De novo transcriptome assembly of a facultative parasitic nematode Pelodera (syn. Rhabditis) strongyloides

Menglei Zhang, Liisa Heikkinen, K. Emily Knott, Garry Wong

PII:	S0378-1119(19)30508-6
DOI:	https://doi.org/10.1016/j.gene.2019.05.041
Reference:	GENE 43890
To appear in:	*Gene*
Received date:	15 June 2018
Revised date:	1 April 2019
Accepted date:	21 May 2019

Please cite this article as: M. Zhang, L. Heikkinen, K.E. Knott, et al., De novo transcriptome assembly of a facultative parasitic nematode Pelodera (syn. Rhabditis) strongyloides, Gene, https://doi.org/10.1016/j.gene.2019.05.041

De novo transcriptome assembly of a facultative parasitic nematode *Pelodera (syn. Rhabditis) strongyloides*

Menglei ZHANG[1], Liisa Heikkinen[2], K. Emily Knott[2], Garry WONG[1*]

1 Centre of Reproduction, Development and Aging, Faculty of Health Sciences, University of Macau, Macau SAR, China

2 University of Jyväskylä, Finland

*Correspondence:

Garry WONG

Faculty of Health Sciences, University of Macau, Macau SAR, China

Tel: +853 88224979; Fax: +853 88222314

GarryGWong@umac.mo

Abstract

Pelodera strongyloides is a generally free-living gonochoristic facultative nematode. The whole genomic sequence of *P. strongyloides* remains unknown but 4 small subunit ribosomal RNA gene (ssrRNA) sequences are available. This project launched a *de novo* transcriptome assembly with 100 bp paired-end RNA-seq reads from normal, starved and wet-plate cultured animals. Trinity assembly tool generated 104,634 transcript contigs with N50 contig being 2,195 bp and average contig length at 1,103 bp. Transcriptome BLASTX matching results of five nematodes (*C. elegans*, *Strongyloides stercoralis*, *Necator americanus*, *Trichuris trichiura*, and *Pristionchus pacificus*) were consistent with their evolutionary relationships. Sixteen genes were identified to be homologous to key elements of the *C.elegans* RNA interference system, such as Dicer, Argonaute, RNA-dependent RNA polymerase and double strand RNA transport proteins. In starved samples, we observed up-regulation of cuticle related genes and 3 dauer formation genes. Dauer morphology was captured with enlarged phasmid under light microscopy, and dauer and normal larvae counts in clumps had a Pearson's product-moment correlation of 0.805 with p-value = 0.0088. Our results demonstrate that *P. strongyloides* could be used for studying nematode-related human or pet parasitic diseases. The sequenced assembled transcriptome reported here may be useful to understand the evolution of parasitism in Nematoda.

Key words: Dauer, facultative parasite, *Pelodera strongyloides*, starvation, transcriptome assembly

Abbreviations: DE, Differential expression; dsRNA, double strand RNA; FDR, false discovery rate; FPKM, Fragments Per Kilobase Million; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; NGM, Nematode Growth Medium; N, Normal; PE, paired ends; S, Starved; RdRP, RNA-dependent RNA polymerase; RSEM, RNA-Seq by Expectation-Maximization; TPM, Transcripts Per Kilobase Million; W, Wet-plate.

1. Introduction

Pelodera (syn. Rhabditis) strongyloides (Scheider, 1860) is a gonochoristic nematode with separate sexes, generally free-living in decaying organic material or wet soil [1]. It is located in clade 9 close to the free-living nematode *C. elegans* and parasitic nematode *Pristionchus pacificus* (**Figure 1A**) [2]. The esophagus of this species is typically rhabditoid shaped with a corpus, isthmus, and bulb, which can be readily observed under light microscopy [3]. It has a life cycle of 3.5 days under laboratory conditions, (*E. coli* agar medium, 22 °C lab room temperature) [4], and the whole life cycle contains 1 adult stage and 4 larval stages (L1-L4). Mature female adults can produce ~500 eggs in one lifetime [5]. The female adults' mean body length is approximately 992 µm, and the body length of male adults is about 885 µm [6]. Although *P. strongyloides* is primarily a free-living animal, it can occasionally invade a host and then exit several days later, which is defined as a directly transmitted parasite [7] or a facultative parasite [8]. The dauer-stage larvae [9] and the third-stage larvae of the dermatitica strain [3] have the ability to invade in hair follicles as well as urine and lacrimal fluid by penetration. The wild post-parasitic larvae (removed from hosts) can molt to the adult stage and return to a free-living condition [9]. Currently, about 50 *P. strongyloides* infection cases have been reported in terrestrial and marine mammals, such as dogs, cats, voles, sheep, horses, black bears, dairy herd, harbor seals and humans [1, 9-16].

Despite many cases reported, the study of this worm are mainly focused on morphology description. Except for 4 sequences published in National Center for Biotechnology Information (NCBI) for evolutionary purposes [17, 18], no more publicly available genetic information exists for this species. Since transcriptome assembly is now possible for species lacking a sequenced genome [19], we aimed to produce a *de novo* transcriptome assembly for this species. This study should provide a genomic resource for the facultative parasite *P. strongyloides* and insights into its evolution and molecular biology that could be exploited to improve human and mammal pet health.

2. Material and methods

2.1 Sample preparation

P. strongyloides (Strain DF5013 obtained from Caenorhabditis Genetics Center (CGC)) was maintained on Nematode Growth Medium (NGM) seeded with NA22 [20]. Control/normal samples were collected after a week of cultivation on NGM plates. Starved samples were collected after 12 days cultivation on NGM plates when all food was consumed and worms started to appear on the plate lid. Wet-plate samples were collected from animals cultured on wet NGM plates that were maintained in a thin layer of liquid for 10 days in a humid box at 22 °C [21]. Samples were mixed cultures and included all 4 larval stages and adults of both genders. All samples were harvested with M9 and homogenized in Trizol (Invitrogen Life Technologies) and then stored in -80 °C.

2.2 RNA isolation, cDNA preparation and RNA-seq

Total RNA was isolated from the frozen normal (control), starved, and wet-plate samples using Trizol following the manufacturer’s protocol. Each treatment had 2 biological replicates. RNA concentration was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies) [22]. The cDNA libraries were prepared with NEBNext Poly(A) mRNA Magnetic Isolation Module for Illumina according to the instructions provided by the manufacturer (New England Biolabs). The libraries were sequenced using Illumina HiSeq2500 in the local Genomics & Bioinformatics Core of the Faculty of Health Sciences, University of Macau. The output reads were set to be 100 base pair (bp) long paired ends (PE). RNA sequencing data has been deposited in the NCBI database under BioProject: **PRJNA408007**, accessible with the following link: http://www.ncbi.nlm.nih.gov/bioproject/408007.

2.3 Quality control and De novo transcriptome assembly

Clean reads were achieved from raw paired-end reads of high quality (Q >30) after adaptor sequence trimming, and *E. coli* reads were removed by aligning them to the *E. coli* genome using bowtie2 (Burrows-Wheeler transform indexing) [23] with parameters “bowtie2 -p 1 --un-conc-gz”. All 6 libraries of clean reads were pooled into a single left-end FASTA file together with a single right-end FASTA file. The *de novo* transcriptome assembly was performed using two methods. Trinity-2.2.0 was used with optimized 25 kmer and the de Bruijn graph algorithm [19], and SOAPdenovo-Trans-bin-v1.03 was used with the same kmer length, but using the de Bruijn graph algorithm with an effective scaffolding module [24]. The assembly resulting in longer average contig was considered to be the primary assembly and used in the down-stream analysis. Further justification of this decision was supported by global length comparison based on *C. elegans* mRNAs (Supplementary Figure 1). Redundancy Filtering of the primary assembly was done with cd-hit-v4.6.1, using greedy incremental clustering algorithm [25] with the parameter “-c 1”, which meant that only 100% matching redundancies were removed, and the remnant was called transcriptome. Assembly quality was assessed by RNA-Seq read representation examined by aligning paired-end clean reads to the transcriptome using a perl script “bowtie_PE_separate_then_join.pl” provided by Trinity, and the alignment statistics were obtained by running another Trinity script “SAM_nameSorted_to_uniq_count_stats.pl” [26]. Transcriptome completeness was measured using BUSCO v2 based on evolutionarily-informed expectations of gene content from near-universal single-copy orthologs provided by OrthoDB v9 [27].

2.4 Functional annotation

The transcripts were examined by aligning the assembly to the reference proteins of *C. elegans* (free-living, clade 9A), *Strongyloides stercoralis* (parasitic, clade 10B), *Necator americanus* (parasitic, clade 9B), *Trichuris trichiura* (parasitic, clade 2A) and *Pristionchus pacificus* (parasitic, clade 9A) (Figure 1) [2] in Wormbase (Version: WS256) using BLASTX with a 10^{-10}

expectation value. The candidate coding regions within transcript sequences were identified using TransDecoder [26] based on minimum length open reading frame and log-likelihood score. The length coverage of predicted peptides was examined using BLASP with the same reference proteins mentioned above. Peptide functional annotation was performed using Trinotate [26]. Various functional annotations were combined together into a SQLite database, including homology search to known sequence data (BLAST+/SwissProt) [28], protein domain identification (HMMER/PFAM) [29], protein signal peptide and transmembrane domain prediction (signalP/tmHMM) [30], and leveraging various annotation databases (eggNOG/GO/KEGG databases). The Gene ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis was conducted by R package clusterProfiler 3.8.0.

