

UC Irvine

UC Irvine Previously Published Works

Title

Hybrid Assembly of the Genome of the Entomopathogenic Nematode *Steinernema carpocapsae* Identifies the X-Chromosome.

Permalink

<https://escholarship.org/uc/item/0pp7685d>

Journal

G3: Genes, Genomes, Genetics, 9(8)

Authors

Serra, Lorryne
Macchietto, Marissa
Macias-Munoz, Aide
et al.

Publication Date

2019-08-08

DOI

10.1534/g3.119.400180

Peer reviewed

Hybrid Assembly of the Genome of the Entomopathogenic Nematode *Steinernema carpocapsae* Identifies the X-Chromosome

Lorrayne Serra,^{*1} Marissa Macchietto,^{†1} Aide Macias-Muñoz,^{*} Cassandra Joan McGill,^{*} Isaryhia Maya Rodriguez,^{*} Bryan Rodriguez,^{*} Rabi Murad,^{*} and Ali Mortazavi^{*,‡,2}

^{*}Department of Developmental and Cell Biology, University of California, Irvine, CA 92697, [†]Institute of Health Informatics, University of Minnesota, Minneapolis, MN, 55455, and [‡]Center for Complex Biological Systems, University of California, Irvine, CA 92697

ORCID ID: 0000-0002-4259-6362 (A.M.)

ABSTRACT Entomopathogenic nematodes from the genus *Steinernema* are lethal insect parasites that quickly kill their insect hosts with the help of their symbiotic bacteria. *Steinernema carpocapsae* is one of the most studied entomopathogens due to its broad lethality to diverse insect species and its effective commercial use as a biological control agent for insect pests, as well as a genetic model for studying parasitism, pathogenesis, and symbiosis. In this study, we used long-reads from the Pacific Biosciences platform and BioNano Genomics Irys system to assemble the most complete genome of the *S. carpocapsae* ALL strain to date, comprising 84.5 Mb in 16 scaffolds, with an N50 of 7.36 Mb. The largest scaffold, with 20.9 Mb, was identified as chromosome X based on sex-specific genome sequencing. The high level of contiguity allowed us to characterize gene density, repeat content, and GC content. RNA-seq data from 17 developmental stages, spanning from embryo to adult, were used to predict 30,957 gene models. Using this improved genome, we performed a macrosyntenic analysis to *Caenorhabditis elegans* and *Pristionchus pacificus* and found *S. carpocapsae*'s chromosome X to be primarily orthologous to *C. elegans*' and *P. pacificus*' chromosome II and IV. We also investigated the expansion of protein families and gene expression differences between adult male and female stage nematodes. This new genome and more accurate set of annotations provide a foundation for additional comparative genomic and gene expression studies within the *Steinernema* clade and across the Nematoda phylum.

KEYWORDS

Steinernema
Steinernema
genome
S. carpocapsae
Genetics of Sex

Nematodes are round, non-segmented worms with a simple body plan that populate all known biological niches. *Caenorhabditis elegans* is by far the best-characterized nematode species and was the first metazoan to have its genome assembled (*C. elegans* Sequencing Consortium

1998). Studies in *C. elegans* have been crucial for the study of a multitude of biological processes such as development, cell specification, cell differentiation, apoptosis and genome evolution (Deppe *et al.*, 1978; Kaletta and Hengartner 2006; Rothman *et al.*, 2014; Segref *et al.* 2010). Other free-living, androdieocious nematode genomes such as *Pristionchus pacificus* and *Panagrellus redivivus* have provided tools for genomic comparisons in developmental processes and phenotypic plasticity (Rödelsperger *et al.*, 2017; Srinivasan *et al.*, 2013).

However, free-living nematodes represent only a subset of the entire nematode phylum as they exclude parasitic nematodes, which are of most agricultural and biomedical interest. Parasitic nematodes have a diverse group of hosts including insects, plants, and humans. *Steinernema carpocapsae* is an entomopathogenic nematode (EPN) that is a lethal insect parasite but is not harmful to humans or plants. *S. carpocapsae* is of great interest because of its wide host range with approximately 200 possible insect hosts and the possible orthology of its toxins

Copyright © 2019 Serra *et al.*

doi: <https://doi.org/10.1534/g3.119.400180>

Manuscript received March 11, 2019; accepted for publication May 10, 2019; published Early Online May 21, 2019.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at FigShare: <https://doi.org/10.25387/g3.8069438>.

¹Equal contribution.

