SUPPLEMENTAL MATERIAL

Wang et al.

Deep Small RNA Sequencing from Nematode *Ascaris* Reveals Conservation, Functional Diversification, and Novel Developmental Profiles

Supplemental Materials and Methods

Ascaris material and RNA preparation. Eggs in the uterus near the vulva (proximal uterus, Zygote-4, Figure 1) of *Ascaris suum* (parasite of pigs) are fertilized (without pro-nuclear fusion) and have developed a mechanically and chemically resistant shell/membrane. These eggs can be isolated in large quantities from the proximal region of uteri (~45 ml of packed eggs/1000 worms) by treatment of dissected uteri with 0.5 N NaOH, and the eggs can be stored at 4°C for many years. Upon incubation at 30°C with high humidity and aeration, the eggs undergo slow and relatively synchronous development ($\geq$ 85%), enabling collection of embryos at various stages (see Figure 1). Early *Ascaris* development (Muller 1903) through the L2 is similar to that of *C. elegans* (Sulston et al. 1983), except that it occurs much slower (10 days to reach L1 in *Ascaris* vs. ~13 hr to L1 in C*. elegans*). Unlike *C. elegans* whose larvae are free-living and feeding, the *Ascaris* L1 and L2 larvae remain within the egg.

Ascaris has separate sexes and their adult male and female reproductive systems are on average ~140 and 240 cm, respectively. Discrete regions of the reproductive systems can be obtained (Figure 1). In males, we collected the germinal zone (Testis-G), two regions of the testis (Testis-1 and Testis-2), and spermatids from the seminal vesicle. In females, we collected the germinal region of the ovary (Ovary-G), the ovary (Ovary-All), and four regions of the uterus that contain the newly fertilized and shelled zygotes (Zygote 1-4) undergoing maturation necessary for pronuclear fusion and subsequent development (Figure 1) (Fairbairn 1957; Kaulenas and Fairbairn 1968).

For the germinal zone of the testis and ovary (Figure 1), the distal 5 cm of the gonad was isolated from live worms and stored at -80°C. Other regions of the gonads were dissected from live worms and either frozen immediately at -80°C or lysed by homogenization in Trizol and then frozen. *Ascaris* fertilized eggs were harvested from 4 regions of female uteri (Zygote 1-4) by treatment with 0.5 N NaOH, followed by extensive washing in water, and then stored at 4 °C with 5 volumes of PBS (pH 2.0). For embryonation, the fertilized eggs from final proximal region of the uterus (Zygote-4 = 0 hr embryos) were incubated at 30 °C in PBS (pH 2.0) with constant (100 rpm) shaking for the desired time. The outer shell/membranes were de-coated with 0.4 N KOH/1.4% sodium hypochlorite at 4 °C (zygotes from regions 1-4) or 30°C (developing embryos and larvae) for 180 or 90 minutes, respectively, prior to RNA

preparation. RNA from fertilized eggs in the uterus and early embryos (0 hr to 8 days) were extracted directly with Trizol (Invitrogen). Frozen L1, L2, testis and ovary samples were first ground with a mortar and pestle in liquid nitrogen and then extracted with Trizol.

Small RNA library preparation, sequencing and reads processing. Libraries were prepared and sequenced as previously described (Brennecke et al. 2007; Grimson et al. 2008) with some modifications. In brief, small RNAs (18–34 or 18-40 nt) were gel purified from 25 µg of total RNA using denaturing polyacrylamide–urea gels. Prior to gel purification of the RNAs, two small RNA spikes (p-AACUGUGUCUUUUCUGAAUAGA and p-UAUUUAGAAUGGCGCUGAUCUG) corresponding to mammal specific small RNAs were added to the samples for subsequent library read normalization. Periodate-treated libraries were generated as described (Vagin et al. 2006; Czech et al. 2009). For 5' all-phosphate libraries, 18-28 or 18-40 nt RNAs were treated with RNA 5' polyphosphatase (Epicentre, Madison, WI) to convert all tri- and dephosphorylated RNAs into 5' monophosphates prior to library construction. Libraries were sequenced on an Illumina GAII obtaining 5-25 million reads of 36-40 nt in length. Following removal of the adaptor sequences from the sequence reads, reads were collapsed to non-redundant datasets and then short reads (<= 17 bp) and low complexity reads (including poly A/C/G/ or T, di-nucleotide repeats, etc) were removed. We mapped all processed reads to *Ascaris* genomic and mRNA sequences (Table S1) using BLAST (E-value cutoff: 1e-2) and perfectly-matched reads (no gaps or mismatches across their entire length) were further analyzed (Table S2). Because the data are derived from wild *Ascaris* populations corresponding to hundreds of individuals, the sequences are highly polymorphic. Therefore, selection of only perfectly matching small RNAs is an extremely stringent criterion. Allowing one or two mismatches did not change the general picture of small RNAs identified.