2.5 Differential expression (DE) analysis

Transcript abundance of counts, FPKM (Fragments Per Kilobase Million), and TPM (Transcripts Per Kilobase Million) were estimated by aligning clean reads of 3 treatments with 2 biological replicates separately back to the transcriptome using Trinity toolkit align_and_estimate_abundance.pl choosing RSEM (RNA-Seq by Expectation-Maximization) method [31]. Six sets of RSEM gene results were combined into a matrix by abundance_estimates_to_matrix.pl script. Differentially expressed genes were identified using the EdgeR Bioconductor package and based on the Empirical Bayes moderated overdispersed Poisson model [32] using run_DE_analysis.pl provided by Trinity.

The DE result was normalized using run_TMM_normalization_write_FPKM_matrix.pl with N_rep1 length as reference. Differential expression analyses were conducted using analyze_diff_expr.pl with parameter "--samples samples_described.txt -C 2 -P 0.001" to extract transcripts that were at least 2^2 fold differentially expressed with false discovery rate (FDR) of at most 1e-3. The genes were then clustered according to their patterns of differential expression across the samples. The differentially expressed genes were annotated based on homology to *C.*

elegans proteins in Wormbase (Version: WS256). GO term enrichment was conducted by clusterProfiler.

2.6 Dauer morphology examination and count statistics

Animal samples for examination of dauer stage were obtained from starved plates. Plates which had clumped animals [33] on the lid, lid edge or on the agar were checked with a platinum wire under bright field light microscopy and photographs were obtained using a Carl Zeiss Axio Observer Z1 camera. Worms were treated with sodium azide (5mM) diluted by M9 from 100mM stock provided by Sigma-Aldrich) and then examined carefully under 40X magnification. The dauer stage was confirmed by presence of an enlarged phasmid in the tail. Those without enlarged phasmid were treated as normal larvae. The number of dauer and normal animal were counted, of which basic statistics and correlation were performed using R 3.4.4. For relative proportion statistics, different developmental stages (adult, larvae, dauer) of each treatment (normal, starved, wet) were calculated from worms grown in 6 cm plates. Worms were washed off with M9 using glass tips. After centrifugation, buffer was removed and volume was brought up to 300 µl with M9. A 2 µl thoroughly mixed sample was taken from each tube and placed into 10 µl sodium azide (5mM) with primed tips (with M9) [34]. Each treatment had 4 plates as replicates and each replicate was randomly sampled 3 times for recording adult, larvae, dauer numbers. Mean value of relative proportions of each replicate were used to draw the error bar plot with R 3.4.4.

3. Results

3.1 Transcriptome assembly

The total number of paired-end (PE) reads used for assembly was 285.0 M obtained from 6 sample libraries: Normal 1 (N1, 37.0 M reads), Normal 2 (N2, 46.2M reads), Starved 1 (S1, 88.3M reads), Starved 2 (S2, 40.5M reads), Wet-plate 1 (W1, 38.6M reads), Wet-plate 2 (W2, 34.4M reads). Read quality (Q>30) were checked by FastQC [35].

Trinity generated 104,962 transcript contigs with N50 of 2,198 bp and average length of 1,104 bp. Correspondingly, SOAPdenovo produced 178,118 transcript contigs, with N50 of 476 bp and average length of 319 bp. The Trinity assembly was chosen for further analysis because of its longer average length, which was supported by the length distribution of SOAPdenovo, Trinity output when compared to the corresponding *C. elegans* mRNA (**Supplementary Figure 1**). After removing redundancy, the remnant contained 104,634 transcript contigs with contig N50 of 2,195 bp and average contig length of 1,103 bp, which was employed as the transcriptome in downstream analysis (**Table 1**). The length distribution of the transcriptome is shown in **Figure 2**. From the analysis, 29.6% of transcriptome was located in the 300 bp bin, 33.4% of the transcriptome was longer than 1000 bp, and the number of transcripts longer than 5000 bp was 2411. The percentage of RNA-seq paired-end reads yielding concordant alignments at least 1 time to the reconstructed transcriptome was over 81% for all the libraries (81.51% (N1), 81.58% (N2), 82.03% (S1), 84.18% (S2), 85.39% (W1) and 84.36% (W2)). Furthermore, the completeness of the transcriptome based on conserved gene contents (protein sequences) of nematoda_odb9 (982 genes), metazoa_odb9 (843 genes) and eukaryota_odb9 (429 genes) in BUSCO annotation were 89%, 83% and 95% respectively.

3.2 Functional annotation

The translated *P. strongyloides* transcriptome was compared to protein sequences from 5 nematodes. There were 28,939 translated transcripts similar to 10,313 *C. elegans* proteins, 26,587 transcripts similar to 7,421 *N. americanus* proteins, 26,445 transcripts similar to 7,659 *P. pacificus* proteins, 24,858 transcripts similar to 6,967 *S. stercoralis* proteins and 14,494 transcripts similar to 3,512 *T. trichiura* proteins. Distribution of *P. strongyloides* transcripts amongst proteins of the above nematodes is illustrated in a Venn diagram (**Figure 1B**). Of these, 12,798 transcripts were commonly similar to all 5 reference nematodes. The numbers of transcripts specifically aligned to only *C. elegans*, *P. pacificus*, *S. ratti*, *N. americanus* and *T. trichiura* were 1,453, 843, 197, 161 and 77, respectively.

Because RNA interference is an important process to regulate gene expression [36], we looked for pathway components in *P. strogyloides*. According to *C. elegans* BLASTX result, the assembled transcriptome contained translated protein sequences that were matched to miRNA and siRNA pathway genes from *C.elegans*, of which 16 were identified to be homologous to 9 proteins from the *C.elegans* RNAi (RNA interference) system and 22 genes were homologous to *C. elegans* phenotype genes (**Table 2**).

The GO term enrichment of *P. strongyloides* transcriptome was analyzed based on homology to *C. elegans* proteins. There were 961 GO terms enriched with P-value < 0.01 and q-value < 0.05, and 53 KEGG pathway terms with P-value < 0.05 were mapped (**Figure 3**).

Furthermore, based on the peptide functional annotation, Transdecoder predicted 39,073 CDS contained in the transcriptome. Except for those transcripts annotated by *C. elegans* BLASTX results, there were 20,,268 more transcript remnants annotated based on the homologous predicted peptide to eggNOG, GO, and KEGG databases via Trinotate. The top 10 biological process, cellular component and molecular function GO terms with most transcript counts were plotted (**Figure 4**). The KEGG pathways and eggNOG orthologous protein groups with top 15 most transcript counts are shown in **Figure 4**.

3.3 Differential expression analysis

The relationship of differentially expressed genes (DE genes) among *P. strongyloides* samples is shown in **Figure 5**. Rows (genes) and columns (samples) were hierarchically clustered based on the gene median-centered log2 transformed expression values (FPKM), which was represented in color ranging from purple (down-regulated with negative value) to yellow (up-regulated with positive value) (**Figure 5A**). Spearman correlation for each pair of samples according to the difference of the transcript expression values (TMM-normalized FPKM) varied from 0.6 (green) to 1 (red) (**Figure 5B**). Replicates of each treatment were clustered together, starved and wet-plate samples were more similar to each other than to normal samples both in DE gene level (**Figure 5B**) and in all transcriptome level (**Figure 5C**).

Compared to normal samples (N), wet-plate samples (W) up-regulated 136 genes (**Supplementary Table 1**) and down-regulated 655 genes (**Supplementary Table 2**), while starved samples (S) had 536 up-regulated genes (**Supplementary Table 3**) and 962 down-regulated genes (**Supplementary Table 4**). Among the starvation up-regulated genes, three of them are closely related to dauer formation (**Table 3**). Furthermore, those up-regulated genes were enriched in 3 categories: "collagen trimer", "structural constituent of cuticle" and "structural molecule activity" (**Figure 6A**). No enriched term was found when analyzing the wet-plate up-regulated genes. Wet-plate and starved down-regulated genes were enriched in 28 and 56 terms respectively (p-value < 0.01, q-value < 0.05). The top 10 GO terms of starved and wet-plate samples are shown in **Figure 6B** (S) and **Figure 6C** (W).

3.4 Dauer morphology examination and count statistics

Dauer animals were only found in newly hatched larvae clumps from starvation plates. The enlarged phasmid of dauer animal can be seen in **Figure 7A**. Dauer number and normal larvae number were recorded for one clump per plate. The distribution of each type of larvae (dauer, normal) and their correlation are illustrated in **Figure 7B**. According to the mean value, there were about 11 dauer and 20 normal larvae in each clump. And the Pearson's product-moment correlation of dauer and normal larvae number was 0.805 with p-value = 0.0088. Mean relative proportions (adult, larvae, dauer) in each experimental condition are shown in **Figure 7C**.