²Corresponding author: 2218 Biological Sciences III University of California Irvine, Irvine, CA 92697-2300. E-mail: ali.mortazavi@uci.edu

to mammalian-parasitic nematodes (Shapiro-Ilan *et al.*, 2017; Yang *et al.*, 2015). In many countries, *S. carpocapsae* are commercialized for use as insect pest control (Dillman *et al.*, 2012; Dito *et al.*, 2016). *Steinernema* has a symbiotic relationship with pathogenic *Xenorhabdus* bacteria where they collectively infect and kill a host within a few days, reproduce separately within the host, and reassociate when forming infective juveniles that will seek out the next host (Hirao *et al.*, 1999). *Steinernema* and its bacterial symbiont have been extensively studied as genetic models to explore symbiosis and pathogenesis (Martens and Goodrich-Blair 2005; Sicard *et al.*, 2003). EPNs are also an excellent satellite organism to study mammalian parasitism as they are closely related to the *Strongyloidea* nematodes (Blaxter *et al.*, 1998). In particular, *Steinernema*'s mechanism of host seeking by olfactory and other sensory cues likely offers a great model for mammalian-parasitic nematodes (Gang & Hallem 2016).

Previously published draft genomes done using Illumina sequencing for five *Steinernema* species opened the door to study evolutionary traits among these parasitic nematodes and compare them to free-living nematodes. These genomes helped elucidate gene families involved in parasitism that were expanded in *Steinernema*, such as proteases and protease inhibitors, while a comparative analysis to *C. elegans* revealed orthologous non-coding regulatory motifs (Dillman *et al.*, 2015). *S. carpocapsae* and *S. feltiae* were further used to study the extent of expression conservation in orthologous genes across nematode families that were associated with embryonic development (Macchietto *et al.*, 2017). This study found a funnel-shaped model of embryonic development based on high conservation of orthologous genes between *Caenorhabditis* and *Steinernema* (Macchietto *et al.*, 2017). The draft genomes also allowed for further studies in neuropeptide sensory perception in *S. carpocapsae*, and the study of lethal venom proteins in both *S. carpocapsae* and *S. feltiae* (Chang *et al.*, 2019, Lu *et al.*, 2017; Morris *et al.*, 2017).

The draft genome assembly of *S. carpocapsae* ALL strain had 1,578 contigs with an estimated genome size of 85.6 Mb, N50 of 300 kb and more than 28,000 predicted genes (Dillman *et al.*, 2015). Another study, also using short-reads assembled *S. carpocapsae* Breton strain into 347 scaffolds with an N50 of 1.24 Mb (Rougon-Cardoso *et al.*, 2016). The draft genome status, although helpful, can hinder further studies of protein families' expansions involved in parasitism. The high fragmentation also deters the analysis of conserved gene sequences, which could play a significantly larger role in the evolution of this phylum. Therefore, a more contiguous assembly of the *S. carpocapsae* genome will contribute to further understanding the mechanisms of *Steinernema* genome evolution. In this study, we reassembled the genome with long-read sequencing from Pacific BioSciences (PacBio) in conjunction with optical mapping using the BioNano Irys system, which produces high throughput physical maps (Lam *et al.*, 2012; Rhoads and Au 2015). The newly assembled genome consists of 16 scaffolds with nearly 31 thousand predicted genes and we identified the largest scaffold as chromosome X. We observed clusters of expansion of proteases families and found gene expression differences in catalytic activity between adult male and female stage nematodes.

MATERIALS AND METHODS

Axenic *S. carpocapsae* culture

S. carpocapsae infective Juveniles (IJs) were cultured on *Xenorhabdus nematophila* colonization defective mutant (HGB315) bacteria to produce axenic IJs. The IJs were first cultured *in vivo* in *Galleria mellonella* (wax-worms) and surface sterilized (Gaugler and Kaya 1990). *X. nematophila* bacteria (HGB315) were cultured in tryptic soy broth (cat. No. A00169,

BD Scientific) and shaken at 220rpm in a 26° incubator overnight. Cultures were plated on lipid agar (LA) plates (8 g/L of nutrient broth, 5 g/L of yeast extract, 2 g/L of MgCl₂, 7 ml/L of corn syrup, 4 ml/L of corn oil, and 15 g/L of Bacto Agar), using 600μL of culture per plate. Approximately 300,000 IJs were plated across 20 lipid agar plates (~15,000 IJs/plate) seeded with lawns of *Xenorhabdus nematophila*. Axenic IJs were collected from LA plates after 2-3 weeks by placing the LA plates in larger petri plates filled with DI water. The IJs were collected from the DI and surface sterilized (Gaugler and Kaya 1990). IJs were stored in Ringer's solution in tissue culture flasks at a density < 15 IJs/μL until enough IJs (~2 million) were collected for the DNA isolation.

