly novel (species-specific)
294 elements. These results support the idea that EV content reflects a selection of
295 specific siRNA sequences.

296 To determine whether exWAGO specifically associates with the EV-enriched
297 sRNAs produced from novel repeats, we immunoprecipitated the adult nematode
298 lysates using an antibody raised against exWAGO and analysed by qRT-PCR the
299 co-purified sRNAs. The siRNAs that derive from EV-enriched clusters
300 immunopurified with exWAGO and were depleted in the unbound fraction (Figure
301 6). We observed the opposite pattern with the IgG bead control. In contrast,
302 siRNAs derived from selected clusters that are abundant in adults but not
303 represented in the EVs did not copurify with exWAGO, nor did Y RNAs or a
304 conserved miRNA (Figure 6). These results suggest that specific siRNA sequences
305 bind to exWAGO, which mediates their encapsulation in EVs and defines the
306 population of vesicular sRNAs that are secreted into the host environment.

307 **Discussion**
308
309 That small RNAs are transferred within organisms, and between organisms, has
310 many implications in cross-species communication and disease. However, there
311 are many questions regarding how RNAs are selected for export from the donor,
312 which RNAs are transferred to the recipient, and whether and how these RNAs
313 function within the recipient. Here we have examined the question of the specificity
314 of packaging of sRNAs in EVs in the model parasite *H. bakeri* using comparative
315 analyses of the origin of sRNAs within the body of the nematode *versus* those
316 selected to be exported in EVs. We compare this to evolutionary analyses of the
317 sRNA machinery in the parasite and the closely related, free-living *C. elegans*. We
318 find that secondary siRNAs within adults of both the free-living and the parasitic
319 nematodes are largely produced to target mRNAs and retrotransposons by
320 antisense pairing along their entire length, and are therefore associated with
321 endogenous gene regulation and defence. In contrast the siRNAs within EVs
322 secreted by the parasite do not appear to be a stochastic sampling of those
323 detected in the adult nematodes but are specifically enriched for those produced
324 from transposons and newly-evolved, repetitive regions in the genome. This
325 suggests both mechanistic and evolutionary selectivity.

326 Our immunoprecipitation experiments show that Y RNAs, which are also abundant
327 in EVs (6), do not associate with exWAGO, suggesting this protein is specific for
328 siRNAs. Mechanistically, we envision three processes that could contribute to the
329 total RNA present in EVs, acting independently or together. The EVs could be
330 passively loaded with the sRNAs present in the cell type from which EVs are
331 exported. It is likely that EVs are released from the intestine (6). Secondly, the

332 sRNAs could be actively loaded by some intrinsic property, perhaps related to their
333 specific biogenesis pathway. Lastly, the sRNAs could associate with a specific
334 RNA-binding protein, as has been shown in some mammalian systems. Our
335 immunoprecipitation data suggest that associative binding occurs for the siRNAs
336 and we identify exWAGO as the mediator of this selective export. Intriguingly
337 exWAGO is highly conserved and abundant in all Clade V parasitic nematodes
338 examined, and we have further shown that it is also secreted in the rodent parasitic
339 nematode *N.brasiliensis*. We propose therefore that the mechanism of exWAGO-
340 mediated siRNA export extends beyond the *H. bakeri* model.

341 The sRNAs selected for export with exWAGO derive from regions of the *H. bakeri*
342 genome that are repetitive and novel, which may reflect recent, dynamic evolution
343 of this putative host manipulation system. Rather than derive sRNAs from
344 conserved loci, and risk self-directed effects, selection may exploit the rapidly
345 evolving non-genic portion of the genome to generate evolutionarily novel but host-
346 relevant sRNA loci. It will be informative to correlate these sequences with host
347 genes and to explore their evolution across parasites.

348 We identified the SAGO and PPW proteins in *C. elegans* and related
349 *Caenorhabditis* species as diverged exWAGO orthologues. Although as yet we
350 have no evidence that the SAGOs (or any other AGOs) are exported extra-
351 organismally from *C. elegans,* preliminary data suggest these may both have
352 common localization in the intestine ((6) and Claycomb, Seroussi, unpublished). It
353 will be of interest to understand whether SAGO and PPW function within *C.*
354 *elegans* can shed light on the roles of exWAGO.

355 Very little is understood regarding the evolution of cross-species communication.
356 That pathogens use small RNAs to modulate their hosts is not unexpected, as it
357 has been well documented in interactions between parasitic fungi and plants (28)
358 and it is similar, conceptually, to the evolution of miRNAs in certain viruses (29). In
359 contrast to viruses, however, extracellular parasites such as *H. bakeri* require a
360 mechanism for transporting specific sRNAs into host cells. The packaging of
361 siRNAs in EVs by exWAGO provides such a mechanism. Notably, EVs have
362 recently been implicated in the transfer of RNA in plants, in this case from the plant
363 cells to fungal parasites (30). We do not yet know if exWAGO is solely involved in
364 export, or is also involved in mediating functional effects inside the recipient cells.
365 Further work is required to understand the individual and collective contributions of
366 all of the EV cargos in host cell modulation. This work establishes a parasite

367 Argonaute as a sorting mechanism for EV RNAs, indicates that focusing only on
368 miRNAs can be misleading and provides an important framework for interrogating
369 new parasite-host interactions and their consequences on infection.

370

371

372 **Materials and Methods are provided in Supplementary Material.**

373

374 **Acknowledgements**

375 This work was supported by HFSP grant RGY0069 to AB, CA and JC. We thank
376 Elaine Robertson for technical support for the *H. bakeri* life cycle, Sujai Kumar for
377 support with genome analysis, Tuhin Maity for preparation of *C. elegans* samples
378 and Rick Maizels for ES from *N. brasiliensis*.

379 **Competing interests**

380 The authors declare no competing interests.

381 **Figure legends**

382

383 **Figure 1: exWAGO is vesicular and present at multiple copies per EV**

384 A) Western blot analysis of equal volumes of sucrose-gradient fractions of EVs
385 from *H. bakeri* using antibody against exWAGO, B) Silver stain blot of same
386 fractions, C) Nanoparticle tracking analysis of EV total number in each fraction (left)
387 and TEM of 1.16 g/cm3 fraction (right). D) Western blot of exWAGO from gradient-
388 purified EVs and following treatment with Proteinase K (5 ug/mL) with or without
389 Triton-X (0.05%). E) Western blot of 3 independent biological replicates of sucrose-
390 gradient purified *H. bakeri* EVs, using recombinant standard of exWAGO for
391 quantification.

392 **Figure 2: Phylogenetic tree of Argonautes in Clade V nematodes and gene**
393 **structure of exWAGO.** A) Grey shading denotes different orthogroups, *C. elegans*
394 protein names in each clade are noted (or absent if no orthologues in that Clade).
395 B) Tree showing phylogenetic relationship and branch lengths of exWAGO
396 orthologues across Clade V, C) Conservation of exons and introns in exWAGO
397 homologues. Each box is an exon with the width denoting length. Boxes with
398 dashed lines denote exons with possible errors in the genome assembly of the
399 species. Colours denote differences in exon size in triplets compared to exWAGO.

400 **Figure 3: Expression of Argonautes across Clade V parasitic nematodes**
401 Relative expression levels of Argonautes from RNAseq data of the adult parasites
402 noted. Data were based on the sum of tpm reads for each orthogroup (defined in
403 Figure 2), normalized to tpm for OG1273 orthogroup (ALG-1/2). The total number
404 of distinct transcripts in each orthogroup in each species is noted below each

405 column. The known *C.elegans* Argonaute names are used where applicable, or
406 exWAGO as defined in this work.

407 **Figure 4: sRNA composition in adult *C. elegans* and *H. bakeri*, compared to**
408 ***H. bakeri* extracellular vesicles.** First nucleotide and length distribution of
409 untreated small RNA libraries for *C. elegans* adults (A), *H. bakeri* adults (D) and *H.*
410 *bakeri* EVs (G), and their corresponding polyphosphatase-treated libraries (B,E,H).
411 The proportions of the 20-25 nt reads mapping within annotated categories in the
412 genome (from Table 1) are shown beneath each barplot. Line plots for *C. elegans*
413 (C) and *H. bakeri* (F) showing the relationship between the percentage of the
414 genome occupied by each annotation category and the percentage of 20-25 nt
415 reads from the polyP-enriched clusters, while (I) shows the relationship between
416 the percentage of the 20-25 nt reads from the adult polyP-enriched clusters to
417 those from the EV polyP-enriched clusters (see Methods).

418 **Figure 5: Clusters with transposons or novel repeats have higher Information**
419 **Content in extracellular vesicles than in adults.** Dot plot comparing Counts Per
420 Million and Information Content of all clusters with transposons or novel repeats
421 (A). Top and side barplots show the number of clusters at each value of the X or Y-
422 axis respectively. Inset: example of read coverage for cluster ncRNA_44089. Read
423 coverage (top), annotation of repeat elements (middle) and zoomed-in read
424 coverage (bottom, in $\log_2$-scale to distinguish individual libraries) for the most highly
425 expressed cluster in EVs (B). In all cases blue indicates EV and red indicates Adult
426 libraries.