4. Discussion

The majority of *P. strongyloides* literature are descriptions of morphology and life history, and only one sequence paper has been published to our knowledge with 4 ssrRNA sequences for phylogenetic purposes [18]. Our study provides transcriptome sequence information of *P. strongyloides* using 285 million 100bp PE reads with coverage >100× based upon 20,000 genes of 1,500 bp size estimation. In the human transcriptome project, 200 million PE reads meets the minimum coverage required [37]. Therefore, we are confident that we have sufficient sequencing coverage for this nematode. We found that the median length of transcript contigs is 489 bp (**Table 1**), which is much shorter than 1,956 bp, the median size of *C. elegans* coding genes. The difference in median length could potentially be due to presence of introns and lack of 3' noncoding regions used to calculate the median size of the C. elegans gene [38]. There are 49,207transcripts (47% of the transcriptome) annotated together by *C. elegans* sequences and unique Trinotate peptide function (**Figure 4**), which is 10,000 more than the predicted CDS number, suggesting that gene annotation is sufficient at the transcriptome level.

The clock hypothesis predicts that unit base substitution in DNA/RNA sequences is proportional to evolutionary time [39]. The more related reference species share more specific genes [40]. According to multi-nematode sequence based BLASTX results, the number of transcripts annotated is coherent with their evolutional relationship, moreover, it also suggests that *P. strongyloides* is more close to *C. elegans* than *P. pacificus*. Moreover, the number of unique genes annotated indicates similar results except that the unique annotated genes of another 9 clade nematode *N. americanus* are less than the 10B clade *S. ratti*.

RNA-seq libraries were constructed from RNA isolated from a well-populated mixture, so the GO annotation of transcriptome includes the biological process of embryo development, larval development, sex differentiation and dauer entry. The KEGG pathways with top 15 least P-value are essential pathways concerning metabolism, nucleic acid synthesis and degradation, and mass transportation (Figure 3).

Since RNAi gene silencing techniques were established in *C. elegans* by Fire et al. (1998) [36], it has been a powerful method in gene function studies. To conduct the RNAi experiment, there are several essential elements should be available: Dicer, Argonaute, RNA-dependent RNA polymerase (RdRP) and double strand RNA (dsRNA) transport proteins [41, 42]. The table of RNAi genes illustrates that *P. strongyloides* has all those mentioned RNAi key elements (**Table 2**). Furthermore, there are 4 genes found in *P. strongyloides* homologous to genes forming phenotypes to validate the success of RNAi in *C. elegans*, including animals with no sperm, lumpy-dumpy larvae, strong twitchers and paralysed behaviors [36]. It is thus very likely that employing RNAi methods to investigate the function of potential "parasitic gene" or other genes in *P. strongyloides* is viable.

Environmental changes can provoke an organism to respond through changes in gene expression [43]. Organisms from the same treatment had similar gene expression patterns (**Figure 5**). The cuticle structure and its collagen components are conserved throughout the nematode phylum [44]. Moreover, cuticle is one of the main self-protecting shields of parasitic nematodes in evading host defenses [45, 46]. Starved up-regulated genes are enriched in cuticle related terms (**Figure 6A**) which provides a clue of parasitism, and dauer is considered to be an infective stage of this facultative parasite [9]. Unlike *C.elegans*, whose body normally straightens to a rod shape following sodium azide exposure, *P. strongyloides* worm body showed a strong shrinkage. All normal larva shrank dramatically, while dauer and adults displayed better tolerance to sodium azide exposure (**Figure 7B**). There was a high correlation between number of normal and dauer larvae in clumps in the starved plates (**Figure 7C**). This suggests that these animals have mixed dauer/non-dauer populations which may be advantageous during infection. Alternatively, starvation may only partially induce dauer formation and other more severe treatments (e.g. dessication, overcrowding) may be necessary to induce full dauer populations. The top 10 significant GO terms of starved down-regulated genes demonstrated that worms decrease the pace of development. Wet-plate down-regulated genes were enriched in cuticle related and extracellular terms and even some death related terms. While wet-plate methods have been used for C. elegans to induce dauers, in the current study, we observed many dead animals, and few and inconsistent formation of dauers, suggesting that not all methodologies can be transferred from different nematode species.

This project aimed to assemble the transcriptome of *P. strongyloides* for the first time. Animals of multi-stage from different treatments were collected and sequenced with the intent of having as many genes expressed as possible. While we were able to produce a *de novo* transcriptome, there are some shortcomings in this project. First, the worms used in this project originated from a single isolated culture from CGC, therefore, it is likely to be fairly homogenous. Second, we used 2 replicate libraries for each treatment, and while we were able to obtain statistical significance, more replicates would increase statistical power. However, we should point out that data from RNA-seq libraries were internally normalized and should improve accuracy of our observations.

In summary, this study obtained a *de novo* assembled transcriptome of *P. strongyloides*. The differential analysis of starved, wet-plate and normal samples allowed us to find the dauer animal. This suggests that this species might be a surrogate system for nematode-related human or pet parasitic diseases. Since there are currently only 96 nematodes studied with genetic information available at genome and transcriptome level to our knowledge, far fewer than the number of species in the Phylum Nematoda (estimated 100 million), this project provides valuable sequence information of a facultative parasite that can be used to investigate the evolution of parasitism in this phylum and provide resources for other comparative studies in the future.

Conflicts of interest

The authors declare that they have no conflict of interest.

Authors' contributions

MZ and GW designed the study. MZ assembled the sequence data and constructed and annotated the assembly. LH examined the assembled data and assessed the quality. KEK provided the essential computing server resources. MZ wrote and GW edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments. Special thanks to Lakhansing Pardeshi and Pooja Sethiya for their patience and help with some of the computational advice. We thank Merja Lakso, Ka Lai Leung, Chenyin Wang for experimental assistance, and Yang Yang, Changliang Wang, Liang Chen for technical support. This research was supported by the Faculty of Health Sciences, University of Macau and the Academy of Finland (Project 258365 to KEK). Computational analyses were supported by High Performance Computing Cluster (HPC Cluster) in University of Macau and super cluster taito in CSC, Finnish IT Center for Science.

Figure Legends.

Table 1. D*e novo* assembled and non-redundant transcripts summary

Table 2. *P. strongyloides* homologous genes to *C.elegans* RNA interference related genes

Table 3. Starvation up-regulated dauer related genes

Figure 1. Phylogenetic location of nematodes and Venn diagram of *P. strongyloides* transcripts BLASTX matches. (**A)** Branches with a number indicate the corresponding clade in

nematoda phylum summarized from Megen et al (2009). (**B)** Venn diagram showing distribution of *P. strongyloides* transcript BLASTX matches by protein among 5 nematode species with different lifestyles. Nematodes matched are: *C. elegans* (free-living), *Strongyloides stercoralis* (parasitic), *Necator americanus* (parasitic), *Trichuris trichiura* (parasitic) and *Pristionchus pacificus* (parasitic).

Figure 2. Contig length distribution of Trinity assembly. Numbers of contigs with lengths from 300 bp to 4,900 bp were counted every 200 bp. The number of contigs over 5,000 bp long were merged into one bin.

Figure 3. Transcriptome KEGG pathway (P < 0.05) based on BLASTX results. The axis is P-value treated by –LOG10. Pathways in the inset area are top 15 lowest p-value, and their mapped transcript number is shown in the colored bar chart.

Figure 4. Trinotate unique transcripts annotation. The transcriptome is first annotated by BLASTX matched by *C.elegans*, then the remnant was annotated by Trinotate based on peptide homologous in GO/KEGG/EGGnog databases. The pie chart shows the distribution of each annotation type. The axis is P-value treated by –LOG10.

Figure 5. Differentially expressed transcripts among different treatments. N: normal treatment (replicate: N1, N2), S: starved sample (replicate: S1, S2); W: wet-plate sample (replicate: W1, W2). **(A)** Differentially expressed transcripts clustered among treated replicates. **(B)** Correlation between replicates of normal, starved and wet-plate samples of differentially expressed transcripts. **(C)** Correlation between replicates of samples across all transcripts.

Figure 6. GO enrichment analysis of differentially expressed genes. (**A)** starvation up-regulated gene enrichment result. (**B)** starvation down-regulated gene enrichment result. (**C)** wet-plate down-regulated gene enrichment result.

Figure 7. Dauer location, morphology and related count statistics. (A) the red box highlights where the dauer animals were found in starved plates; enlarged phasmid of dauer with both lateral

and vertical views. (**B)** Boxplot and correlation of dauer and normal larva. (**C)** Relative proportions of adult, larvae, and dauer animals in each treatment. Data shown are average ± S.D. from 4 individual plates for each treatment. Three technical replicates were sampled from each plate. N, normal; S, starving; W, wet.