Small RNA read normalization. In order to compare small RNA reads between different stages in 5' monophosphate libraries, we added two small RNA spikes at difference concentrations. Following sequencing, we observed an average of 255,531 high spike reads and an average of 7,429 low spike reads in the libraries. Using one of the 5' monophosphate libraries (Ovary-G) as a standard by setting its read number to 10 million, we compared all the spike reads in other libraries to the spike reads in the standardized Ovary-G. For each library, we normalized its read numbers by multiply the raw value by the

fold change to standard (Table S2), which was derived from the average $\log_2$ ratio fold changes of the two spikes in this library to spikes in the standardized library, and all normalized reads are rounded to integer (Table S2). While the fold change between spikes (spike1/spike2) in most stages (26 out of 28) is close to the average (34.4 fold), we observed an ~6.3X fluctuation in L2 and ~20 X in 8 day. As spike reads become lower, we expect larger fluctuation between spikes (the cases in 6 day, 7 day, 8 day, L1 and L2). Day 8 was significantly out of range from average and thus is not included in our comparative analysis of read frequencies between stages.

We normalized the read frequencies in the 5' all-phosphate libraries using miRNA reads. Since for 5' all-phosphate libraries on the same samples, we also prepared 5' monophosphate libraries, the miRNAs reads should be represented equally in each of the two pairs of libraries. For analysis of 0 - 64 cell and 128 cell – L2 5' all-phosphate libraries, their 5' monophosphate miRNAs data were derived from 0 hr - 116 hr 5' monophosphate libraries and 6 day - L2, respectively (Table S2). All reads in these 28 5' monphosphate and 19 5' all-phosphate libraries are comparable as they are normalized to our standard library (10 million 5' monophosphate reads in Ovary-G).

Expression profiles of miRNAs and siRNAs were clustered by using the Cluster 3.0 program (Eisen et al. 1998) (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm) followed by manual adjustment. The TreeView program (http://jtreeview.sourceforge.net/) was used to display the heatmaps.

MicroRNA identification and analysis. Following removal of small RNA sequence corresponding to *Ascaris* rRNA, tRNA, snRNA, snoRNA, and mitochondrial DNA, the remaining reads were clustered to identify potential miRNA hairpins (presence of neighboring matches within 50 nucleotides of each other with at least one of the matches $\geq$ 2,000 reads). Sequences of the clusters were then examined by mfold (Mathews et al. 1999). If the free energy of a potential hairpin was less than -20 kcal/mol, the most frequent reads located on one arm of a predicted hairpin, and the reads were involved in $\geq$ 15 base-pairings, the potential miRNA hairpin was further examined manually. miRNAs were defined using a combination of the following criteria: 1) RNAs derived from a unique and relatively stable genomic hairpin predicted by RNA folding, 2) the RNA sequence is within one arm of the genomic hairpin, 3) at least 15 nt of the RNA is involved in hairpin base-pairing, 4) presence of a RNA* reads in the other arm of the

hairpin, 5) RNA and RNA* have 2 nt overhangs, and 6) 5' terminal homogeneity of the small RNA. Additional low-frequency miRNAs were identified by sequence homology to miRBase and by searching 500 bp flank regions of established miRNAs to look for miRNAs in polycistronic loci.

