Supplementary methods

H. bakeri culture and DNA extraction

CBA x C57BL/6 F1 (CBF1) mice were infected with L3 infective-stage *H. bakeri* larvae by gavage and adult nematodes were collected from the small intestine 14 days post infection. The nematodes were washed and maintained in serum-free media *in vitro* as described (1). To extract genomic DNA, nematodes were washed twice with sterile PBS and resuspended in Puregene cell lysis buffer (Qiagen) before being triturated by hand using a sterile mortar and pestle under liquid nitrogen. The ground nematode extract was thawed and digested with 100 µg proteinase K (Qiagen) at 65°C with gentle shaking overnight. RNA was removed by subsequent digestion at 37°C for 1hr with RNase A (100 µg; Qiagen). Puregene Protein Precipitation Solution (Qiagen) was added and the digests incubated on ice for 5 min. Precipitate was removed by centrifugation at 13,000 rpm at 4°C for 10 min, and the supernatant removed to a new tube. Isopropanol was added to the supernatant and samples placed at −20°C for at least 1 hr, followed by centrifugation and ethanol precipitation. Re-suspended DNA was treated with 5 µl RiboShredder mix (Epicentre) at 37°C for 2 hr prior to subsequent purification with the Zymo Research Genomic DNA Clean & Concentrator kit following manufacturer's instructions.

Sequencing

Genomic DNA integrity and molecular weight were verified by agarose gel electrophoresis and two template libraries with 10 kb inserts were prepared. PacBio sequencing was carried out on the PacBio RS II platform following the standard protocol with a C2 sequencing kit at the 1×120-min acquisition mode (Centre for Genomic Research at University of Liverpool, UK). The run was carried out with diffusion-based loading and analysed using standard PacBio primary data analysis. Illumina short-read data were generated from independent DNA preparations, using 350 base and 550 base insert libraries on a HiSeq2500 instrument (paired end 125 base reads) (Edinburgh Genomics at University of Edinburgh, UK).

RNA preparation from adult worms and RNA sequencing

C. elegans cultures were grown at a density of 100,000 worms per plate, on 150mm plates with 2ml of 5X concentrated OP50 to the gravid adult stage (containing embryos; 65h post L1 at 20°C), and harvested and flash frozen as in

(2). *H. bakeri* were collected as above and disrupted in 700 µl of Qiazol (Qiagen) using mechanical disruption with 5 mm stainless steel beads (Qiagen) on a Tissue Lyser II (30hz for 2 min twice; Qiagen). Total RNA was extracted using a miRNAeasy mini kit (Qiagen) following the manufacturer's instruction. RNA was treated with Turbo DNA-free kit (Thermo Fisher) to remove residual DNA. Libraries for small RNA sequencing were prepared using the CleanTag small RNA library prep kit according to manufacturer's instruction, using total RNA from adult worms (30 ng), and sucrose-gradient purified EVs (equivalent $1X10^{10}$ EVs measured by Nanosight, Malvern). For all samples, 1:12 dilutions of both adapters were used with 18 amplification cycles (TriLink BioTechnologies). Libraries of the length between 140-170bp were size selected and sequenced on an Illumina HiSeq high output v4 50bp SE in Edinburgh Genomics. For stranded mRNA sequencing Illumina TruSeq was carried out by Edinburgh Genomics and the libraries sequenced on an MiSeq (reagent Kit v2) with 100PE in Edinburgh Genomics.