References

1. Tanaka, A., et al., *Pelodera strongyloides infestation presenting as pruritic dermatitis.* J. Am. Acad. Dermatol., 2004. **51**(5): p. S181-S184. DOI: 10.1016/j.jaad.2004.05.010
2. van Megen, H., et al., *A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences.* Nematology, 2009. **11**(6): p. 927-950. DOI: 10.1163/156854109X456862
3. Saari, S.A. and S.E. Nikander, *Pelodera (syn. Rhabditis) strongyloides as a cause of dermatitis–a report of 11 dogs from Finland.* Acta Vet. Scand., 2006. **48**(1): p. 18. DOI: 10.1186/1751-0147-48-18
4. Vangestel, S., *Comparative and phylogenetic analysis of the early embryonic development in the phylum Nematoda*. 2008, Ghent University. DOI: 1854/13577
5. Bird, A.F. and J. Bird, *The structure of nematodes*. 2012: Academic Press. pp. 230.
6. Cliff, G. and R. Anderson, *Development of Pelodera strongyloides (Schneider, 1860) Schneider, 1866 (Nematoda: Rhabditidae) in culture.* J. Helminthol., 1980. **54**(02): p. 135-146. DOI: 10.1017/S0022149X00006489
7. Blaxter, M. and G. Koutsovoulos, *The evolution of parasitism in Nematoda.* Parasitology, 2015. **142**(S1): p. S26-S39. DOI: 10.1017/S0031182014000791
8. Poinar Jr, G.O., *Nematodes as facultative parasites of insects.* Annu. Rev. Entomol., 1972. **17**(1): p. 103-122. DOI: 10.1146/annurev.en.17.010172.000535
9. Poinar, G., *Life history of Pelodera strongyloides (Schneider) in the orbits of murid rodents in Great Britain.* Proc Helminthol Soc Wash, 1965. **32**: p. 148-151. DOI: 10.1017/S0022149X00006489
10. Kipnis, R. and K. Todd Jr, *Pelodera strongyloides in the urine of a cat.* Feline Pract., 1977.
11. Jones, C., T. Rosen, and C. Greenberg, *Cutaneous larva migrans due to Pelodera strongyloides.* Cutis, 1991. **48**(2): p. 123-126. PMID: 1935236
12. Ramos, J., et al., *Pelodera dermatitis in sheep.* Vet. Rec., 1996. **138**: p. 474-474. PMID: 8735541
13. Rashmir-Raven, A.M., et al., *Papillomatous pastern dermatitis with spirochetes and Pelodera strongyloides in a Tennessee Walking Horse.* J. Vet. Diagn. Invest., 2000. **12**(3): p. 287-291. DOI: 10.1177/104063870001200320
14. Yeruham, I. and S. Perl, *Dermatitis in a dairy herd caused by Pelodera strongyloides (Nematoda: Rhabditidae).* Zoonoses and Public Health, 2005. **52**(4): p. 197-198.
15. Fitzgerald, S.D., T.M. Cooley, and M.K. Cosgrove, *Sarcoptic mange and Pelodera dermatitis in an American black bear (Ursus americanus).* J. Zoo Wildl. Med., 2008. **39**(2): p. 257-259. DOI: 10.1638/2007-0071R.1
16. McHuron, E.A., et al., *Pelodera strongyloides infection in pacific Harbor seals (Phoca vitulina richardii) from California.* J. Zoo Wildl. Med., 2013. **44**(3): p. 799-802. DOI: 10.1638/2013-0027.1

17. Fitch, D., B. Bugaj-Gaweda, and S.W. Emmons, *18S ribosomal RNA gene phylogeny for some Rhabditidae related to Caenorhabditis.* Mol. Biol. Evol., 1995. **12**(2): p. 346-358. DOI: 10.1093/oxfordjournals.molbev.a040207
18. Kiontke, K., et al., *Trends, stasis, and drift in the evolution of nematode vulva development.* Curr. Biol., 2007. **17**(22): p. 1925-1937. DOI: 10.1016/j.cub.2007.10.061
19. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome.* Nat. Biotechnol., 2011. **29**(7): p. 644-652. DOI: 10.1038/nbt.1883
20. Stiernagle, T., *Maintenance of C. elegans.* C. elegans, 1999. **2**: p. 51-67. DOI: 10.1895/wormbook.1.101.1
21. Weller, A.M., *The wet plate protocol: an efficient way to obtain dauer larvae.* A rapid nematode preparation for microscopy, 2010. http://wbg.wormbook.org/2010/12/06/the-wet-plate-protocol-an-efficient-way-to-obtain-dauer-larvae/. Accessed 1 APRIL 2019.
22. Kumar, M., et al., *De novo transcriptome sequencing and analysis of the cereal cyst nematode, Heterodera avenae.* PloS one, 2014. **9**(5): p. e96311.
23. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome biol., 2009. **10**(3): p. R25. DOI: 10.1371/journal.pone.0096311
24. Li, R., et al., *De novo assembly of human genomes with massively parallel short read sequencing.* Genome Res., 2010. **20**(2): p. 265-272. DOI: 10.1101/gr.097261.109
25. Li, W. and A. Godzik, *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.* Bioinformatics, 2006. **22**(13): p. 1658-1659. DOI: 10.1093/bioinformatics/btl158
26. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.* Nat. Protoc., 2013. **8**(8): p. 1494-1512. DOI: 10.1038/nprot.2013.084
27. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015: p. btv351. DOI: 10.1093/bioinformatics/btv351
28. Boeckmann, B., et al., *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.* Nucleic Acids Res., 2003. **31**(1): p. 365-370. PMID: 12520024
29. Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic Acids Res., 2011: p. gkr367. DOI: 10.1093/nar/gkr367
30. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J. Mol. Biol., 2001. **305**(3): p. 567-580. DOI: 10.1006/jmbi.2000.4315
31. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**(1): p. 1-16. DOI: 10.1186/1471-2105-12-323
32. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-140. DOI: 10.1093/bioinformatics/btp616
33. Inoue, T. and J.H. Thomas, *Targets of TGF-β signaling in Caenorhabditis elegans dauer formation.* Dev. Biol., 2000. **217**(1): p. 192-204. DOI: 10.1006/dbio.1999.9545
34. Scanlan, L.D., et al., *Counting Caenorhabditis elegans: Protocol Optimization and Applications for Population Growth and Toxicity Studies in Liquid Medium.* Sci. Rep., 2018. **8**. DOI: 10.1038/s41598-018-19187-3
35. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 1 APRIL 2019.
36. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* nature, 1998. **391**(6669): p. 806-811. DOI: 10.1038/35888

37. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses.* Nat. Rev. Genet., 2014. **15**(2): p. 121-132. DOI: 10.1038/nrg3642
38. Spieth, J., et al., *Overview of gene structure in C. elegans.* 2005. WormBook. 2014; doi/10.1895/wormbook.1.65.2. DOI: 10.1895/wormbook.1.65.2
39. Thorpe, J.P., *The molecular clock hypothesis: biochemical evolution, genetic differentiation and systematics.* Annu Rev Ecol Syst, 1982. **13**(1): p. 139-168.
40. Snel, B., P. Bork, and M.A. Huynen, *Genome phylogeny based on gene content.* Nat. Genet., 1999. **21**(1): p. 108-110. DOI: 10.1038/505
41. Hannon, G.J., *RNA interference.* Nature, 2002. **418**(6894): p. 244-251. DOI: 10.1038/418244a
42. Selkirk, M.E., et al., *The development of RNA interference (RNAi) in gastrointestinal nematodes.* Parasitology, 2012. **139**(5): p. 605-612. DOI: 10.1017/S0031182011002332
43. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat. Genet., 2003. **33**: p. 245-254. DOI: 10.1038/ng1089
44. Stepek, G., G. McCormack, and A.P. Page, *Collagen processing and cuticle formation is catalysed by the astacin metalloprotease DPY-31 in free-living and parasitic nematodes.* Int. J. Parasitol., 2010. **40**(5): p. 533-542. DOI: 10.1016/j.ijpara.2009.10.007
45. Maizels, R.M., M.L. Blaxter, and A.L. Scott, *Immunological genomics of Brugia malayi: filarial genes implicated in immune evasion and protective immunity.* Parasite Immunol., 2001. **23**(7): p. 327-344. PMID: 11472553
46. Davis, E.L., R.S. Hussey, and T.J. Baum, *Getting to the roots of parasitism by nematodes.* Trends Parasitol., 2004. **20**(3): p. 134-141. DOI: 10.1016/j.pt.2004.01.005

Credit Author Statement

MZ and GW designed the study. MZ assembled the sequence data and constructed and annotated the assembly. LH examined the assembled data and assessed the quality. KEK provided the essential computing server resources. MZ wrote and GW edited the manuscript. All authors read and approved the final manuscript.