We used miRBase (version 16.0) to identify *Ascaris* miRNAs orthologs. We first searched pairs of miRNAs that shared the seed region (2-8). For all such pairs, we aligned the mature miRNAs with EMBOSS:needle software (http://emboss.sourceforge.net/apps/cvs/emboss/apps/needle.html) that uses Needleman-Wunsch (NW) algorithm, with the default scoring matrix (match score 5, mismatch -4, gap penalty 10, and gap extension penalty 0.5). miRNA hairpin sequences were also aligned after the hairpin sequences were trimmed according to miRNA and miRNA* positions flanking 2 nucleotides. miRNA pairs with both high sequence similarities for their mature miRNAs and hairpins were considered as homologous. The cutoff scores for high similarities were defined as following: for each NW alignment, the ratio of the alignment score over the alignment length reflects how similar the sequences are. We used all miRNA pairs that do not share the seed sequences as the background, and their ratios were calculated and the median value (0.60 for mature miRNA alignment and 0.57 for hairpin alignment) was defined as the background. Seed-matched pair alignments 2-fold or more above background (1.20 for mature miRNA and 1.14 for hairpin) were considered having high sequence similarities and considered orthologous miRNAs.

Endo-siRNA identification and analysis. After removal of small RNAs corresponding to structural RNAs (rRNAs, tRNAs, snRNAs/snoRNAs, etc.) and miRNAs, the remaining genome- and mRNA-matched small RNAs were categorized into two types, siRNAs and Other small RNAs. The *siRNAs* are defined as small RNAs mapped to our *Ascaris* polyA-selected mRNA database and the *Other RNAs* are genome-matched RNAs that don't have mRNA matches. In order to distinguish sense and antisense small RNAs, we annotated our mRNA data (Table S1) and used them for the analysis. siRNA positions on the mRNAs were analyzed by using the 8,279 near-full length mRNAs (Table S1). We used a cutoff of 100 reads of small RNA/mRNA/million reads for each individual 5' all-phosphate library to define a curated set of 26G and 22G-targets (see text). A closer examination of 16 overlapping mRNAs targets of 26G and 22G-RNAs revealed that they have distinct 22G/22G ratios: 14 of them have a high 26G/22G

ratio, similar to the other 168 26G-RNA targets, while the other 2 overlap mRNAs have a high 22G/26G ratio, similar to the other 157 22G-RNA targets (Figure 6E and Table S7). We found the overlap is largely due to the "low" cutoff we have selected, and when lower ones were used, we see more overlap targets, but the ratio of 26G- and 22G RNAs remains constant (data not shown). We applied this constant 26G/22G ratio feature to refine the targets (see text).

21U/piRNA analysis. Several approaches were used to identify potential *Ascaris* orthologs of *C. elegans* 21U RNAs (piRNAs). First, we looked for *Ascaris* genomic regions enriched with 21nt RNAs starting with U (*Ascaris* 21U-RNAs) from the germline libraries. We used a cutoff of 21U-RNA % > 95% and reads per kb > 25 and identified only 92 *Ascaris* genomic loci that met these criteria. Using the same approach *on C. elegans*, we were able to identify 1,198 potential piRNA loci in the smaller *C. elegans* genome producing 4,749 unique small RNA reads with a total of 30,710 reads, and 94% of these small RNAs were defined as 21U-RNAs (Ruby et al. 2006; Batista et al. 2008). Thus, although our approach can identify 21U-RNAs clusters with high probability in *C. elegans*, few can be identified in *Ascaris*. *C. elegans* 21U-RNAs are derived from across an entire 21U genomic locus with each small RNA present at low frequency. In contrast, most of the potential 92 *Ascaris* 21U-RNA's loci are represented by only a few individual species at high-frequency, further suggesting they are unlikely to be orthologous to *C. elegans* 21U-RNA. Second, after we removed all RNAs types we could define, we looked for U-initiated RNA enrichment in all libraries. No U-intiated RNAs were identified (Figure 3 A-B, Fig. S1-4). Third, we used a position-specific scoring matrix based on a 21U-RNA motif defined in *C. elegans* (http://web.wi.mit.edu/bartel/pub/softwareDocs/TouRnaMotif/Introduction.html) to predict 21U-RNAs loci. Using moderate stringency (score cutoff, 15.5), we predicted 33,411 *C. elegans* potential 21U-RNAs of which 10,468 (31.3%) are true 21U-RNAs, representing 67% (total read number) of all known 21U-RNAs. In contrast, only 8,746 potential 21U-RNAs were predicted for the *Ascaris* genome and only 32 of these were present in our small RNA libraries. The 21U-RNA motif among different nematodes is diverse in regions outside the core GTTTC motif (de Wit et al. 2009). To check whether motifs from other nematodes can be used to predicted 21U-RNA in *Ascaris*, we used the matrix based on the diverged 21U-RNA motif derived from *Pristionchus pacificus* (de Wit et al. 2009). We were able to predict 88%