Genome assembly and annotation

The quality of Illumina reads was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapters were removed with Cutadapt (3), low-quality bases were trimmed with Trimmomatic (4), and reads were error-corrected with BLESS (5) using a kmer size of 21. A preliminary assembly was performed with Velvet (6) and library insert sizes confirmed after read mapping using bowtie2 (7). The Velvet assembly was inspected for contamination with blobtools (8), with no contamination being detected. The short Illumina data were assembled and gapfilled using Platanus (9), and scaffolded using transcriptome evidence (described below) with SCUBAT2 (Koutsovoulos G. SCUBAT2. https://github.com/GDKO/SCUBAT2) Long-read PacBio data (a total of 10.2 Gb from reads with an N50 of 9,411 bases) were used to further scaffold and gapfill the assembly with PBJelly (10). RNA-Seq reads were assembled with Trinity (11) resulting in 38,777 transcripts in 30,483 gene components. The transcriptome assembly was filtered based on expression (>1 TPM) and isoform percentage (>0%) calculated by kallisto (12), reducing the dataset to 32,595 transcripts. Transcripts that could encode a protein with at least 50 amino acids were selected to use with SCUBAT2. RepeatModeler (Smit, AFA, Hubley, R. RepeatModeler http://www.repeatmasker.org) and RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker http://www.repeatmasker.org) were used to identify and mask repeated regions in the genome prior to gene finding. We used the BRAKER (13) pipeline to predict protein-coding genes using the RNA-

Seq reads as evidence. We combined the BRAKER general feature format (gff) file and the transcriptome assembly within MAKER2 (14) to predict untranslated regions of transcripts (UTRs) and remove low quality gene predictions. This curated set was used for downstream analyses.

Identification of *H. bakeri* exWAGO orthologues in other Rhabditomorpha

Loci orthologous to the *H. bakeri* secreted WAGO (exWAGO) were identified using BLAST (15) searches of the genome-derived proteomes of *Haemonchus contortus*, *Necator americanus* and *Pristionchus pacificus*. The gene model of each identified orthologue was then evaluated with RNA-Seq data from the relevant species, and corrected if necessary. Using an alignment of these four proteins, a custom hidden Markov model (HMM) was constructed to identify homologues in other nematode genomes using HMMer (16) within GenePS (Koutsovoulos G https://github.com/jgraveme/GenePS). The discovered gene models were corrected based on alignment of the HMM profile, *de novo* Augustus (17) prediction, the previous predicted gene models and RNA-Seq data if available. Orthologues in *Caenorhabditis* species were validated through analysis of reciprocal best BLAST matches. Protein sequences of exWAGO orthologues were aligned with MAFFT (18) and the alignment was analysed with PHYML (using the LG+G model) (19, 20). Bootstrap support was calculated from 100 bootstrap replicates.

Ortholog clustering

Protein sequences of 21 nematode species were retrieved from the sources specified in Supplemental Table 3. Sequences below a length of 30 residues and containing more than one non-terminal stop codon were removed. Manually annotated *H. bakeri* exWAGO orthologues were added to the corresponding proteomes. Sequence similarity searches were performed using BLAST v2.4.0+ (-evalue 1e-5 -outfmt '6' -seg yes -soft_masking true -use_sw_tback). Protein clustering was carried out using OrthoFinder v1.1.4 (21) under the MCL inflation value of 3.0. Functional annotation of proteins was carried out via InterProScan v5.22-61.0 (22) against PFAM v30.0 (23) and SignalP-Euk v4.1 (24). Orthogroups were analysed using KinFin v1.0.3 (25) by providing functional annotation and the phylogenetic tree of the taxa. The orthogroups were screened based on previously described *C. elegans* WAGOs (26) using the KinFin script get_count_matrix.py. Output files are deposited at https://github.com/DRL/chow2018.