427 **Figure 6: Immunoprecipitation of exWAGO and detection of associated**
428 **sequences.** A) Western blot to detect exWAGO following immunoprecipitation of
429 10 ug adult worm lysates with exWAGO anti sera or control (naïve) sera.
430 Equivalent volumes input and unbound were loaded (unbound is defined as first
431 flow-through from beads). B) qRT-PCR analysis of samples from (A) for siRNAs
432 derived from EV-enriched or adult-enriched clusters as well as Y-RNA and miR-
433 100. To ensure equivalent recovery of RNA, a synthetic spike was included prior to
434 extraction which varied <2 fold across all sample types. Data are shown as mean
435 with standard deviation for n=3.

436

437

438

439 **Supplementary Figure 1 Conservation of RNAi pathway in Clade V**

440 Presence or absence of orthologues to *C. elegans* genes associated with RNAi
441 pathways in Clade V organisms. Phylogenetic relationship is shown at the top of
442 table, gene identities across each species are detailed in
443 https://github.com/DRL/chow2018.
444

445 **Supplementary Figure 2:** Expression of all sRNA clusters (dots) comparing the
446 average counts-per-million (X-axis) to the fold-change of polyphosphatase-treated
447 relative to untreated libraries (Y-axis). Blue and red dots highlight those clusters
448 identified respectively as polyP (enriched in polyphosphatase treated libraries) or
449 monoP (similar normalised expression between treated and untreated libraries) for
450 *C. elegans* adult nematodes (A), and *H. bakeri* adults (B) or extracellular vesicles
451 (C). Clusters containing known miRNAs (expected to be monoP) are highlighted in
452 gold.

453 **Supplementary Figure 3:** Distribution of reads in annotated categories for each
454 type of library, comparing the reads falling within monoP or polyP clusters to all
455 reads. Plots are shown for untreated libraries for *C. elegans* (A) and *H. bakeri* (B),
456 and polyphosphatase-treated libraries for *C. elegans* (C) and *H. bakeri* (D).

457 **Supplementary Figure 4:** Comparison of *C. elegans* and *H. bakeri* genome sizes
458 and fractions of each genome devoted to annotated.

459 **Supplementary Table 1**: New *H.bakeri* genome assembly information

460 **Supplementary Table 2:** Details of RNAseq data used for Figure 3

461

462

463

464

465 **Table 1. Non-overlapping genome composition**

	C. elegans		*H. bakeri*	
	bases	%	*bases*	%
intergenic	29,355,343	29.272	168,174,849	24.130
exons	28,409,938	28.329	23,682,169	3.398
introns	25,956,467	25.882	170,798,580	24.506
transposons	11,880,919	11.847	92,998,044	13.343
novel repeats	1,402,898	1.399	13,1452,375	18.861
retroelements	1,307,772	1.304	104,917,538	15.054
satellite repeats	594,478	0.593	377,947	0.054
simple repeats	581,880	0.580	2,960,599	0.425
other ncRNA	429,956	0.429	749,784	0.108
piRNA	256,574	0.256	63,990	0.009
tRNA	62,284	0.062	675,650	0.097
miRNA	35,312	0.035	43,955	0.006
rRNA	9,520	0.009	53,983	0.008
yRNA	3,060	0.003	5,940	0.001
TOTAL	100,286,401	100%	696,955,403	100%

466

467 **Table 2. ExWAGO identified in EV products by mass spectrometry**

Species	Protein	Accession number	Length (aa)	Predicted MW (kDa)	Predicted PI	Unique peptide sequence (Start position)
H. bakeri	exWAGO	HPOL_0000298601-mRNA-1	912	102	9.23	TGMGQLSVGAVALPEKR (6)
						SAAVAVYK (86)
						AAVLFSAQR (114)
						QFMLPASVVSSAGPDATGIR (132)
						ISQMSIFFDQR (277)
						NAMQPFNQK (297)
						VTLQQQTPDQVASMIK (393)
						ASATLPQTR (409)
						IMKDALDITPR (423)
						AATTIAPR (716)
						LVNDGDLK (899)
N. brasiliensis	exWAGO	NBR_exWAGO	913	102	9.25	QDFVCNLTALK (32)
						DIFPQDSALFYDR (102)
						ILPTPTILYGER (457)

468 **References**

469

470 1. Swarts DC, *et al.* (2014) The evolutionary journey of Argonaute proteins.
471 *Nature structural & molecular biology* 21(9):743-753.
472 2. Tkach M & Thery C (2016) Communication by Extracellular Vesicles: Where
473 We Are and Where We Need to Go. *Cell* 164(6):1226-1232.
474 3. Liu S, *et al.* (2016) The Host Shapes the Gut Microbiota via Fecal
475 MicroRNA. *Cell host & microbe* 19(1):32-43.
476 4. LaMonte G, *et al.* (2012) Translocation of sickle cell erythrocyte microRNAs
477 into Plasmodium falciparum inhibits parasite translation and contributes to
478 malaria resistance. *Cell host & microbe* 12(2):187-199.
479 5. Behnke JM, Menge DM, & Noyes H (2009) Heligmosomoides bakeri: a
480 model for exploring the biology and genetics of resistance to chronic
481 gastrointestinal nematode infections. *Parasitology* 136(12):1565-1580.
482 6. Buck AH, *et al.* (2014) Exosomes secreted by nematode parasites transfer
483 small RNAs to mammalian cells and modulate innate immunity. *Nature*
484 *communications* 5:5488.
485 7. Eichenberger RM, *et al.* (2018) Hookworm Secreted Extracellular Vesicles
486 Interact With Host Cells and Prevent Inducible Colitis in Mice. *Frontiers in*
487 *immunology* 9:850.
488 8. Coakley G, *et al.* (2017) Extracellular Vesicles from a Helminth Parasite
489 Suppress Macrophage Activation and Constitute an Effective Vaccine for
490 Protective Immunity. *Cell reports* 19(8):1545-1557.
491 9. Shears RK, Bancroft AJ, Hughes GW, Grencis RK, & Thornton DJ (2018)
492 Extracellular vesicles induce protective immunity against Trichuris muris.
493 *Parasite immunology*:e12536.
494 10. Buck AH & Blaxter M (2013) Functional diversification of Argonautes in
495 nematodes: an expanding universe. *Biochemical Society transactions*
496 41(4):881-886.
497 11. Youngman EM & Claycomb JM (2014) From early lessons to new frontiers:
498 the worm as a treasure trove of small RNA biology. *Frontiers in genetics*
499 5:416.
500 12. Holz A & Streit A (2017) Gain and Loss of Small RNA Classes-
501 Characterization of Small RNAs in the Parasitic Nematode Family
502 Strongyloididae. *Genome biology and evolution* 9(10):2826-2843.
503 13. Pak J & Fire A (2007) Distinct populations of primary and secondary
504 effectors during RNAi in C. elegans. *Science* 315(5809):241-244.
505 14. Sarkies P, *et al.* (2015) Ancient and novel small RNA pathways compensate
506 for the loss of piRNAs in multiple independent nematode lineages. *PLoS*
507 *biology* 13(2):e1002061.
508 15. Yigit E, *et al.* (2006) Analysis of the C. elegans Argonaute family reveals
509 that distinct Argonautes act sequentially during RNAi. *Cell* 127(4):747-757.
510 16. Ashe A, *et al.* (2012) piRNAs can trigger a multigenerational epigenetic
511 memory in the germline of C. elegans. *Cell* 150(1):88-99.
512 17. Lee HC, *et al.* (2012) C. elegans piRNAs mediate the genome-wide
513 surveillance of germline transcripts. *Cell* 150(1):78-87.
514 18. Shirayama M, *et al.* (2012) piRNAs initiate an epigenetic memory of nonself
515 RNA in the C. elegans germline. *Cell* 150(1):65-77.
516 19. Rechavi O & Lev I (2017) Principles of Transgenerational Small RNA
517 Inheritance in Caenorhabditis elegans. *Current biology : CB* 27(14):R720-
518 R730.
519 20. Blaxter ML, *et al.* (1998) A molecular evolutionary framework for the phylum
520 Nematoda. *Nature* 392(6671):71-75.

521 21. Shurtleff MJ, Temoche-Diaz MM, Karfilis KV, Ri S, & Schekman R (2016) Y-
522 box protein 1 is required to sort microRNAs into exosomes in cells and in a
523 cell-free reaction. *eLife* 5.
524 22. Santangelo L*, et al.* (2016) The RNA-Binding Protein SYNCRIP Is a
525 Component of the Hepatocyte Exosomal Machinery Controlling MicroRNA
526 Sorting. *Cell reports* 17(3):799-808.
527 23. Villarroya-Beltri C*, et al.* (2013) Sumoylated hnRNPA2B1 controls the
528 sorting of miRNAs into exosomes through binding to specific motifs. *Nature*
529 *communications* 4:2980.
530 24. McKenzie AJ*, et al.* (2016) KRAS-MEK Signaling Controls Ago2 Sorting into
531 Exosomes. *Cell reports* 15(5):978-987.
532 25. Consortium IHG (2018) Comparative genomics of the major parasitic
533 worms. *bioRxiv* 36539; doi: https://doi.org/10.1101/236539.