(988 of the 1,123) of the known 21U-RNAs in *P. pacificus* (score cutoff, 20). While for *Ascaris*, only 96 out of 17,944 potential 21U-RNAs were found in our libraries. Last, in an effort to enrich for piRNAs, we prepared periodate treated small RNA libraries to enrich for small RNAs with 3'-modifications, but did not identify any 21U RNAs enriched in any of the libraries.

Genomic DNA library preparation, sequencing and assembly. The *Ascaris* genome is 290 ± 30 and 394 ± 40 Mb (somatic and germline, respectively) based on our sequencing and measurement of DNA content of nuclei (fuelgen staining). The *A. suum* genome was sequenced by a whole-genome shotgun strategy using three sequencing platforms: paired-end Sanger sequencing, direct 454 titanium sequencing and Illumina paired-end sequencing (Table S1). For Sanger sequencing, DNA was size fractioned to enrich for ~5.5 kb fragments as previously described (Abubucker et al. 2008). The Sanger sequencing generated 451,630 reads with an average length of 826 bases. For 454 Titanium sequencing, 6.6 µg of genomic DNA was fragmented via nebulization, DNA fragments of 500-800 bp in size were isolated, and their ends were polished and ligated to 454 Titanium library adaptors using reagents from the Titanium General Library Kit (454 Roche). The DNA library was sequenced on the Genome Sequencer instrument using the GS FLX Titanium Sequencing Kit XLR70 (Roche 454). A total of 2.66 million reads were generated with an average length of 350 bases (see Table S1). For Illumina sequencing, genomic DNA was fragmented using a Bioruptor sonicator (Diagenode), Solexa paired-end adaptors were added to the DNA as described by Illumina, DNA fragments with an insertion of 300-400 bp in length were gel purified, and then PCR amplified. Forty-one million paired-end reads with an average length of 100 nt were generated on the Illumina GAII.

Genomic reads from Sanger and 454 platforms were combined and assembled by using Newbler (Roche), and genomic reads from Illumina were assembled with Velvet (Zerbino and Birney 2008). We combined the two assemblies by using Phrap (http://www.phrap.org/), and all the assemblies were evaluated by using Tablet assembly viewer (http://bioinf.scri.ac.uk/tablet/). Scaffolds were built by using the paired-end Sanger reads and Illumina reads. The final assembly consists of 86,678 scaffolds with an N50 of 11,773 bp (the largest scaffold is of 185,246 bp) and ~273 Mb of assembled sequence (Table S1). This assembly was used to define the genome origin of small RNAs and their potential biogenesis.

Ascaris rRNA sequences were identified using database sequences and tRNA sequences were identified by using tRNAscan-SE program (Lowe and Eddy 1997).

cDNA library preparation, sequencing, assembly and annotation. cDNA libraries were prepared from polyadenylated RNA (2x oligo-dT purified using μMACS™ mRNA Isolation Kit, Miltenyi Biotec). Two types of libraries were prepared: 1) sheared, full-length cDNA synthesized using a combination of oligo-dT and random hexamer priming and 2) cDNA prepared from RNA first chemically sheared and then double-stranded cDNA prepared using ramom hexamer priming. For the chemically sheared library, RNA was treated for 75-150 sec with RNA fragmentation buffer (Ambion) at 70°C to generate RNA fragments 150-400 nt in length. First stand cDNA was prepared by random priming (3-6 μg random hexamers/1 μg of A+ RNA) and second stranded cDNA synthesis was carried out using a SuperScript® Double-Stranded cDNA Synthesis Kit (Invitrogen). Solexa paired-end adaptors were added to the blunt cDNA as described by Illumina, cDNA fragments of 300-400 bp in length were gel purified, the fragments amplified using 17 PCR cycles, and paired-end, 76-120 nt reads were generated as described on the Illumina GAII. All the libraries we have sequenced are listed in Table S1. Six libraries are chemically sheared first; these are Embryo (pooled RNA from early embryos development (0, 24, 46, 64, 96, and 116 hr embryos)), Larvae (larval development and larvae (6, 7, and 8 day embryos and L1 and L2 larvae)), Testis (testis germimal zone), Ovary (ovary germinal zone), complete female reproductive system, and whole males. One library is full-length sheared cDNAs from mixed samples of all of the RNAs in other six libraries.