Proteomics of excretory-secretory products

Total protein from three replicates *N. brasiliensis* ES (2.5-5ug EVs and 5 ug supernatant) was loaded on a 4-12% Tris-Bis NuPAGE gel (Invitrogen) and electrophoresed before overnight in-gel digestion as described (27). Peptide extracts were dried by Speedvac and the dried peptide samples were re-suspended in MS-loading buffer (0.5% trifluoroacetic acid in water) then filtered before HPLC-MS analysis. Analysis was performed using an online system of a nano-HPLC (Dionex Ultimate 3000 RSLC, Thermo-Fisher) coupled to a QExactive mass spectrometer (Thermo-Fisher) with a 300µm x 5mm pre-column (Acclaim Pepmap, 5µm particle size) joined with a 75µm x 50cm column (Acclaim Pepmap, 3µm particle size). Peptides were separated using a multi-step gradient of 2–98% buffer B (80% acetonitrile and 0.1% formic acid) at a flow rate of 300 nL/min over 90 min. Data from MS/MS spectra was searched using MASCOT against a *N. brasiliensis* databases (WormBase Parasite). The parameters used in each search were: (i) missed cut = 2, (ii) fixed cysteine carbamidomethylation modification, (iii) variable methionine oxidation modification, (iv) peptide mass tolerance of 10ppm, (v) fragment mass tolerance of 0.05 Da. Search results were exported using a significance threshold (p-value) of less than 0.05 and a peptide score cut off of 20.

EV purification, proteinase K sensitivity and western blot analysis

For collection of *H. bakeri* EV, culture media from the adult worms was collected and purified following 3 days in culture every 3 days for a maximum of 9 days and purified as previously described (excluding the first 24 hours in culture) [15]: eggs were removed by spinning at 400 g and media was then filtered through 0.22mm filter (Millipore) followed by ultracentrifugation at 100,000 g for 2 h in polyallomer tubes at 4 °C in a SW40 rotor (Beckman Coulter). Pelleted material was washed two times in filtered PBS at 100,000 g for 2 h and re-suspended in PBS. The pelleted materials were mixed with 1.5 mL 2.5 M sucrose solution, and overlaid with a linear sucrose gradient (2.0 M – 0.4 M sucrose in PBS). Gradients were centrifuged 18-20h at 192,000g in a SW40 rotor (k-factor 144.5) (Beckman Coulter, Brea, CA). For RNA extraction and small RNA library preparation, the two fractions with densities of 1.16 – 1.18 g/cm^3 (as calculated from measured reflective index by refractometry) were pooled, diluted 10 times in PBS and centrifuged again at 192,000g for 90 min in a SW40 rotor (k-factor 144.5) prior to resuspension in PBS.

Proteinase K sensitivity experiments were performed using 1.5ug of gradient-purified EVs. These were mixed with either 5 ug /mL Proteinase K (Epicentre) alone, or in the presence of 0.05% Triton X-100 at 37°C for 30 min. Untreated

samples were also included as a control. Western Blot analysis was used to check the presence of exWAGO. Samples were separated by SDS-PAGE and transferred to Immobilon-FL membranes (Millipore) using a Trans-Blot System (Bio-Rad). Membranes were blocked in Tris-buffered saline containing 1% Tween 20 and 5% BSA (Fraction V, Fisher Scientific) for 2 hr at room temperature prior to incubation with polyclonal exWAGO antibody (generated and purified against peptides TKQTKDDFPEQERK, Eurogentec) overnight at 4°C, followed by incubation with a IRDye 680RD Goat anti-Rabbit IgG (LI-COR) secondary antibody for 1 hr at room temperature. Odyssey (LI-COR Biosciences) was used for visualization. Quantification was carried out with a recombinant version of full length exWAGO containing an N-terminal 3XFlag-His tag.

Nanoparticle tracking analysis (NTA)

NTA was carried out using a NanoSight LM14 instrument (Malvern Instruments, Malvern, UK). Vesicles were diluted with 0.1 µm-filtered PBS prior to analysis. NTA 2.2 software was used to record and analyse the samples. The camera level was set to 15 and the detection threshold to 5. Minimum expected particle size, blur and minimum track length were set to auto. The background extraction was set to On. Three measurements with 60s recording were taken for each samples. The mean and standard mean error (SEM) were calculated and plotted using GraphPad Prism 7 (GraphPad Software).