Table 1. De novo assembled and non-redundant transcripts summary

	Trinity_cl_100	Trinity	SOAPdenovo
Total trinity 'genes' number	72805	73071	155384
Total trinity transcripts no.	104634	104962	178118
Percent GC	43.66	43.66	42.62
Transcript contig N50/bp	2195	2198	476
Median transcript contig length	489	490	168
Average transcript contig	1102.67	1104.51	319
Total transcript assembled bases	115376920	115931948	56993506
GENE' contig N50	1579	1583	1038
Median 'GENE' contig length	360	361	166
Average 'GENE' contig	783.87	785.65	407
Total 'GENE' assembled bases	57069323	57408459	63220198

Table 2. P. strongyloides homologous genes to *C.elegans* RNA interference related genes

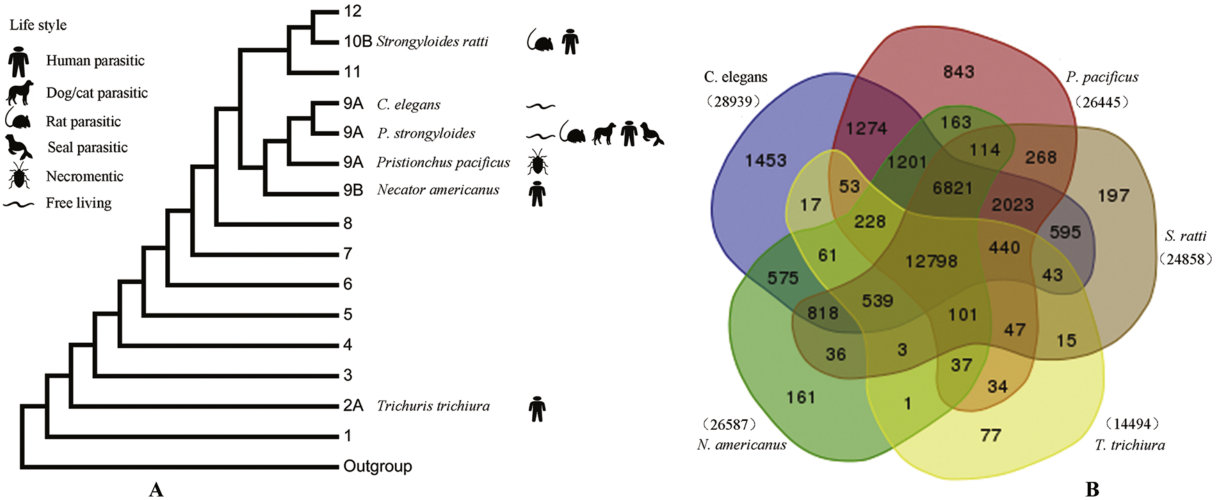
Transcript ID	C. elegans gene ID	Function/phenotype
TRINITY_DN10682_c0_g1 TRINITY_DN10682_c0_g2 TRINITY_DN10682_c0_g3 TRINITY_DN20412_c2_g1	*alg-1*	Argonaut ortholog
TRINITY_DN24358_c0_g1	*alg-4*	Argonaute (AGO)
TRINITY_DN18852_c0_g1	*drh-1*	Dicer-related helicase that contains a DExD/H-box helicase domain
TRINITY_DN21014_c1_g1	*drh-3*	DEAH/D-box helicase
TRINITY_DN4627_c0_g1	*rrf-3*	RNA-directed RNA polymerase (RdRP) homolog
TRINITY_DN14696_c0_g2	*drsh-1*	RNase III-type ribonuclease orthologous to Drosophila and human Drosha
TRINITY_DN10471_c0_g1 TRINITY_DN17651_c1_g1 TRINITY_DN21553_c1_g1	*ego-1*	Homolog of RNA-directed RNA polymerase
TRINITY_DN3780_c0_g2 TRINITY_DN3780_c0_g1	*eri-1*	SAP/SAF box domain and a DEDDh-like 3'-5' exonuclease domain
TRINITY_DN20333_c0_g1	*prg-1*	Piwi subfamily protein of highly conserved Argonaut/Piwi proteins
TRINITY_DN20784_c1_g3	*sid-3*	Regulate the import of dsRNA into cells
TRINITY_DN20492_c1_g1	*fem-1*	femal (no sperm)
TRINITY_DN14313_c0_g1	*hlh-1*	lumpy-dumpy larvae
TRINITY_DN16389_c0_g1	*unc-22*	Strong twitchers
TRINITY_DN16775_c0_g1 TRINITY_DN16775_c0_g2 TRINITY_DN16775_c1_g1 TRINITY_DN16775_c2_g2 TRINITY_DN16775_c3_g1 TRINITY_DN20497_c0_g1 TRINITY_DN20497_c0_g2 TRINITY_DN20497_c0_g3 TRINITY_DN20497_c0_g5 TRINITY_DN20497_c1_g1 TRINITY_DN20497_c2_g1 TRINITY_DN20497_c3_g1 TRINITY_DN20497_c4_g1 TRINITY_DN21170_c0_g1 TRINITY_DN21170_c0_g2 TRINITY_DN21170_c1_g2 TRINITY_DN21170_c2_g1 TRINITY_DN21170_c5_g1 TRINITY_DN21170_c7_g1	*unc-54*	Paralysed

Table 3. Starvation up-regulated dauer related genes

ID	Locus	Symbol	Description	FDR
TRINITY_DN11239_c0_g2	*hsp-12.6*	F38E11.2	may contribute to prolonged lifespan in dauer larvae	2.46E-38
TRINITY_DN20807_c3_g2	*col-40*	T13B5.4	present in L1 larvae and at the L2d-dauer molt	3.49E-19
TRINITY_DN18259_c0_g1	*cut-1*	C47G2.1	alae formation and radial shrinking in dauer differentiation	1.92E-10

Highlights

- The transcriptome of a facultative nematode *Pelodera strongyloides* was sequenced.
- Assembled transcriptome BLASTX alignments with 1 free-living and 4 parasitic nematodes were consistent with their evolutionary relationships.
- A total of 104,634 transcript contigs with N50 contig length of 2,195 nucleotides were obtained.
- Starving conditions resulted in up-regulation of cuticle related and dauer formation genes.



Life style
Human parasitic
Dog/cat parasitic
Rat parasitic
Seal parasitic
Necromentic
Free living
12
10B Strongyloides ratti
11
9A C. elegans
9A P. strongyloides
9A Pristionchus pacificus
9B Necator americanus
8
7
6
5
4
3
2A Trichuris trichiura
1
Outgroup
A
C. elegans
（28939）
P. pacificus
（26445）
S. ratti
（24858）
（26587）
N. americanus
（14494）
T. trichiura
843
1274
163
1453
114
268
1201
53
6821
197
17
2023
228
595
61
440
12798
575
43
539
818
101
47
15
36
3
37
34
161
1
77
B

Figure 1

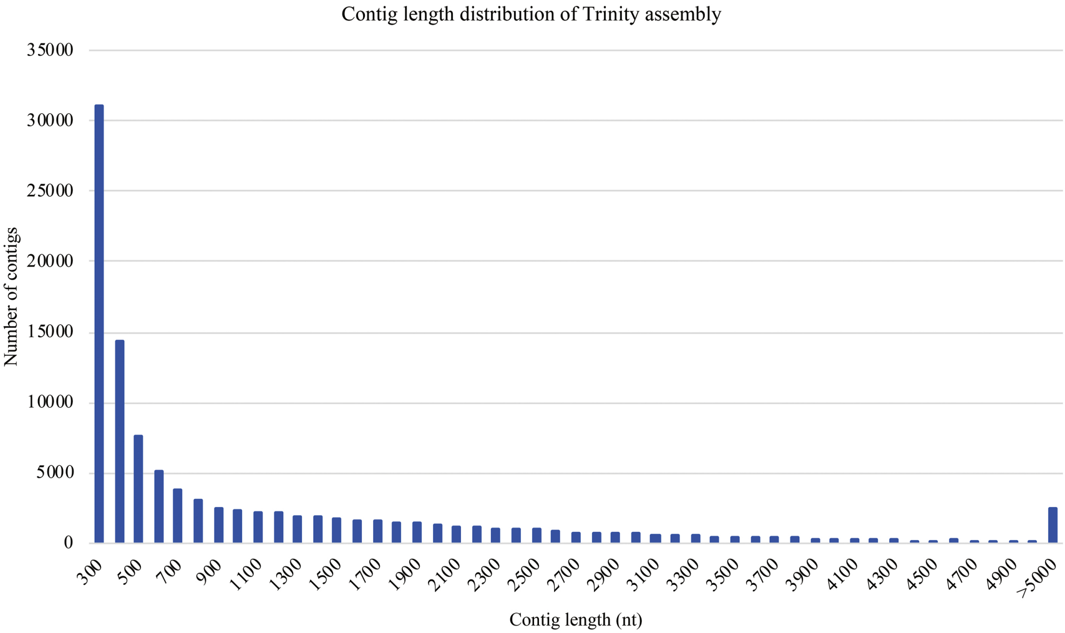

Contig length distribution of Trinity assembly
35000
30000
25000
20000
15000
10000
5000
0
Number of contigs
300
500
700
900
1100
1300
1500
1700
1900
2100
2300
2500
2700
2900
3100
3300
3500
3700
3900
4100
4300
4500
4700
4900
>5000
Contig length (nt)