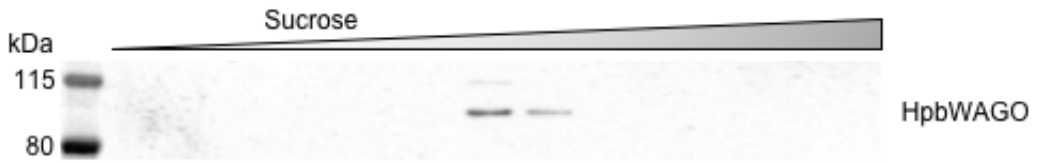
534 26. Emms DM & Kelly S (2015) OrthoFinder: solving fundamental biases in
535 whole genome comparisons dramatically improves orthogroup inference
536 accuracy. *Genome biology* 16:157.
537 27. Laetsch DR & Blaxter ML (2017) KinFin: Software for Taxon-Aware
538 Analysis of Clustered Protein Sequences. *G3* 7(10):3349-3357.
539 28. Weiberg A*, et al.* (2013) Fungal small RNAs suppress plant immunity by
540 hijacking host RNA interference pathways. *Science* 342(6154):118-123.
541 29. Pfeffer S*, et al.* (2004) Identification of virus-encoded microRNAs. *Science*
542 304(5671):734-736.
543 30. Cai Q*, et al.* (2018) Plants send small RNAs in extracellular vesicles to
544 fungal pathogen to silence virulence genes. *Science* 360(6393):1126-1129.