PacBio DNA isolation

Sucrose float to remove dead IJs and bacterial contamination. Axenic IJs stored in the tissue culture flasks were transferred to conical tubes, spun down, and washed 3x with DI water. IJs were re-surface sterilized by adding 7.5 mL of distilled (DI) water to the worm pellet and 7.5 mL of egg solution (3 mL DI water, 4.5 mL 1 M NaOH, 2.5 mL of fresh Clorox bleach) and incubating for 5 min. The egg solution was immediately removed, and the IJs were rinsed 3 times by centrifuging at 2000 rpm for 1-2 min and then suspended in 7 mL of molecular grade water. 7 mL of cold 60% sucrose was added, and the sample was mixed and spun at 50x g (630 rpm/660 rpm) for 1 min at 4°, and then immediately at 1150x g (3000 rpm /3180 rpm) for 3 min at 4°. The live IJs were collected from the top layer and transferred to a sterile conical tube and washed 3x (centrifuging at 1650g (3600 rpm / 3810rpm) for 5 min) with molecular grade water.

Grinding of the IJs with a mortar and pestle in liquid nitrogen to break nematode cuticles. A mortar, pestle, and four ultracentrifuge tubes were wrapped in aluminum foil and autoclaved for at least 20 min at 121°. The cleaned IJs were added to the autoclaved mortar and liquid nitrogen was poured into the mortar and ground with the pestle until the liquid nitrogen evaporated. IJs were ground for 10 min adding liquid nitrogen as needed. The ground powder was transferred into a 50mL conical tube on ice.

IJ cell/tissue lysis and protein and RNA digestion. In a 15 mL conical, 9.5 mL Qiagen Buffer G2 from the Genomic DNA Buffer Set (Cat. No. 19060, Qiagen), 19 μL RNase A (100mg/mL; Cat. No. EN05231, Thermo Fisher) and 375 μL of proteinase K (> 800mili-absorbance units (mAu), Cat. No. P4850-5ML, Sigma-Aldrich) were mixed, added to the ground IJ sample, and incubated at 50° in a water bath for 3-4 hr until the lysate was clear. After incubated the sample was spun down at 5,000x g for 10 min at 4° to separate out any particulate matter that can clog the genomic tip column. The supernatant was transferred to a clean 15 mL conical tube.

Genomic DNA measurements. The total amount of genomic DNA was measured with the Qubit fluorometer to determine which genomic tip size to use. We used the 100G genomic-tip because we had between 80-100 μg of gDNA.

Genomic-tip protocol. A 100G Qiagen Genomic-tip (100/G) was equilibrated with 4 mL of Buffer QBT. Separately, the DNA sample was diluted with an equal volume of buffer QBT and vortexed for 10s at maximum speed and immediately applied to the equilibrated genomic-tip. The following steps of the genomic-tip were followed according to the manufacturer's instructions, except that the ethanol and isopropanol precipitation centrifugation steps were performed at 10,000x g for 45 min each. The DNA was eluted in EB buffer overnight at 55°. The DNA was sheared using 10 pumps of a blunt 24-gauge needle followed by 10 pumps using a blunt 21-gauge needle. The gDNA concentration, absorbance ratios, and fragment size range were determined before SMRT-bell library preparation.

S. carpocapsae genome assembly

The *S. carpocapsae* genome was assembled with Illumina reads and PacBio reads. Then, further improved in contiguity with BioNano genomics Iris system. Previously published Illumina libraries were assembled into high quality contigs using the Platanus (version 1.2.4) command *platanus assemble* with a k-mer size of 51 (-k 51) and a minimum k-mer coverage cut-off of 5 (-c 5) (Dillman *et al.*, 2015; Kajitani *et al.*, 2014). Then, we sequenced six SMRT cells on the PacBio RS II and obtained a total of 500,026 reads (616,611 sub reads) with a N50 read length of 18,308 bp and a mean subread length of 13,036 bp. This translated to 76.8X PacBio read coverage of the 85 Mb genome. All PacBio reads (76.8X genome coverage) were used with the PBcR pipeline into a self-assembly (PB assembly) (Chakraborty *et al.*, 2015; Koren *et al.*, 2012). Then, a hybrid assembly was assembled using the longest error-corrected PacBio reads that provided 25X coverage of the genome, which was output from the PBcR pipeline, with the Illumina assembly. The hybrid assembly was generated using the following DBG2OLC parameters: k 17, KmerCovTh 10, MinOverlap 100, AdaptiveTh 0.02, PathCovTh 3, RemoveChimera 1. The hybrid assembly and self-assembly was corrected for errors and chimeras using the Illumina and PacBio reads with quiver from the smrtanalysis package from PacBio (version 2.3.0p5) and subsequently Pilon (version 1.16) (Walker *et al.* 2014). Lastly, the PB-assembly and hybrid assembly were merged with quickmerge and MUMmer (version 3.23) using the default settings to generate a *merged assembly* (Chakraborty *et al.* 2015; Kurtz *et al.* 2004).