Figure 2

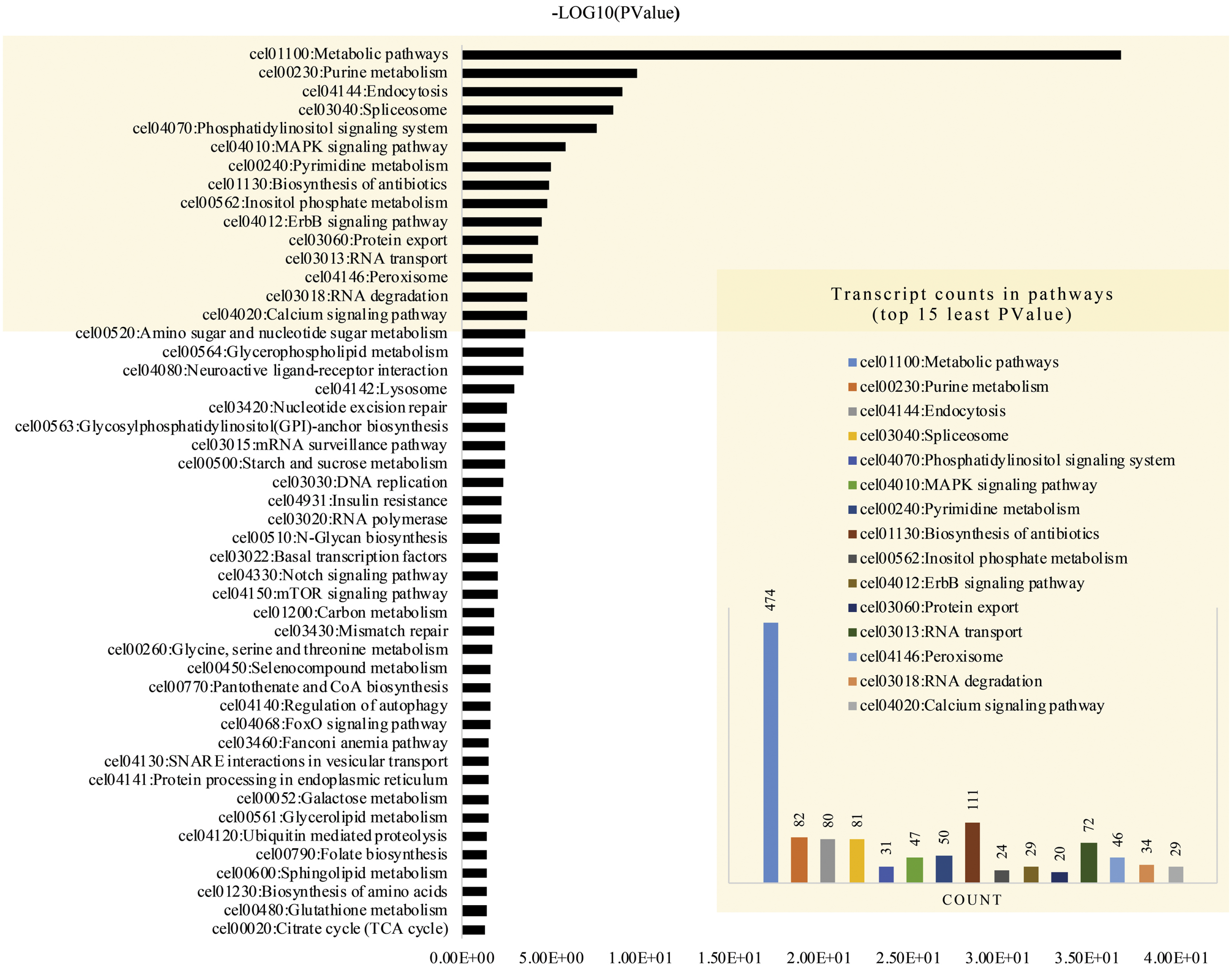

-LOG10(PValue)
cel01100:Metabolic pathways
cel00230:Purine metabolism
cel04144:Endocytosis
cel03040:Spliceosome
cel04070:Phosphatidylinositol signaling system
cel04010:MAPK signaling pathway
cel00240:Pyrimidine metabolism
cel01130:Biosynthesis of antibiotics
cel00562:Inositol phosphate metabolism
cel04012:ErbB signaling pathway
cel03060:Protein export
cel03013:RNA transport
cel04146:Peroxisome
cel03018:RNA degradation
cel04020:Calcium signaling pathway
cel00520:Amino sugar and nucleotide sugar metabolism
cel00564:Glycerophospholipid metabolism
cel04080:Neuroactive ligand-receptor interaction
cel04142:Lysosome
cel03420:Nucleotide excision repair
cel00563:Glycosylphosphatidylinositol(GPI)-anchor biosynthesis
cel03015:mRNA surveillance pathway
cel00500:Starch and sucrose metabolism
cel03030:DNA replication
cel04931:Insulin resistance
cel03020:RNA polymerase
cel00510:N-Glycan biosynthesis
cel03022:Basal transcription factors
cel04330:Notch signaling pathway
cel04150:mTOR signaling pathway
cel01200:Carbon metabolism
cel03430:Mismatch repair
cel00260:Glycine, serine and threonine metabolism
cel00450:Selenocompound metabolism
cel00770:Pantothenate and CoA biosynthesis
cel04140:Regulation of autophagy
cel04068:FoxO signaling pathway
cel03460:Fanconi anemia pathway
cel04130:SNARE interactions in vesicular transport
cel04141:Protein processing in endoplasmic reticulum
cel00052:Galactose metabolism
cel00561:Glycerolipid metabolism
cel04120:Ubiquitin mediated proteolysis
cel00790:Folate biosynthesis
cel00600:Sphingolipid metabolism
cel01230:Biosynthesis of amino acids
cel00480:Glutathione metabolism
cel00020:Citrate cycle (TCA cycle)
0.00E+00
5.00E+00
1.00E+01
1.50E+01
2.00E+01
2.50E+01
3.00E+01
3.50E+01
4.00E+01
Transcript counts in pathways
(top 15 least PValue)
cel01100:Metabolic pathways
cel00230:Purine metabolism
cel04144:Endocytosis
cel03040:Spliceosome
cel04070:Phosphatidylinositol signaling system
cel04010:MAPK signaling pathway
cel00240:Pyrimidine metabolism
cel01130:Biosynthesis of antibiotics
cel00562:Inositol phosphate metabolism
cel04012:ErbB signaling pathway
cel03060:Protein export
cel03013:RNA transport
cel04146:Peroxisome
cel03018:RNA degradation
cel04020:Calcium signaling pathway
474
82
80
81
31
47
50
111
24
29
20
72
46
34
29
COUNT