545

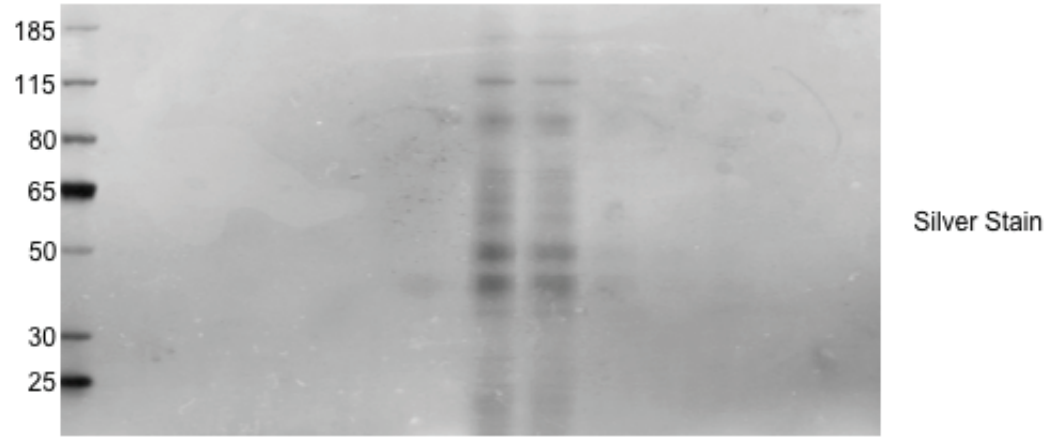
546

Figure 1

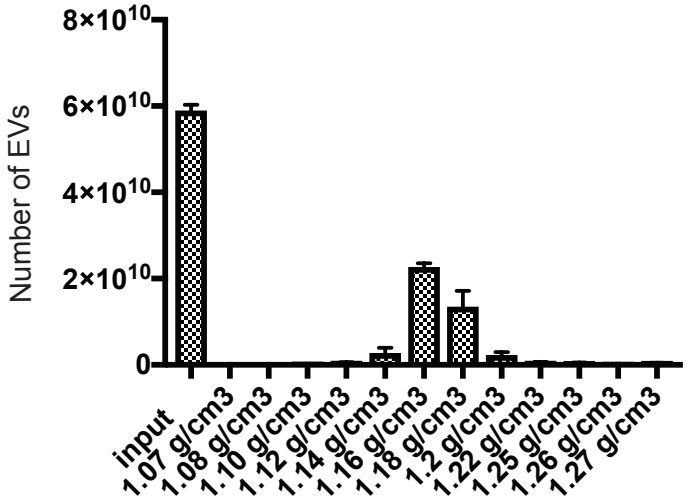
A



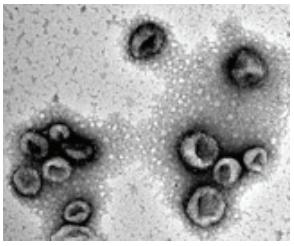
B



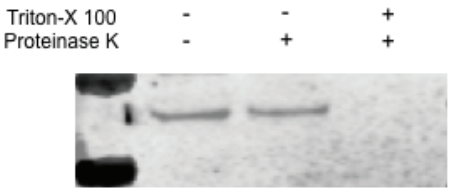
C



Fraction at 1.16 g/cm3



D



E

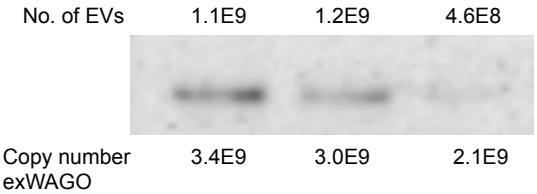


Figure 2

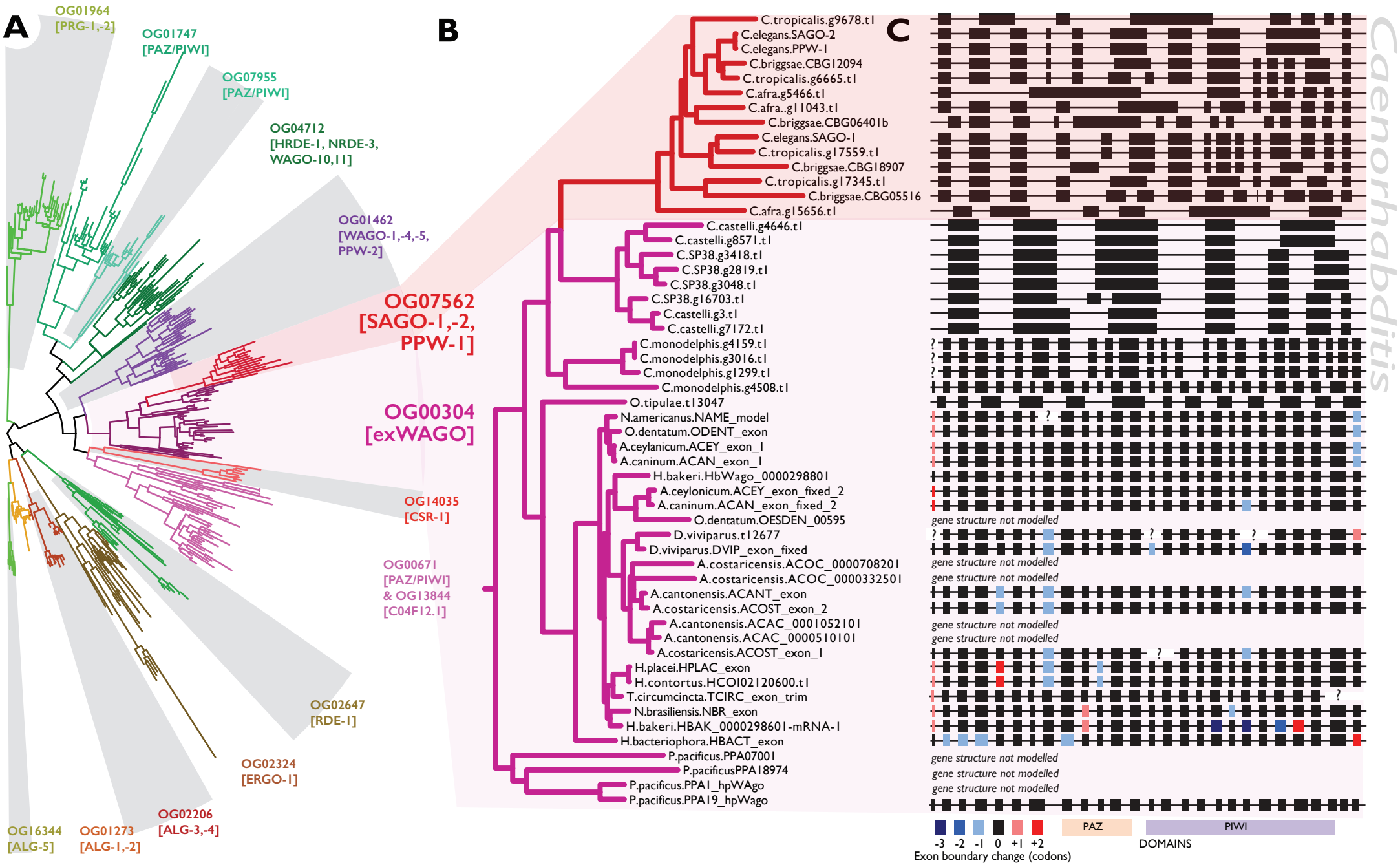

A
OG01964
[PRG-1,-2]
OG01747
[PAZ/PIWI]
OG07955
[PAZ/PIWI]
OG04712
[HRDE-1, NRDE-3,
WAGO-10,11]
OG01462
[WAGO-1,-4,-5,
PPW-2]
OG07562
[SAGO-1,-2,
PPW-1]
OG00304
[exWAGO]
OG14035
[CSR-1]
OG00671
[PAZ/PIWI]
& OG13844
[C04F12.1]
OG02647
[RDE-1]
OG02324
[ERGO-1]
OG02206
[ALG-3,-4]
OG16344
[ALG-5]
OG01273
[ALG-1,-2]
B
C.tropicalis.g9678.t1
C.elegans.SAGO-2
C.elegans.PPW-1
C.briggsae.CBG12094
C.tropicalis.g6665.t1
C.afra.g5466.t1
C.afra..g11043.t1
C.briggsae.CBG06401b
C.elegans.SAGO-1
C.tropicalis.g17559.t1
C.briggsae.CBG18907
C.tropicalis.g17345.t1
C.briggsae.CBG05516
C.afra.g15656.t1
C.castelli.g4646.t1
C.castelli.g8571.t1
C.SP38.g3418.t1
C.SP38.g2819.t1
C.SP38.g3048.t1
C.SP38.g16703.t1
C.castelli.g3.t1
C.castelli.g7172.t1