Transmission electron microscopy (TEM)

For visualization of the vesicles, the purified *H. bakeri* EVs were fixed in 2% paraformaldehyde (PFA), deposited on Formvar-carbon-coated EM grids and treated with glutaraldehyde before treatment with uranyl oxalate and methyl cellulose as described elsewhere previously (28) and then viewed in a Philips CM120 TEM. Images were taken on a Gatan Orius CCD camera.

Annotation of known families of ncRNA in *H. bakeri* by homology

To expand the annotation of *H. bakeri* into the non-coding realm, we predicted ncRNA families with Infernal 1.1.1 (29) using covariance models from Rfam 12.0 (30). For transfer RNA annotation, we used tRNAscan-SE 1.3.1 (31), and RNAmmer 1.2 (32) for ribosomal RNA. We downloaded all mature sequences from miRBase 21 (33) to annotate known miRNAs with MapMi 1.5.9 (34). The sequences of predicted and curated miRNAs and piRNAs from a previous

publication (35) were also transferred to the new genome using BLAST and requiring perfect hits. yRNA annotation is based on our previous report (35).

Processing of sRNA-seq data

For this publication, we sequenced and analysed a total of 14 new sRNA-seq libraries from *H. bakeri.* These consisted of: adult worms (3 standard, 3 with 5' RNA polyphosphatase treatment), purified extracellular vesicles (EVs) by sucrose gradients (2 standard, 2 with 5' RNA polyphosphatase), and we also included the input pellet and supernatant before sucrose purification (2 each with 5' RNA polyphosphatase treatment). In parallel we also prepared and sequenced 6 new *C. elegans* sRNA-seq libraries from adult worms (3 standard and 3 with 5' RNA polyphosphatase treatment). Before proceeding we checked all libraries for their quality with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We then used reaper (36) to remove the Illumina small RNA Adapter sequences, and PullSeq 1.0.2 (https://github.com/bcthomas/pullseq) to keep reads that were at least 16 nucleotides long. In order to map all 14 trimmed libraries to the *H. bakeri* genome and the 6 libraries to the *C. elegans* genome, we used the alignment component of ShortStack 3.8.3 (37), with parameters: --nostitch, --mismatches 2, --mmap u, --bowtie_m 500 and --ranmax 500. The reference genome for *H. bakeri* genome was the one described in this paper, and for *C. elegans* we downloaded the c_elegans.PRJNA13758.WBPS7.genomic.fa.gz file from https://parasite.wormbase.org.

Annotation of ncRNA in *H. bakeri* using sRNA-seq data

We also used all our sRNA-seq results to predict novel ncRNA producing regions. For this, we used two different approaches: ShortStack 3.8.3 (37) and miRDeep2 (38). For mapping and analysis by ShortStack, we combined all libraries for each genome as a single input, and used the following parameters: --pad 10, --mincov 10, --dicermin 18, --dicermax 32. ShortStack will thus predict regions or clusters on the genome where many sRNAs are mapped. We also used miRDeep2 to discover new miRNAs. For the miRDeep2 alignment procedure we used the combination of all libraries and parameters: -c -j -l 18 –m -q. For the miRDeep identification procedure, we used parameters: -a 100, -g 500. As input sequences, we used the high confidence *H. bakeri* mature sequences from our previous publication (35) and all mature and hairpin sequences from miRBase 21 (33).

Annotation segmentation of the genome

In order to understand how different parts of the genome led to sRNA production, we produced a non-overlapping segmentation of the genomic annotation. At the end, every base of the genome was assigned to a single type of annotation. For bases that had more than one type of overlapping annotation, we defined a simple hierarchy to choose the preferred one. The hierarchy consisted of: miRNA > yRNA > piRNA > tRNA > rRNA > snRNA > snoRNA > other_ncRNA > known retrotransposons > known transposons > mRNA exons > mRNA introns > satellite repeats > novel repeats > low complexity and simple repeats. For example, if a base overlapped with a miRNA and an intron on the same strand, it was assigned as a miRNA. Different annotation was allowed on either strand, so a single base pair could be assigned as an intron on one strand, and as a miRNA on the other strand. This was accomplished within R, extensively using objects and functions from the GenomicRanges package (39). The total number of bases assigned to each annotation type in both genomes is presented in Table 1. For all analyses that required defining a single annotation for each read (Figure 4), we used the annotation of the base aligned to the central position of the read.