Figure 3

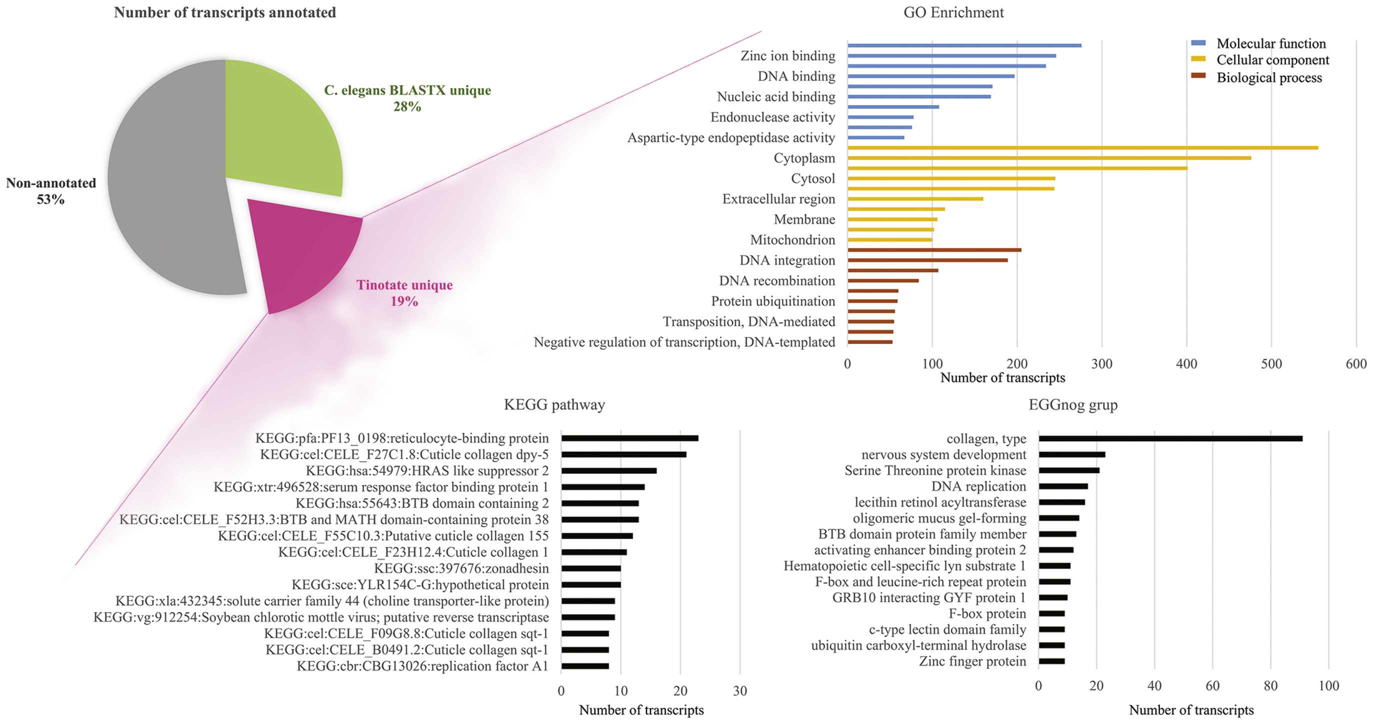

Number of transcripts annotated
C. elegans BLASTX unique
28%
Non-annotated
53%
Tinotate unique
19%
GO Enrichment
Molecular function
Cellular component
Biological process
Zinc ion binding
DNA binding
Nucleic acid binding
Endonuclease activity
Aspartic-type endopeptidase activity
Cytoplasm
Cytosol
Extracellular region
Membrane
Mitochondrion
DNA integration
DNA recombination
Protein ubiquitination
Transposition, DNA-mediated
Negative regulation of transcription, DNA-templated
0
100
200
300
400
500
600
Number of transcripts
KEGG pathway
KEGG:pfa:PF13_0198:reticulocyte-binding protein
KEGG:cel:CELE_F27C1.8:Cuticle collagen dpy-5
KEGG:hsa:54979:HRAS like suppressor 2
KEGG:xtr:496528:serum response factor binding protein 1
KEGG:hsa:55643:BTB domain containing 2
KEGG:cel:CELE_F52H3.3:BTB and MATH domain-containing protein 38
KEGG:cel:CELE_F55C10.3:Putative cuticle collagen 155
KEGG:cel:CELE_F23H12.4:Cuticle collagen 1
KEGG:ssc:397676:zonadhesin
KEGG:sce:YLR154C-G:hypothetical protein
KEGG:xla:432345:solute carrier family 44 (choline transporter-like protein)
KEGG:vg:912254:Soybean chlorotic mottle virus; putative reverse transcriptase
KEGG:cel:CELE_F09G8.8:Cuticle collagen sqt-1
KEGG:cel:CELE_B0491.2:Cuticle collagen sqt-1
KEGG:cbr:CBG13026:replication factor A1
0
10
20
30
Number of transcripts
EGGnog grup
collagen, type
nervous system development
Serine Threonine protein kinase
DNA replication
lecithin retinol acyltransferase
oligomeric mucus gel-forming
BTB domain protein family member
activating enhancer binding protein 2
Hematopoietic cell-specific lyn substrate 1
F-box and leucine-rich repeat protein
GRB10 interacting GYF protein 1
F-box protein
c-type lectin domain family
ubiquitin carboxyl-terminal hydrolase
Zinc finger protein
0
20
40
60
80
100
Number of transcripts

Figure 4

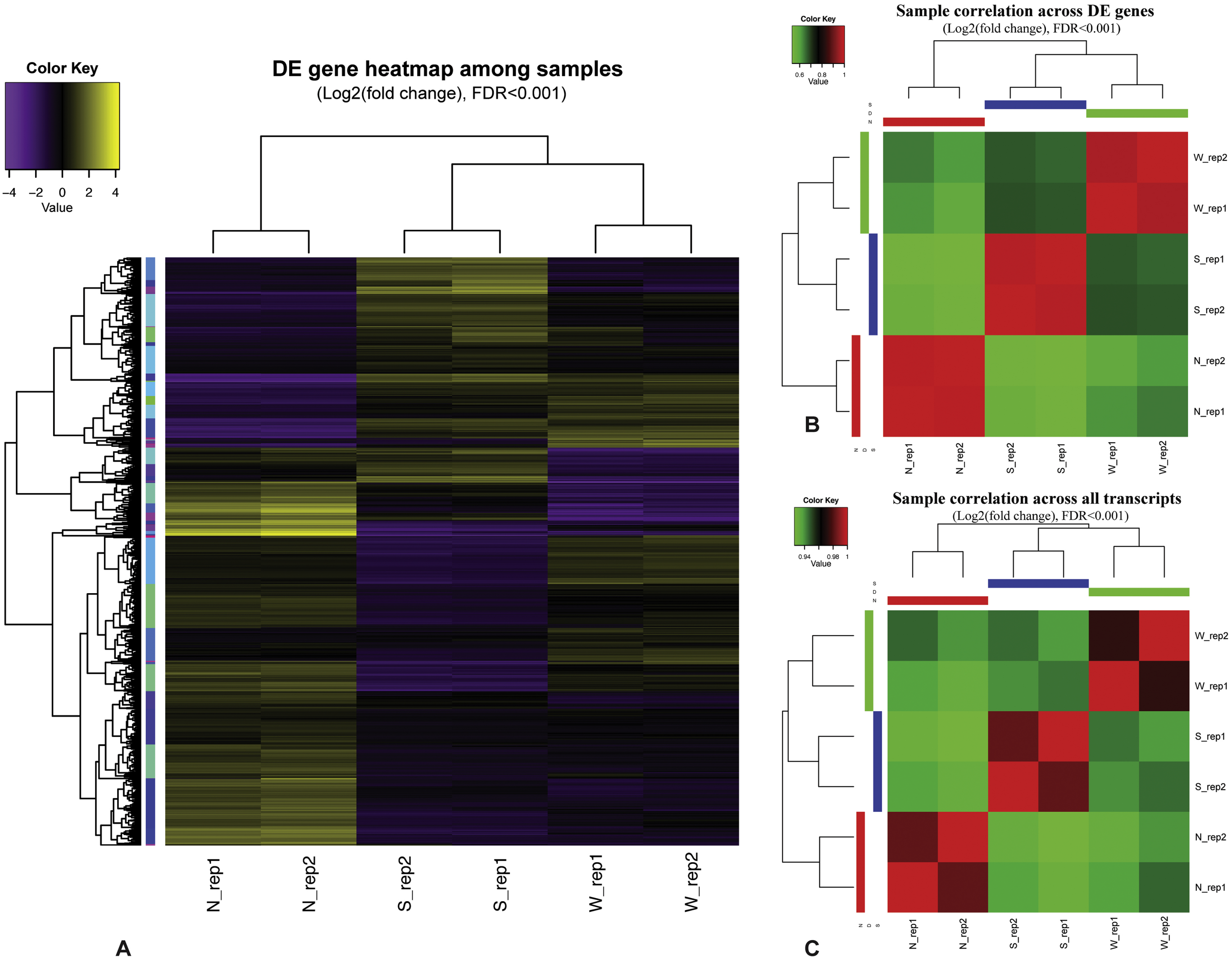

Color Key
−4 −2 0 2 4
Value
DE gene heatmap among samples
(Log2(fold change), FDR<0.001)
N_rep1
N_rep2
S_rep2
S_rep1
W_rep1
W_rep2
A
Color Key
0.6 0.8 1
Value
Sample correlation across DE genes
(Log2(fold change), FDR<0.001)
S
D
N
W_rep2
W_rep1
S_rep1
S_rep2
N_rep2
N_rep1
B
N D S
N_rep1
N_rep2
S_rep2
S_rep1
W_rep1
W_rep2
Color Key
0.94 0.98 1
Value
Sample correlation across all transcripts
(Log2(fold change), FDR<0.001)
S
D
N
W_rep2
W_rep1
S_rep1
S_rep2
N_rep2
N_rep1
C
N D S
N_rep1
N_rep2
S_rep2
S_rep1
W_rep1
W_rep2