C.monodelphis.g4159.t1
C.monodelphis.g3016.t1
C.monodelphis.g1299.t1
C.monodelphis.g4508.t1
O.tipulae.t13047
N.americanus.NAME_model
O.dentatum.ODENT_exon
A.ceylanicum.ACEY_exon_1
A.caninum.ACAN_exon_1
H.bakeri.HbWago_0000298801
A.ceylonicum.ACEY_exon_fixed_2
A.caninum.ACAN_exon_fixed_2
O.dentatum.OESDEN_00595
D.viviparus.t12677
D.viviparus.DVIP_exon_fixed
A.costaricensis.ACOC_0000708201
A.costaricensis.ACOC_0000332501
A.cantonensis.ACANT_exon
A.costaricensis.ACOST_exon_2
A.cantonensis.ACAC_0001052101
A.cantonensis.ACAC_0000510101
A.costaricensis.ACOST_exon_1
H.placei.HPLAC_exon
H.contortus.HCOI02120600.t1
T.circumcincta.TCIRC_exon_trim
N.brasiliensis.NBR_exon
H.bakeri.HBAK_0000298601-mRNA-1
H.bacteriophora.HBACT_exon
P.pacificus.PPA07001
P.pacificusPPA18974
P.pacificus.PPA1_hpWAgo
P.pacificus.PPA19_hpWago
C
Caenorhabditis
gene structure not modelled
PAZ
PIWI
DOMAINS
-3 -2 -1 0 +1 +2
Exon boundary change (codons)

Figure 3

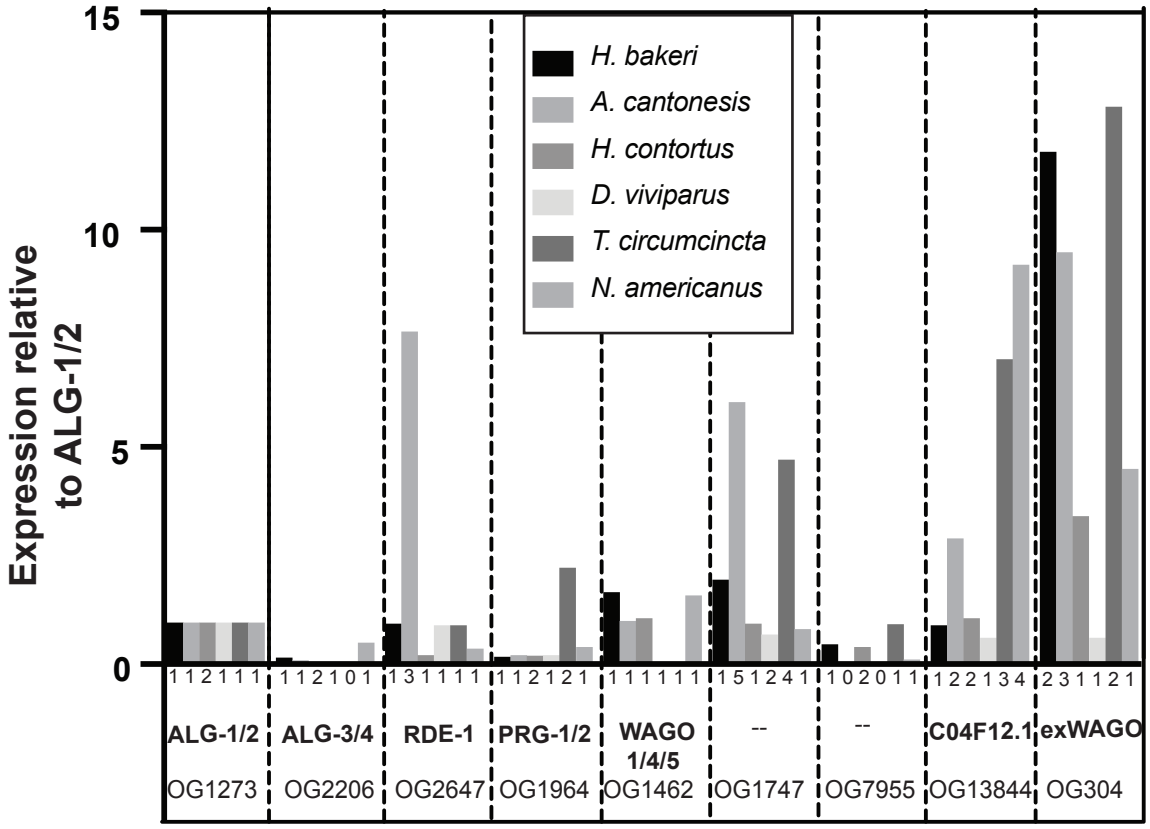
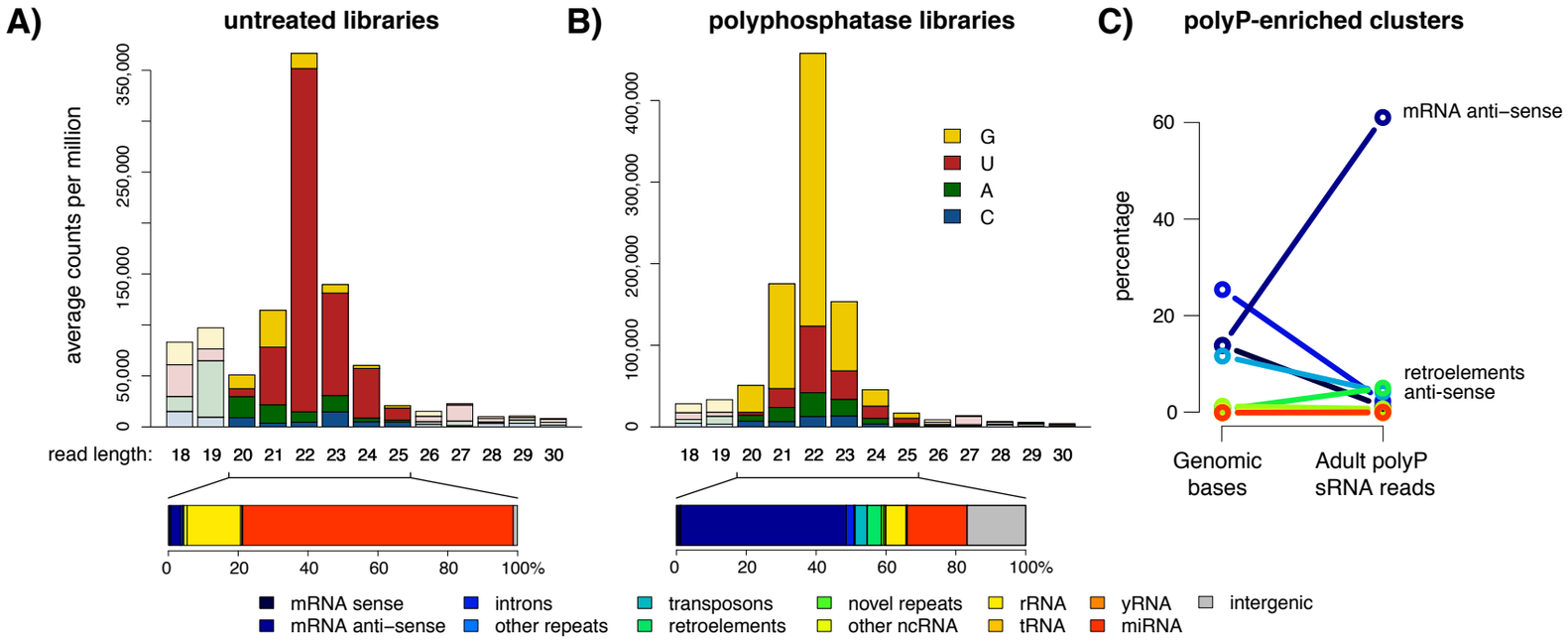
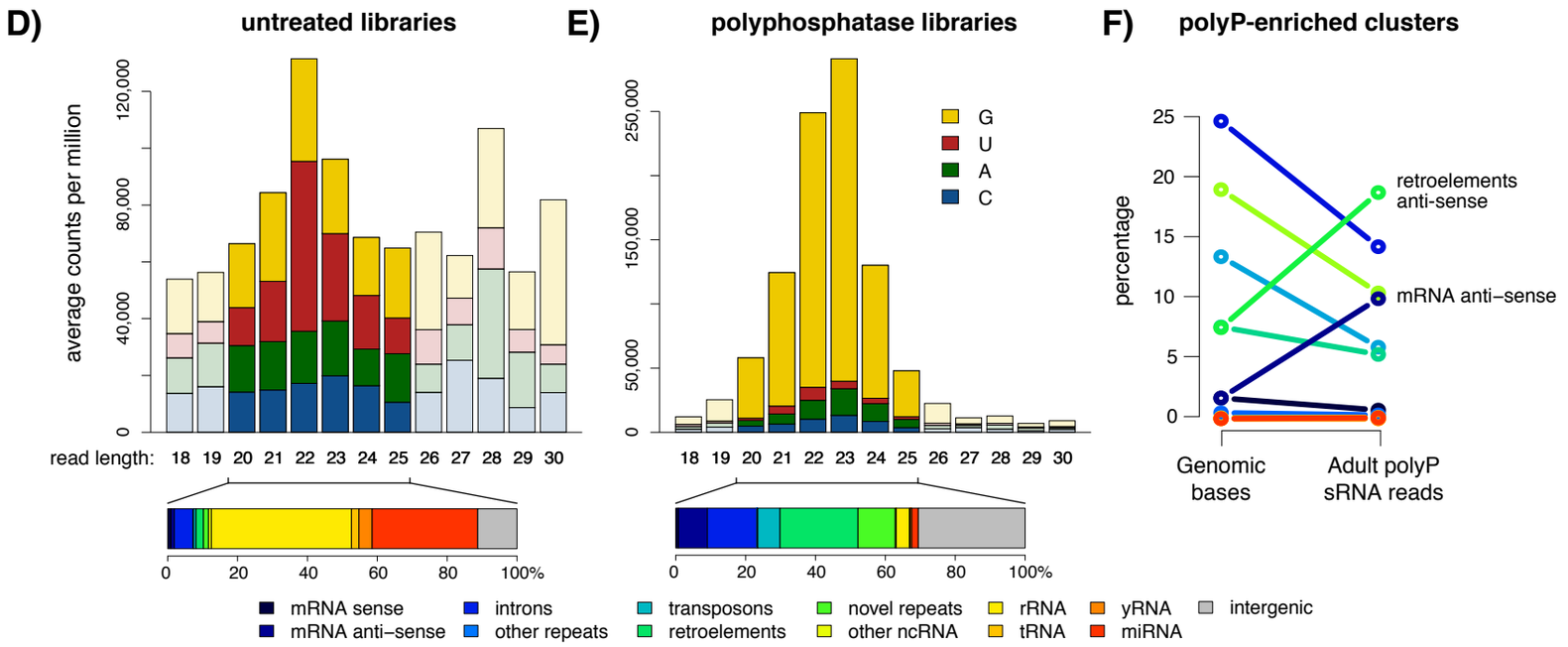