Reads from each cDNA library were initially assembled with Velvet (Zerbino and Birney 2008). We built a set of reference sequences by combining all these assemblies together. Using these reference sequences, we assembled each library again and an improved assembly for each library was obtained. We then built an improved reference sequences and iterated this cycle until no further improvement was achieved. All the assemblies were checked with Tablet (http://bioinf.scri.ac.uk/tablet/) by using a cutoff of 5% mismatch for mis-assembled contigs. Other obvious chimera contigs (after annotation, see below) were split. The final assembly (Table S1) was improved by scaffolding, genomic sequences, and available ESTs.

We used the UniProt (http://www.uniprot.org/) and WormBase (http://www.wormbase.org/) to annotate the cDNAs by using BlastX. cDNAs with database matches with an E-value <= 1e-10 were annotated. cDNAs were further categorized using Pfam 24.0 (http://pfam.janelia.org/). The strand of cDNAs were identified using BlastX annotation, the presence of a polyA tail and/or splice leader sequence (Nilsen et al. 1989).

Supplemental Results and Discussion

***Ascaris* miRNA gene organization and expression, processing, editing, and nucleotide addition.** We identified 13 putative polycistronic miRNA loci encoding 33 miRNAs (see Figure S7). miRNAs derived from polycistronic loci are predominantly expressed before day 8 of development and often simultaneously express miRNAs with the same seed sequence. RT-PCR analysis demonstrated that most of the predicted polycistronic loci generate polycistronic primary transcripts, and a number of the primary transcripts are spliced leader trans-spliced and polyadenylated (data not shown). For most of these loci, the individual miRNAs from a single locus are not equally expressed (see Table S3 and Figure S7). While it remains possible that some of these miRNAs have their own promoters, **the data suggest they are either differentially processed or have different half-lives or both.**

No *Ascaris* miRNAs were identified as mirtrons (Ruby et al. 2007) whereas C. elegans has four. *Ascaris* miR-100a-2 is an offset RNA, as recently described in tunicates (see Figure S5) (Shi et al. 2009). Variation in the processing of miRNAs was also observed. As illustrated in Figure 4, the asu-miR-5347 hairpin produces two different sizes of mature miRNAs, 21 and 23 nt. The 23 nt miRNA is maternally contributed and the 21 nt miRNA is produced from the 96 hr – L2 stages of development. Independent miRNAs are derived from the same miRNA hairpin precursor as illustrated by the significant expression of several *Ascaris* miRNA*s (Table S3). miRNA* levels are typically observed at very low frequencies compared to the miRNA. This is observed for most of miRNA* detected in the libraries (on average ~1% of the miRNA reads). However, we observed that 16 *Ascaris* mRNA* reads (16.5% of the miRNA*) accumulate to 15% or more of the mature miRNA reads (Table S3). Several recent and comprehensive studies have made similar observations in flies, including evidence that the miRNA* is functional (Okamura et al. 2008; Czech et al. 2009; Okamura et al. 2009). Interestingly, miRNA strand selection (Liu et al. 2008; Okamura et al. 2008; de Wit et al. 2009) also appears to change in the evolution of the *Ascaris* miRNA locus. *Ascaris* miR-79 and miR-9 hairpins are paralogs having apparently been derived from duplication, yet the mature miRNAs produced from the two hairpins are derived from opposite sides of the hairpin. Thus, the mature miR-79 miRNA is 3p whereas the mature miR-9 is 5p (Figure S8).

Five *Ascaris* miRNAs undergo miRNA editing, four of which have A->I (G) editing, one exhibits C->U editing, and two are edited within their seed sequences (asu-miR-100d and asu-miR-5354) (Table S6). miRNA 3'-uridylation and 3'-adenylation (~3% of the reads) were observed for most of the *Ascaris* miRNAs. Eleven miRNAs exhibited significant levels of 3'-addition (Table S6) and a combination of polyuridylation and polyadenylation was often observed on the same miRNA. When truncated miRNAs are observed (asu-miR-1/2a/43e/56/750/5347), they exhibit much higher levels of 3' addition than the full-length miRNAs. This may correlate with marking the truncated miRNA for degradation (Chatterjee and Grosshans 2009).