Definition and classification of sRNA-producing clusters in the genome

To define discrete sRNA-producing regions in the genome, we selected ShortStack clusters as our main reference loci. Some of these clusters overlapped exon-intron boundaries. Since we were interested in distinguishing the sRNA reads coming from introns (possible degradation products) and exons (potential siRNAs), we split all overlapping clusters at exon-intron boundaries. A total of 417,292 clusters were defined for the *H. bakeri* genome, and 103,278 for *C. elegans*. We next added the annotation predicted by Rfam, tRNAscan-SE, RNAmmer, MapMi and miRDeep2 by calculating overlaps between GRanges objects, using the GenomicFeatures R package (39). For the clusters that overlapped with any annotation, we gave them a *biotype* according to the overlapping annotation. Thus, we define the cluster biotypes as follows: yRNA if predicted by Rfam or our custom tool; miRNA if predicted by Rfam, MapMi, miRDeep2 or ShortStack; tRNA if predicted by Rfam or tRNAscan-SE; rRNA if predicted by Rfam or RNAmmer.

Expression quantification of sRNA-producing clusters

In order to obtain a level of expression for each sRNA-producing cluster, we counted mapped reads using the *findOverlaps* function from GenomicRanges R package (39) with parameters minoverlap=16 and ignore.strand=TRUE. With this we obtained count tables, where each cluster was a row and each library a column.

We also calculated coverage mapping to clusters using the *coverage* function from GenomicRanges R package using counts as weights.

Differential expression analysis of sRNA-producing clusters

Exploratory analysis of the expression levels of all clusters across libraries, led us to the conclusion that we had two distinct types of clusters: those producing mostly sRNAs with a 5' mono-phosphate (expressed in both types of libraries) and those producing mostly sRNAs with a 5' poly-phosphate (expressed mostly in the polyphosphatase-treated libraries). In order to consistently predict these two types of clusters, we performed differential expression analyses using the edgeR package (40). *C. elegans* libraries were analysed on their own, and since *H. bakeri* adult and EV libraries were quite distinct (MDS plots), they were analysed separately. We first imported the count tables, chose the desired libraries, and selected only clusters with at least 0.5 counts-per-million in at least 2 libraries. Next, we performed an edgeR analysis with TMM normalisation and using only the estimated common dispersion. To define the monoP-enriched clusters (with higher relative expression in untreated libraries, since untreated libraries do not contain 5' poly-phosphate reads) we performed a *glmTreat* test, searching for clusters that were significantly more abundant in the untreated libraries. We visually confirmed that the selected clusters contained the expected types of annotation (rRNA, tRNA, miRNA), and formed a consistently horizontal cloud in abundance vs fold-change (MA) plots. To define these monoP-enriched clusters, we used different cutoffs: for *C. elegans* a fold-change of 5 and False Discovery Rate (FDR) of 0.01, for *H. bakeri* adult libraries a fold-change of 2 and FDR of 0.01, for *H. bakeri* EV libraries a fold-change of 1.05 and FDR of 0.25. In all cases we calculated the FDR using the Benjamini-Hochberg method (41). Supplemental Figure 1 shows these results, highlighting the overlap with miRNAs. We next used the miRNA-containing monoP-enriched clusters as a baseline (since they should be equally expressed in untreated and treated libraries) to re-normalise the data. To define the polyP-enriched clusters, we again used a *glmTreat* test, searching for those with significantly more expression in the polyphosphatase-treated libraries. For the *C. elegans* and *H. bakeri* adult libraries we used a fold-change cutoff of 2 and 0.01 FDR, and for the EV libraries a fold-change of 1.5 and FDR of 0.2. The resulting polyP-enriched clusters are also indicated in Supplemental Figure 2. A full table with all the analysed clusters, genome coordinates, read counts for all libraries, the differential expression results and the overlapping annotation, is available as Supplemental Table 4.