Figure 5

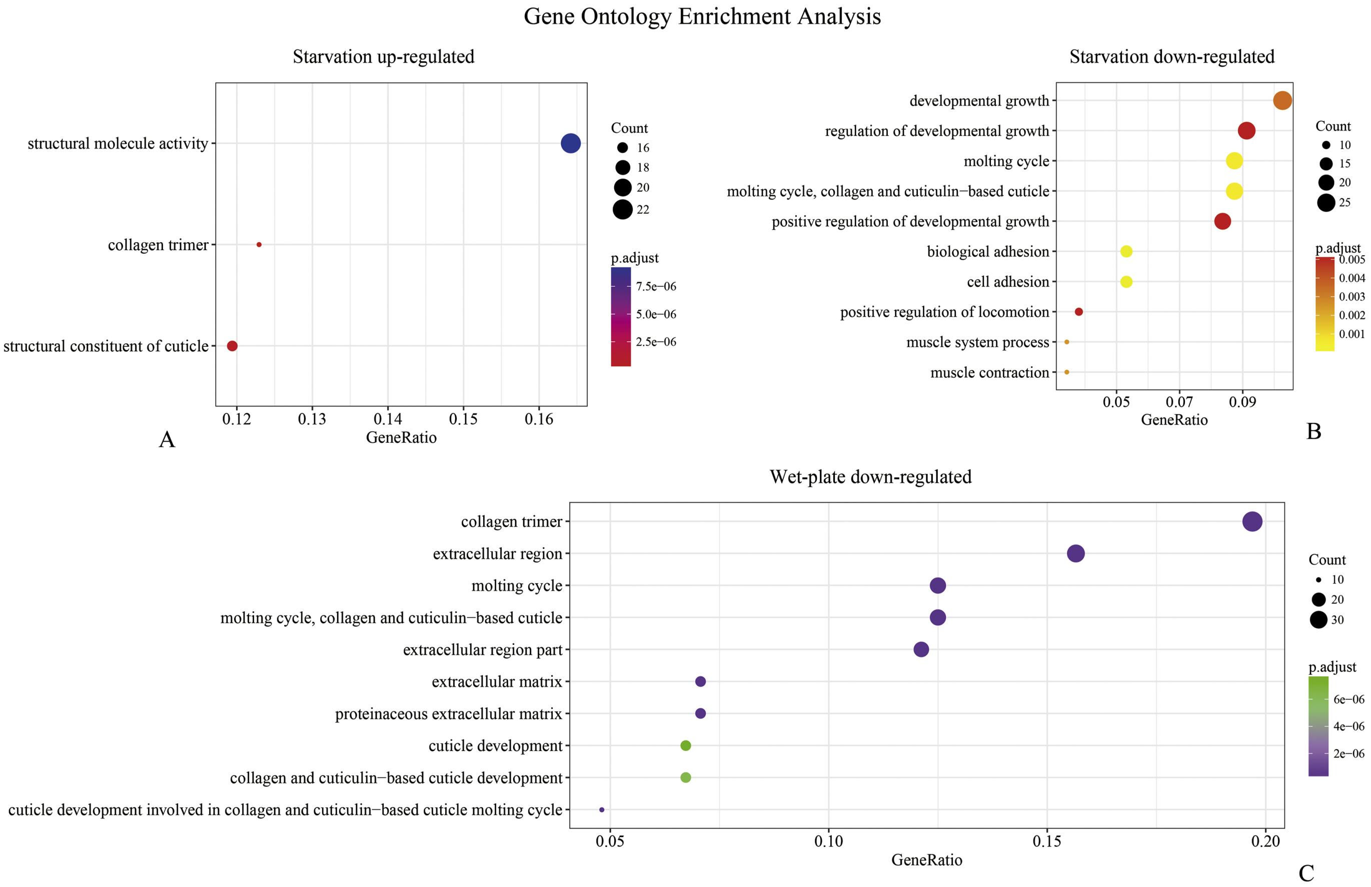

Gene Ontology Enrichment Analysis
Starvation up-regulated
structural molecule activity
collagen trimer
structural constituent of cuticle
0.12
0.13
0.14
0.15
0.16
GeneRatio
Count
16
18
20
22
p.adjust
7.5e−06
5.0e−06
2.5e−06
A
Starvation down-regulated
developmental growth
regulation of developmental growth
molting cycle
molting cycle, collagen and cuticulin−based cuticle
positive regulation of developmental growth
biological adhesion
cell adhesion
positive regulation of locomotion
muscle system process
muscle contraction
0.05
0.07
0.09
GeneRatio
Count
10
15
20
25
p.adjust
0.005
0.004
0.003
0.002
0.001
B
Wet-plate down-regulated
collagen trimer
extracellular region
molting cycle
molting cycle, collagen and cuticulin−based cuticle
extracellular region part
extracellular matrix
proteinaceous extracellular matrix
cuticle development
collagen and cuticulin−based cuticle development
cuticle development involved in collagen and cuticulin−based cuticle molting cycle
0.05
0.10
0.15
0.20
GeneRatio
Count
10
20
30
p.adjust
6e−06
4e−06
2e−06
C

Figure 6

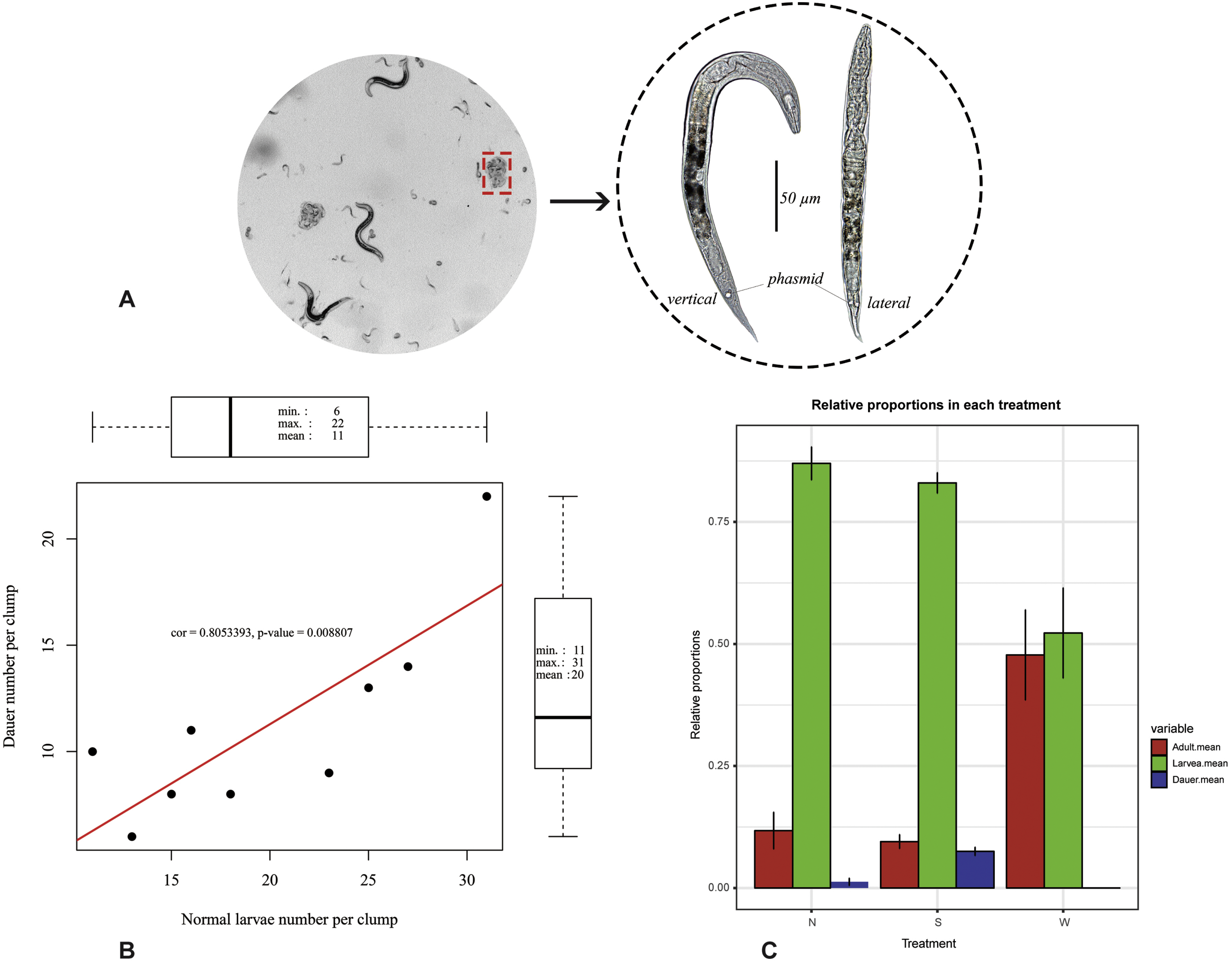

A
50 µm
phasmid
vertical
lateral
min. : 6
max. : 22
mean : 11
cor = 0.8053393, p-value = 0.008807
min. : 11
max.: 31
mean : 20
Dauer number per clump
10
15
20
15
20
25
30
Normal larvae number per clump
B
Relative proportions in each treatment
Relative proportions
0.00
0.25
0.50
0.75
N
S
W
Treatment
variable
Adult.mean
Larvea.mean
Dauer.mean
C

Figure 7

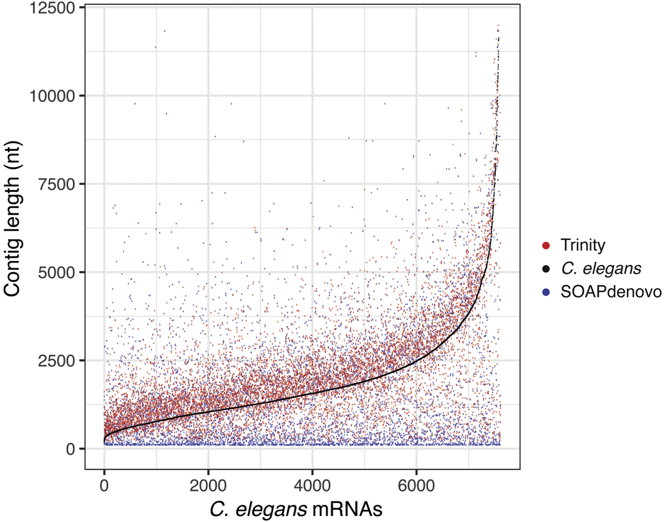

Contig length (nt)
12500
10000
7500
5000
2500
0
0
2000
4000
6000
C. elegans mRNAs
● Trinity
● *C. elegans*
● SOAPdenovo

Figure 8