Expression relative to ALG-1/2
15
10
5
0
H. bakeri
A. cantonesis
H. contortus
D. viviparus
T. circumcincta
N. americanus
1 1 2 1 1 1
1 1 2 1 0 1
1 3 1 1 1 1
1 1 2 1 2 1
1 1 1 1 1 1
1 5 1 2 4 1
1 0 2 0 1 1
1 2 2 1 3 4
2 3 1 1 2 1
ALG-1/2
ALG-3/4
RDE-1
PRG-1/2
WAGO 1/4/5
--
--
C04F12.1
exWAGO
OG1273
OG2206
OG2647
OG1964
OG1462
OG1747
OG7955
OG13844
OG304

Figure 4

C. elegans, adult libraries



H. bakeri, adult libraries



H. bakeri, vesicle libraries

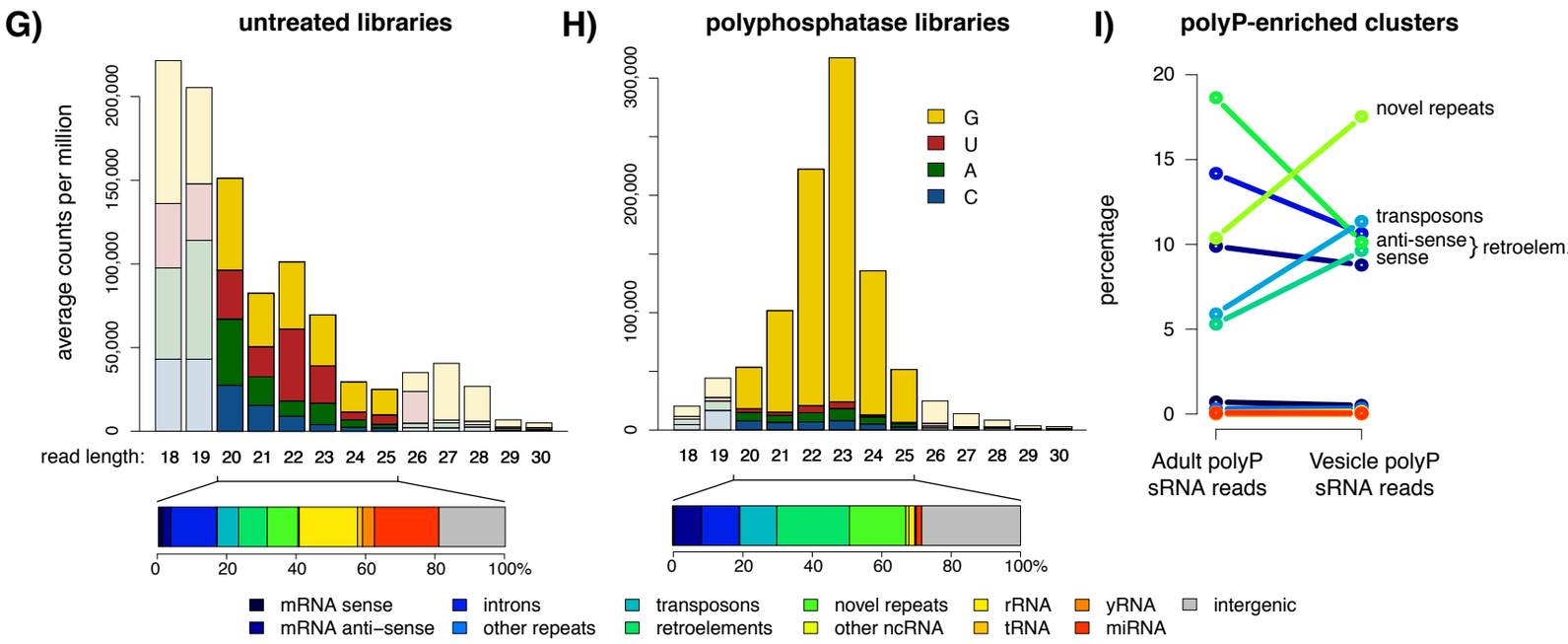


Figure 5

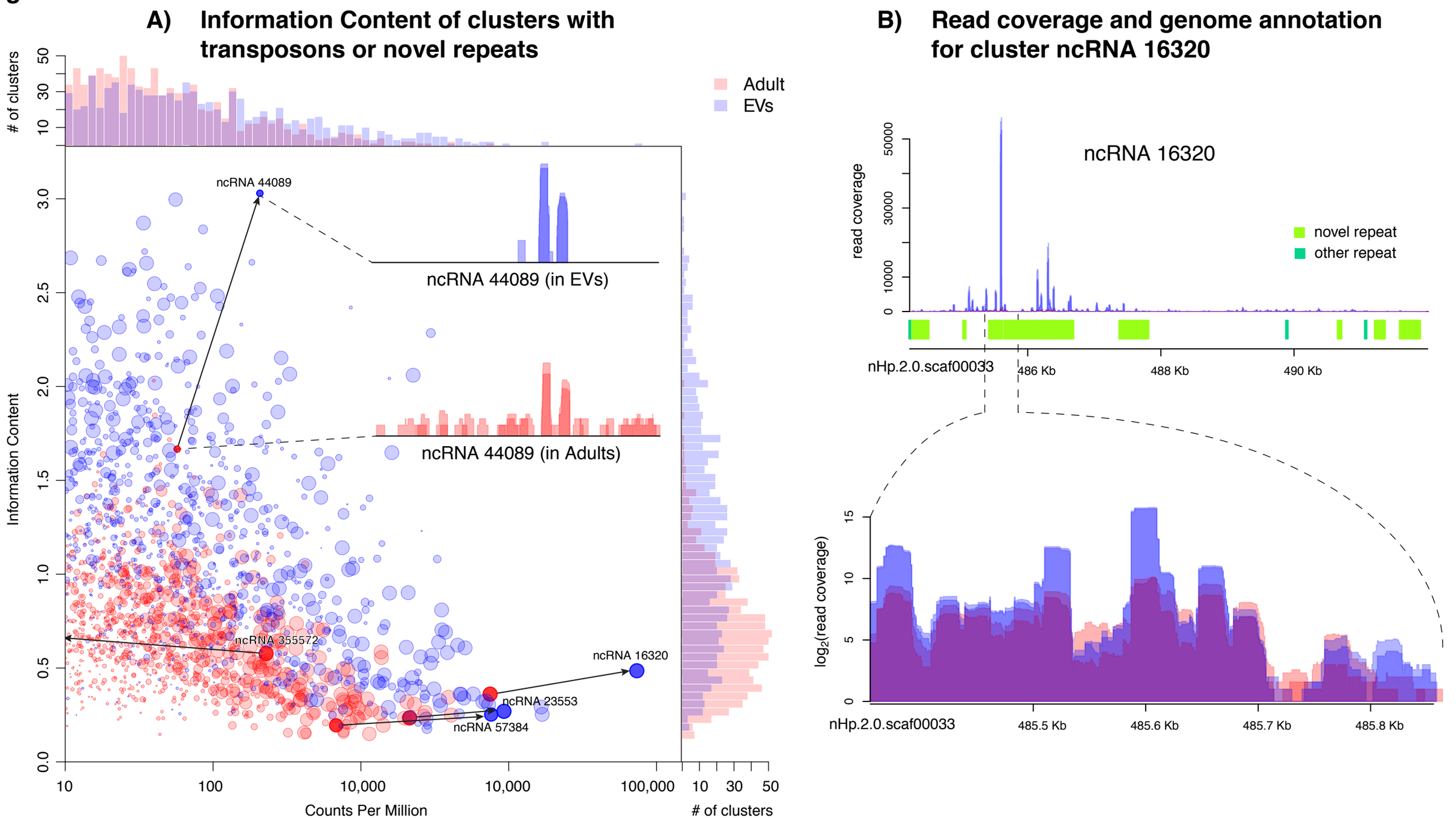
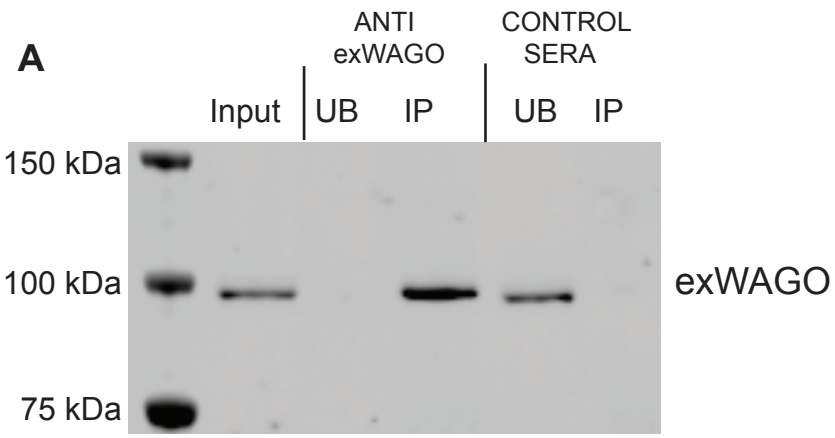
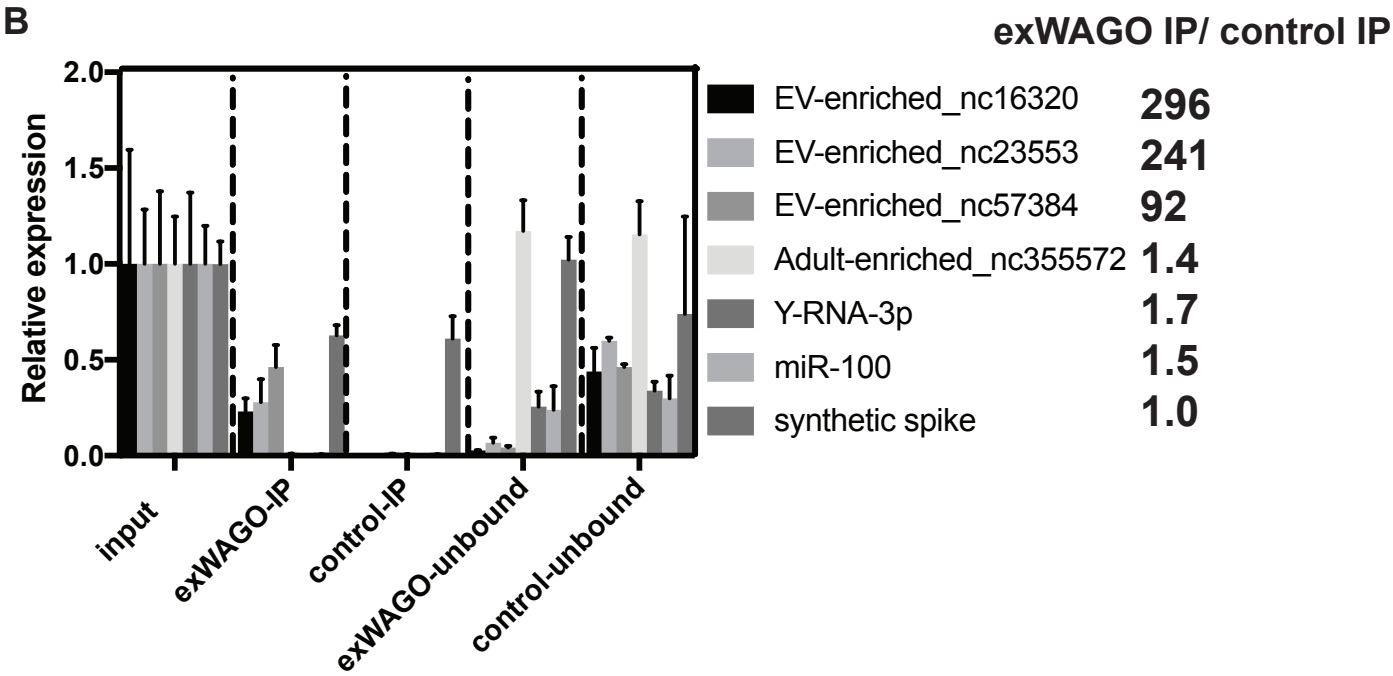

A) Information Content of clusters with transposons or novel repeats
B) Read coverage and genome annotation for cluster ncRNA 16320
Adult
EVs
# of clusters
Information Content
Counts Per Million
ncRNA 44089
ncRNA 44089 (in EVs)
ncRNA 44089 (in Adults)
ncRNA 355572
ncRNA 16320
ncRNA 23553
ncRNA 57384
ncRNA 16320
read coverage
novel repeat
other repeat
nHp.2.0.scaf00033
486 Kb
488 Kb
490 Kb
log2(read coverage)
485.5 Kb
485.6 Kb
485.7 Kb
485.8 Kb

Figure 6


A
ANTI
exWAGO
CONTROL
SERA
Input
UB
IP
UB
IP
150 kDa
100 kDa
75 kDa
exWAGO

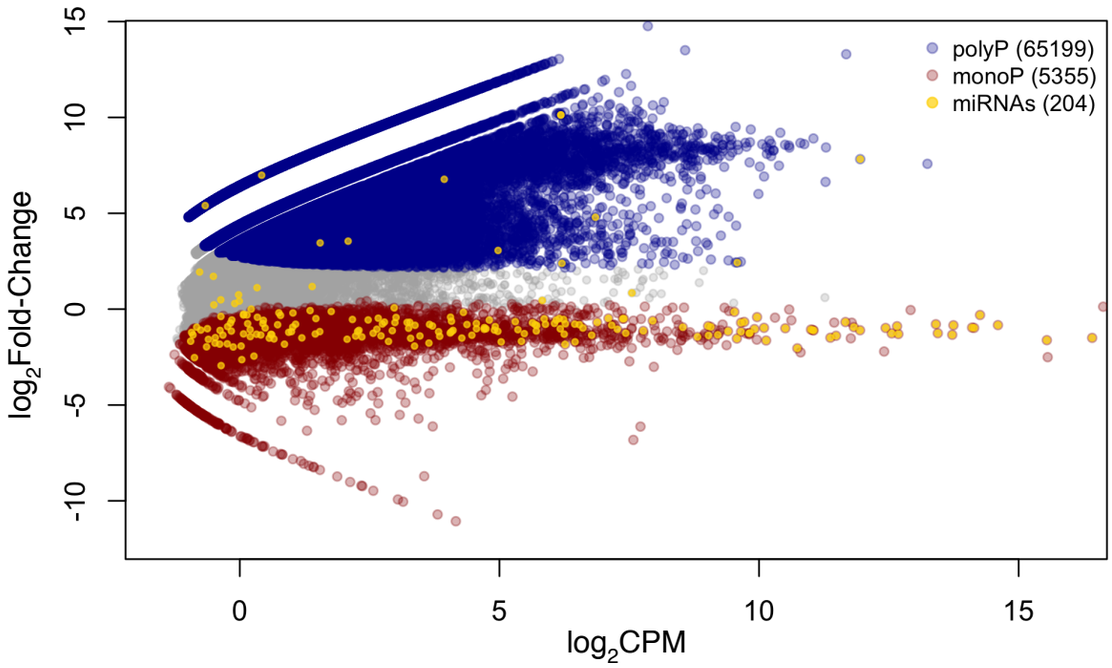

B
exWAGO IP/ control IP
Relative expression
2.0
1.5
1.0
0.5
0.0
input
exWAGO-IP
control-IP
exWAGO-unbound
control-unbound
EV-enriched_nc16320 296
EV-enriched_nc23553 241
EV-enriched_nc57384 92
Adult-enriched_nc355572 1.4
Y-RNA-3p 1.7
miR-100 1.5
synthetic spike 1.0