***Ascaris* small RNA protein machinery.** *C. elegans* has a diverse set of Argonaute proteins including 5 Ago proteins, 3 PIWI Argonautes, and 18 worm specific Argonautes (WAGOs) (Yigit et al. 2006). A search for Argonaute-related proteins in the *Ascaris* genomic and transcriptome sequence led to the identification of 5 general Ago-clade proteins (2 corresponding to *C. elegans* Alg-1/Alg-2 and Alg-3/Alg-4), 5 WAGOs (corresponding to *C. elegans* csr-1, Wago-1, and nrde-3; the other two *Ascaris* WAGOs cannot be assigned to a *C. elegans* ortholog), but no PIWI orthologs (see Figure 3B and Table S9). Ago-, PIWI-, and WAGO-clade Argonautes associate with miRNAs and 26G-RNAs, piRNAs, and 22G-RNAs, respectively. *Ascaris* orthologs of pasha, drosha, dicer, three RNA-dependent RNA polymerases, and orthologs of many key proteins in *C. elegans* RNA interference pathways were also identified (eri-1, dcr-1, cde-1, mut-7, drh-3, ekl-1, ain-1, nlh-2, and nrde-2).

Ascaris small RNAs and transposons. Only 1-2% of all *Ascaris* small RNAs correspond to mobile elements. Over 90% of these correspond to LT and non-LTR retrotransposons. Small RNAs corresponding to these elements are expressed at relatively high frequency compared to other non-miRNA small RNAs (see Table S8).

References

Abubucker S, Martin J, Yin Y, Fulton L, Yang SP, Hallsworth-Pepin K, Johnston JS, Hawdon J, McCarter JP, Wilson RK et al. 2008. The canine hookworm genome: analysis and classification of Ancylostoma caninum survey sequences. *Mol Biochem Parasitol* **157**(2): 187-192.

Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S et al. 2008. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell* **31**(1): 67-78.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**(6): 1089-1103.

Chatterjee S, Grosshans H. 2009. Active turnover modulates mature microRNA activity in Caenorhabditis elegans. *Nature* **461**(7263): 546-549.

Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ. 2009. Hierarchical Rules for Argonaute Loading in Drosophila. **36**(3): 445-456.

de Wit E, Linsen SE, Cuppen E, Berezikov E. 2009. Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res* **19**(11): 2064-2074.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**(25): 14863-14868.

Fairbairn D. 1957. The biochemistry of Ascaris. *Exp Parasitol* **6**(5): 491-554.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**(7217): 1193-1197.

Kaulenas MS, Fairbairn D. 1968. RNA metabolism of fertilized Ascaris lumbricoides eggs during uterine development. *Exp Cell Res* **52**(1): 233-251.

Liu N, Okamura K, Tyler DM, Phillips MD, Chung WJ, Lai EC. 2008. The evolution and functional diversification of animal microRNA genes. *Cell Res* **18**(10): 985-996.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**(5): 955-964.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**(5): 911-940.

Muller H. 1903. Beitrag zur Embryonalentwickelung der Ascaris megalocephala. . *Zoologica* **41**: 1-30.

Nilsen TW, Shambaugh J, Denker J, Chubb G, Faser C, Putnam L, Bennett K. 1989. Characterization and expression of a spliced leader RNA in the parasitic nematode Ascaris lumbricoides var. suum. *Mol Cell Biol* **9**(8): 3543-3547.

Okamura K, Liu N, Lai EC. 2009. Distinct Mechanisms for MicroRNA Strand Selection by Drosophila Argonautes. **36**(3): 431-444.

Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC. 2008. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* **15**(4): 354-363.

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127**(6): 1193-1207.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**(7149): 83-86.

Shi W, Hendrix D, Levine M, Haley B. 2009. A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nat Struct Mol Biol* **16**(2): 183-189.

Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* **100**(1): 64-119.

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**(5785): 320-324.

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC. 2006. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**(4): 747-757.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**(5): 821-829.