Information Content of sRNA-producing clusters

To describe the pattern of reads covering each cluster, we calculated the coverage cluster Information Content (IC). IC is based on the coverage entropy of each cluster, compared to a uniform coverage distribution (maximum entropy) and is defined as *IC = log2(length(y)-entropy(y))*, where *y* is the cluster coverage. Entropy values were calculated with the *entropy* function from *entropy* R package [J. Hausser, S. Korbinian, entropy: Estimation of Entropy, Mutual Information and Related Quantities (2014)]. The total number of bases for each coverage value (discretized by the function y = *round(log10(coverage+1),1)*10*) were tabulated and converted into an entropy value. This was subtracted from the maximum entropy given the length of each cluster. For a perfectly uniform distribution (either no reads, or the same depth across the whole cluster, IC = 0. If IC > 0, there is more coverage variability, with the cluster presenting one or more "peaks". To avoid biases caused by higher sequencing depth of the Adult libraries, we used the same number of amplification cycles for both types of libraries, and during the bioinformatic analysis we randomly sampled exactly 2.5 million mapped reads from two Adult and two EV libraries, before calculating the IC values as mentioned above.

Argonaute Immunoprecipitations

H. bakeri adult worms were lysed with worms lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 0.5% NP40, 0.5 mM EDTA, cOmplete™ Protease Inhibitor Cocktail Tablets from Roche, pH 7) using mechanical disruption with 5 mm stainless steel beads (Qiagen) on a Tissue Lyser II (30hz for 2 min twice; Qiagen). The lysates were cleared by centrifugation (16,000×g) for 10 min at 4°C. The supernatants were immunoprecipitated with rat polyclonal anti-exWAGO antibody (raised against full length protein) or rat normal IgG followed by protein L magnetic beads (Fisher Scientific). After immunoprecipitation, equivalent amounts (from 200 uL) of the input, flow-through and the immunoprecipitated product were kept for RNA extraction using a miRNeasy Kit (Qiagen). Synthetic spiked-in of 0.1pM RT4 was added to the Qiazol before RNA extraction as internal control. The small RNAs that were associated with *H. bakeri* exWAGO were analysed by qRT-PCR. The rabbit polyclonal anti-exWAGO antibody (as described above) was used for the western blot analysis of exWAGO.

Detection of *H. bakeri* siRNAs by qRT-PCR

For reverse transcription of RNA from exWAGO IP, a fixed volume of 5 mL of total RNA was used as input for reverse transcription reactions using the miScript RT II System (Qiagen) according to the manufacturer's protocol. Quantitative PCR was carried out with the QuantiTec SYBR Green PCR kit (Qiagen), which includes a universal primer, according to the manufacturer's protocol. Primers for *H. bakeri* specific siRNAs and synthetic spike-in RNAs were used at a final concentration of 200 nM and were purchased from IDT. Two technical replicates were included, as well as a nuclease-free water ("no template") control. The q-RT-PCR condition used was as follows: 1) pre-denaturation for 15 min at 95^{o}C, 2) 40 cycles of denaturation 15s at 94^{o}C, annealing 30s at 55^{o}C, and elongation 30s at 70^{o}C. Fluorescence data collection was performed at the end of each annealing step. Data was collected on a Light Cycler 480 System (Roche).