Supplementary Figure 1

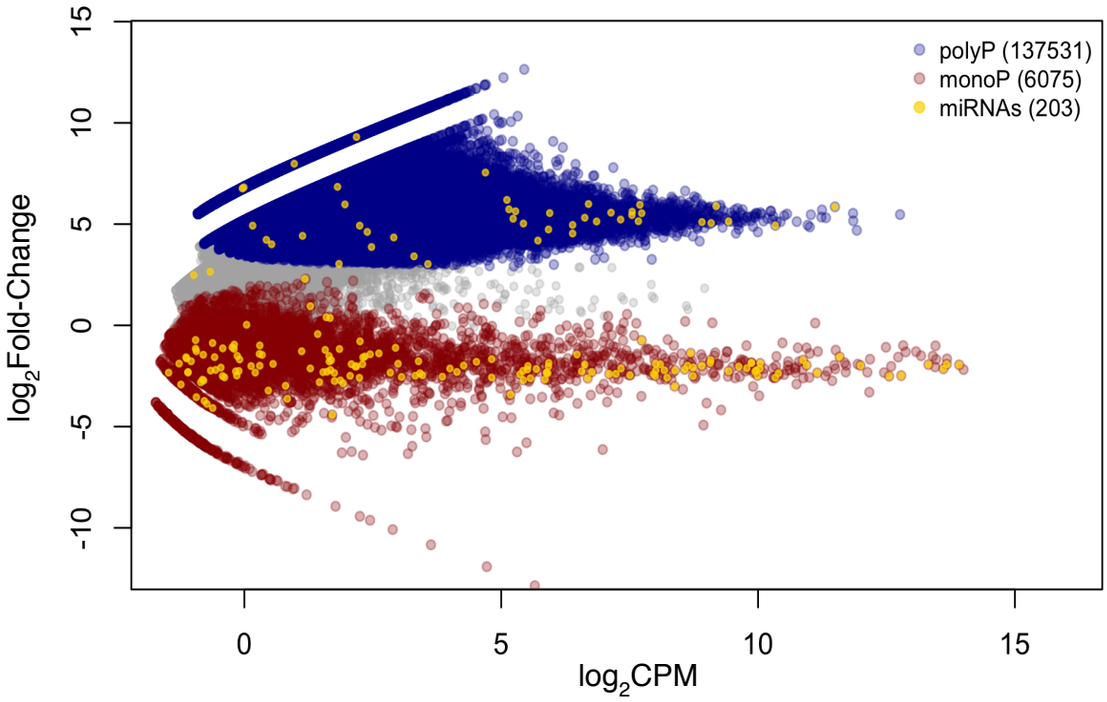
Pathway	Description	Gene	Orthogroup	CELEG	CBRIG	CTROP	CAFRA	CCAST	CSP38	CSP1	OTIPU	HBAKE	NBRAS	TCIRC	HPLAC	HCONT	ACOST	ACANT	DVIVI	ODENT	NAMER	ACEYL	HBACT	PPACI
miRNA	Primary Argonaute	alg-2	OG0001273	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Primary Argonaute	alg-1	OG0001273	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Primary Argonaute	alg-5/hpo-24	OG0016344	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Dicer	dcr-1	OG0001117	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	nuclease	drsh-1	OG0005079	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□
	RBP	vig-1	OG0004036	■	■	■	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	RBP	pash-1	OG0005640	■	■	■	■	■	■	□	■	■	■	■	■	■	■	■	■	□	■	■	□	■
	nuclease	tsn-1	OG0006505	■	■	■	■	■	■	■	■	■	□	■	■	■	■	■	■	■	■	■	■	■
	cofactors	nhl-2	OG0001917	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■	■
	mRNA decapping	dcs-1	OG0008205	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	□	■	■	■	■	■
	helicase	cgh-1	OG0002514	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	ain-2	OG0010872	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	ain-1	OG0003857	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
26G (ERGO-1)	Primary Argonaute	ergo-1	OG0002324	■	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	■
	helicase	eri-7	OG0014061	■	■	□	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	helicase	eri-6	OG0014716	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	RNA transferase	eri-9	OG0014727	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	mut-16	OG0023428	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
26G (ALG-3/4)	Primary Argonaute	alg-4	OG0002206	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■
	Primary Argonaute	alg-3	OG0002206	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■
26G	methyltransferase	henn-1 (ERI-complex)	OG0002498	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Dicer	dcr-1 (ERI-complex)	OG0001117	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	dsRNA binding	rde-4 (ERI-complex)	OG0005821	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■	■	■
	cofactors	eri-5 isoform a (ERI-complex)	OG0026790	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	eri-5 isoform b (ERI-complex)	OG0023542	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	RDRP	rrf-3 (ERI-complex)	OG0003948	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■	■	■	□	■
	exonuclease	eri-1 (ERI-complex)	OG0001143	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■	■
	helicase	drh-3 (ERI-complex)	OG0001120	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	eri-3 (ERI-complex)	OG0010894	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
piRNA	Primary Argonaute	prg-2	OG0001964	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Primary Argonaute	prg-1	OG0001964	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	transcription factor	fkh-5	OG0018311	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	transcription factor	fkh-4	OG0018311	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	transcription factor	fkh-3	OG0018311	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	methyltransferase	henn-1	OG0002498	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	prde-1	OG0014911	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	pid-1	OG0014928	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	tofu-1	OG0015117	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	tofu-2	OG0007430	■	■	■	■	■	■	■	□	■	■	□	■	■	■	■	■	□	□	■	■	■
	cofactors	tofu-3/ulp-5	OG0016132	■	■	■	■	□	□	■	■	■	■	■	■	■	■	■	■	■	□	□	■	□
	cofactors	tofu-4	OG0015040	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	tofu-5	OG0005839	■	■	■	■	■	■	■	□	■	■	■	■	■	■	□	■	■	■	■	■	■
	cofactors	tofu-6	OG0014857	■	■	■	■	□	□	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	tofu-7	OG0013824	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
Exo RNAi	Primary Argonaute	rde-1	OG0002647	■	■	■	■	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	RBP	sid-1	OG0006373	■	■	■	■	■	■	■	□	■	■	□	■	■	□	□	□	■	■	■	■	□
	cofactors	sid-2	OG0013828	■	■	■	■	■	■	■	□	□	□	■	■	■	□	□	□	■	□	□	□	□
	tyrosine kinase	sid-3	OG0007024	■	■	■	■	■	■	■	■	■	■	□	□	■	□	■	■	□	■	■	□	□
	Dicer	dcr-1	OG0001117	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	RBP	rde-4	OG0005821	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■	■	■
	RDRP	rrf-1	OG0000233	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	helicase	drh-3	OG0001120	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	helicase	drh-1	OG0001604	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	□
CSR-1 22G	CSR-1 Argonaute	csr-1	OG0014035	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	RDRP	ego-1	OG0000233	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	Tudor domain	ekl-1	OG0003151	■	■	□	■	■	■	■	■	□	■	□	■	■	■	■	■	■	■	■	■	■
	Poly(U) Polymerase	cid-1/cde-1	OG0011716	■	■	■	■	■	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□
	helicase	drh-3	OG0001120	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Secondary 22G	Secondary Argonaute	wago-5	OG0001462	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	□	■
	Secondary Argonaute	wago-4	OG0001462	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	□	■
	Secondary Argonaute	wago-1	OG0001462	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	□	■
	Secondary Argonaute	sago-2/wago-6	OG0007562	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	sago-1/wago-8	OG0007562	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	ppw-2	OG0001462	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	□	■
	Secondary Argonaute	ppw-1	OG0007562	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	C04F12.1	OG0013844	■	□	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	wago-11	OG0004712	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	wago-10	OG0004712	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	nrde-3 (nuclear)	OG0004712	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	Secondary Argonaute	hrde-1 (nuclear)	OG0004712	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	helicase	mut-14	OG0002648	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	helicase	smut-1	OG0002648	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	nuclease	mut-7	OG0001312	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	nucleotidyltransferase	mut-2/rde-3	OG0004469	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■
	Tudor domain	ekl-1	OG0003151	■	■	□	■	■	■	■	■	□	■	□	■	■	■	■	■	■	■	■	■	■
	cofactors	mut-8/rde-2	OG0023472	■	□	□	□	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	mut-16	OG0023428	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	helicase	drh-3	OG0001120	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	RDRP	rrf-1	OG0000233	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	RDRP	ego-1	OG0000233	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	rde-11	OG0008695	■	■	■	■	■	□	■	■	■	■	□	■	■	■	■	■	□	■	■	■	□
	cofactors	rde-10	OG0010879	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	mut-15	OG0016083	■	■	■	■	□	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	helicase	mut-14	OG0002648	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	hpl-2 isoform a (nuclear machinery)	OG0001228	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	hpl-2 isoform b (nuclear machinery)	OG0015237	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	mes-3 (nuclear machinery)	OG0020222	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	mes-4 (nuclear machinery)	OG0002364	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■	■	■	■
	methyltransferase	set-32 (nuclear machinery)	OG0001718	■	■	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	methyltransferase	set-25 (nuclear machinery)	OG0004709	■	■	■	■	■	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	■
	cofactors	nrde-1 (nuclear machinery)	OG0012361	■	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	nrde-4 (nuclear machinery)	OG0022866	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	nrde-2 (nuclear machinery)	OG0014088	■	■	■	□	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□
RNAi Spreading	cofactors	rsd-6	OG0018055	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	rsd-2	OG0012321	■	■	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
	cofactors	rsd-3	OG0003052	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cofactors	sid-5	OG0017703	■	■	■	■	□	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□

Supplemental Figure 2

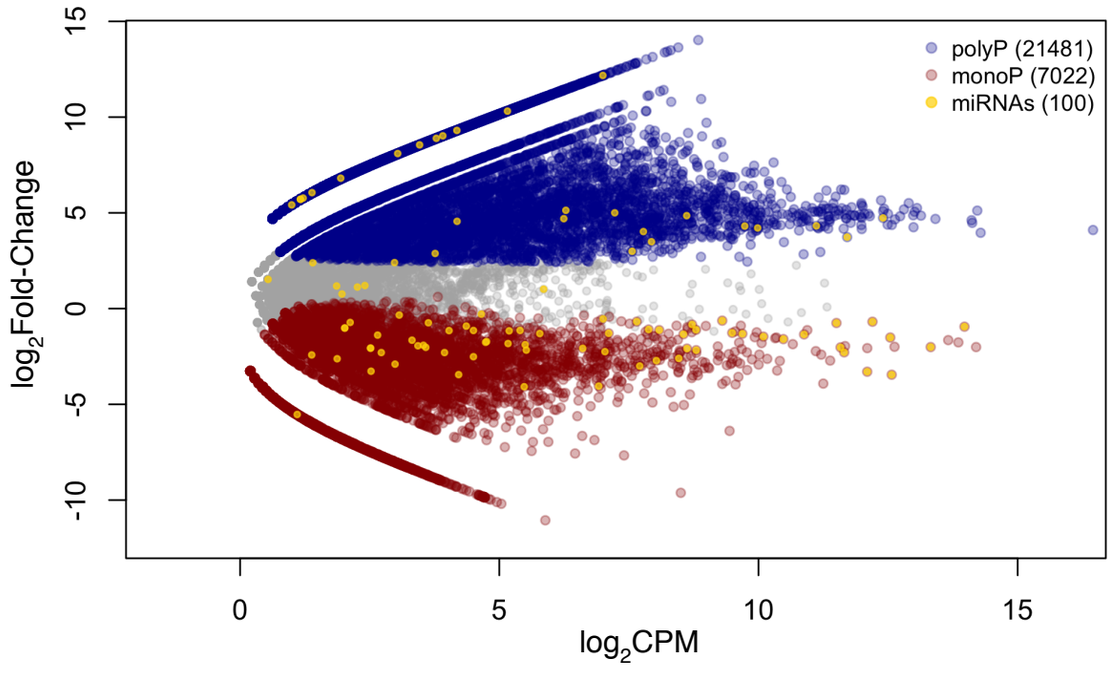
A) *C. elegans* Adult libs, monoP and polyP-enriched clusters



B) *H. bakeri* Adult libs, monoP and polyP-enriched clusters

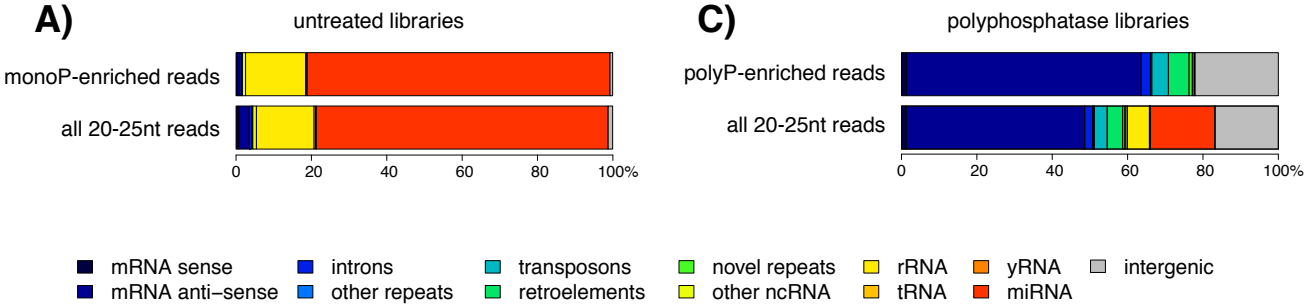


C) *H. bakeri* Vesicle libs, monoP and polyP-enriched clusters

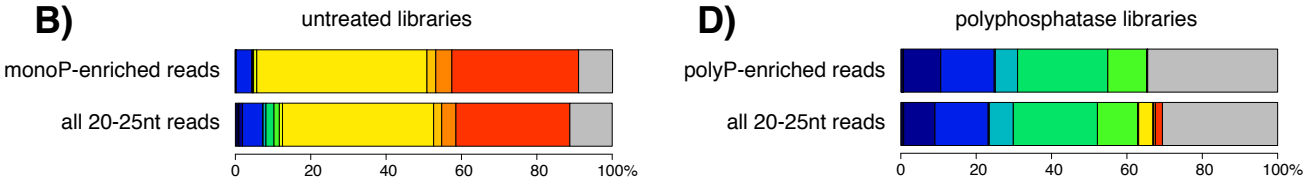


Supplemental Figure 3

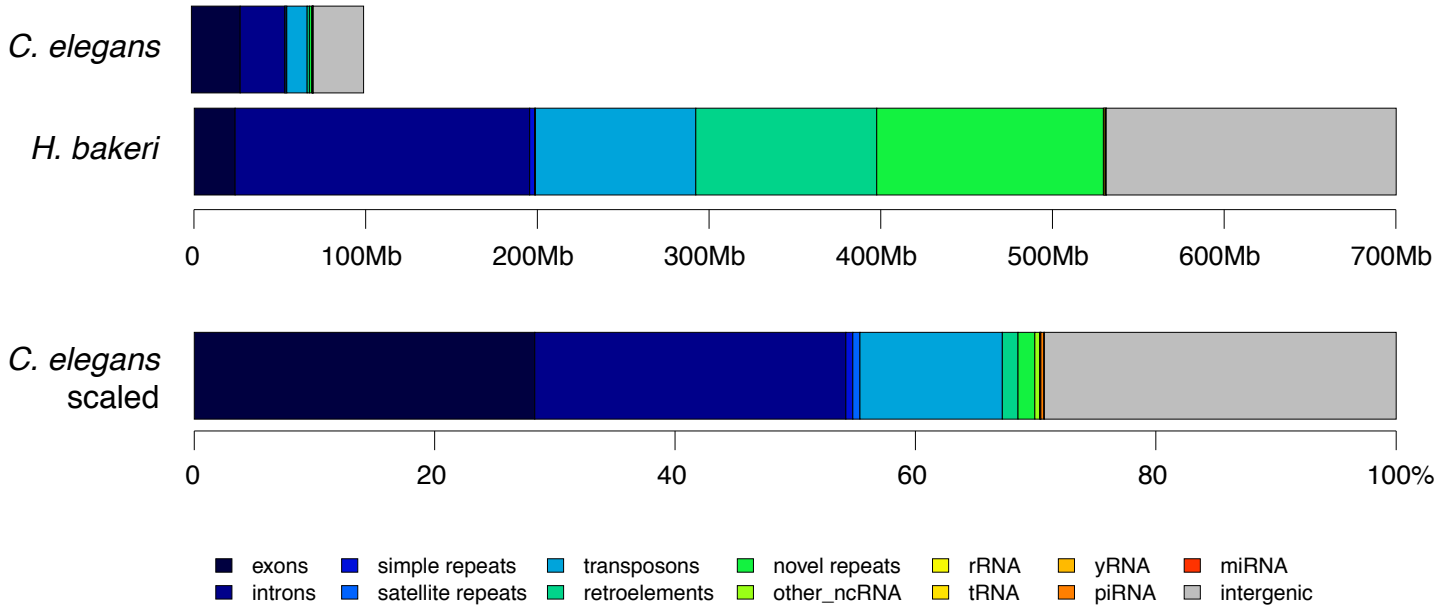
C. elegans, adult libraries



H. bakeri, adult libraries



Supplemental Figure 4



Supplemental Table 1: *Heligmosomoides bakeri* genome assembly

Feature	***Heligmosomoides bakeri* genome assembly v2.0**	***Heligmosomoides bakeri* genome assembly v1.0**
Reference	This work	WTSI
Span (Mb)	696	560
G+C content (%)	45.6	45.0
Scaffold / contig N50 (kb)	179.6 / 42.6	35.8 / 12.8
Number of contigs	23647	44728
Genome CEGMA complete / partial (%)	88.7 / 8.1	78.8 / 18.1
Genome BUSCO (Nematoda) complete / partial (%)	87.1 / 7.2	67.8 / 10.7
Genome BUSCO (Eukaryota) complete / partial (%)	87.8 / 1.7	74.3 / 8.9
Transcriptome mapping	96.3%	72.3%
Number of protein-coding genes	24371	27459