List of qRT-PCR DNA primers used:

EV-enriched_nc16320	GATGACCAACCGGCTGTGGAAGC
EV-enriched_nc57384	GTAGTTGGGGTGGTTGTAGG
EV-enriched_nc23553	GAACGACTGCTTCTATGCCACCCGA
Adult-enriched_nc355572	GGAACTCCCAACGGGCCCGGG
Y-RNA-3p	CGACAAAAGCTCGACCGGCGC
miR-100	AACCCGTAGATCCGAACTTGTGT

References

1. Harcus Y*, et al.* (2009) C-type lectins from the nematode parasites Heligmosomoides polygyrus and Nippostrongylus brasiliensis. *Parasitology international* 58(4):461-470.
2. Claycomb JM*, et al.* (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* 139(1):123-134.
3. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1).
4. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
5. Heo Y, Wu XL, Chen D, Ma J, & Hwu WM (2014) BLESS: bloom filter-based error correction solution for high-throughput sequencing reads. *Bioinformatics* 30(10):1354-1362.
6. Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 18(5):821-829.

7. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4):357-359.
8. Kumar S, Jones M, Koutsovoulos G, Clarke M, & Blaxter M (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Frontiers in genetics* 4:237.
9. Kajitani R*, et al.* (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research* 24(8):1384-1395.
10. English AC*, et al.* (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS one* 7(11):e47768.
11. Haas BJ*, et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8(8):1494-1512.
12. Bray NL, Pimentel H, Melsted P, & Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34(5):525-527.
13. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, & Stanke M (2016) BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32(5):767-769.
14. Holt C & Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics* 12:491.
15. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology* 215(3):403-410.
16. Johnson LS, Eddy SR, & Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC bioinformatics* 11:431.
17. Stanke M & Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research* 33(Web Server issue):W465-467.
18. Katoh K, Misawa K, Kuma K, & Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30(14):3059-3066.
19. Guindon S, Lethiec F, Duroux P, & Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic acids research* 33(Web Server issue):W557-559.
20. Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* 52(5):696-704.
21. Emms DM & Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* 16:157.
22. Jones P*, et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236-1240.
23. Finn RD*, et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* 44(D1):D279-285.
24. Petersen TN, Brunak S, von Heijne G, & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 8(10):785-786.
25. Laetsch DR & Blaxter ML (2017) KinFin: Software for Taxon-Aware Analysis of Clustered Protein Sequences. *G3* 7(10):3349-3357.
26. Grishok A (2013) Biology and Mechanisms of Short RNAs in Caenorhabditis elegans. *Advances in genetics* 83:1-69.
27. Le Bihan T*, et al.* (2011) Shotgun proteomic analysis of the unicellular alga Ostreococcus tauri. *Journal of proteomics* 74(10):2060-2070.

28. Thery C, Amigorena S, Raposo G, & Clayton A (2006) Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Current protocols in cell biology* Chapter 3:Unit 3 22.
29. Nawrocki EP, Kolbe DL, & Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25(10):1335-1337.
30. Nawrocki EP*, et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic acids research* 43(Database issue):D130-137.
31. Lowe TM & Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research* 25(5):955-964.
32. Lagesen K*, et al.* (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* 35(9):3100-3108.
33. Kozomara A & Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* 42(Database issue):D68-73.
34. Guerra-Assuncao JA & Enright AJ (2010) MapMi: automated mapping of microRNA loci. *BMC bioinformatics* 11:133.
35. Buck AH*, et al.* (2014) Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nature communications* 5:5488.
36. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, & Enright AJ (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63(1):41-49.
37. Axtell MJ (2013) ShortStack: comprehensive annotation and quantification of small RNA genes. *Rna* 19(6):740-751.
38. Friedlander MR, Mackowiak SD, Li N, Chen W, & Rajewsky N (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* 40(1):37-52.
39. Lawrence M*, et al.* (2013) Software for computing and annotating genomic ranges. *PLoS computational biology* 9(8):e1003118.
40. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
41. Y BYH (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57:289-300.