BioNano: DNA was extracted from *S. carpocapsae* Infective Juveniles (IJs) using the Animal Tissue DNA Isolation kit (Bionano Genomics). Bionano Irys optical data were generated and assembled with IrysSolve 2.1. We then merged the Bionano assembly with the merged assembly using IrysSolve, retaining Bionano assembly features when the two assemblies disagreed (Figure S1A).

The new *S. carpocapsae* genome assembly also went through HaploMerger to create an assembly with minimum possible number of haplotypes. First, the assembly was softmasked with Windowmasker (version 2.2.22) (Morgulis *et al.*, 2006). Next, the softmasked genome was cleaned using the *faDnaPolishing.pl* script provided by HaploMerger2 (Huang *et al.*, 2012). Then, the longest 5% of the genome was used as a target and the remaining 95% was used as a query for alignment with Lastz, which created a score matrix. The alignment threshold was kept at 95% to best identify heterozygosity. Lastly, the *S. carpocapsae* ALL strain was compared to the Breton strain through genomeevolution.org using synmap function with default settings (Lyons & Freeling 2008).

de novo developmental transcriptome assembly with Oases

Smartseq2 RNA-seq datasets spanning 16 developmental stages (zygote, 2-cell, 4-cell, 8-cell, 24-44-cell, 64-78-cell, comma, 1.fivefold, twofold, moving, L1, nonactivated IJ, 9-15h activated IJ, L4, male, and female) in replicates of 2-4 were assembled into stage-specific transcriptomes using Oases 0.2.8 and Velvet 1.2.10. Stage-specific transcriptomes were assembled using different kmer sizes from 22-35 in steps of 2, and then merged into a final assembly. The merged final transcriptome assemblies were compiled into a single fasta file to use for gene model prediction.

Genome annotation training and prediction with Augustus

539 *S. carpocapsae* genes sequences that matched 458 *C. elegans* CEGMA genes (OrthoMCL) were used together with *de novo* assembled

stage-specific transcriptomes to automatically train Augustus gene model exon and intron prediction (Stanke *et al.* 2008). *C. elegans* UTR models provided by Augustus were used for predicting the *S. carpocapsae* UTRs. A compiled set of *de novo* assembled stage-specific transcriptomes (from above) was mapped onto the merged genome using Blat version 36 with the following settings: -maxIntron = 70000 -minScore = 100 -minIdentity = 94. The best alignments were taken using the command *pslReps* with the setting: -singleHit, and were then sorted and converted into hints file for Augustus using *blat2hints.pl* from Augustus (version 3.2.1).

Assessing genome assembly completeness with Benchmarking Universal Single-Copy Orthologs (BUSCO)

Genome completeness was checked with BUSCO v3 software with default settings for genome and using near-universal single-copy orthologs selected from OrthoDB v9 nematoda_odb9 (Simão *et al.*, 2015). Nematode_odb9 has 982 groups for which 854 groups were found in the *S. carpocapsae* assembly, which is an 87% completeness (Figure S1B).

S. carpocapsae female and male DNA collection

IJs were cultured with *X. nematophila* for approximately 54 hr to collect 100 males and 62 hr to collect 50 females. Female and male DNA were extracted with DNeasy blood and Tissue kit (Qiagen Cat No. 69504). The DNA of females and males were processed following the steps F through J from protocol of Serra *et al.* (2018). Briefly, DNA was tagged using the Nextera DNA library prep kit (Illumina, FC-121-1030). Then, tagged DNA was amplified using Phusion High Fidelity PCR master mix with the amplification program set to 1) 72° 5 min. 2) 98° 30 sec. 3) 98° 10 sec, 63° 30 sec., 72° 1 min. (repeat 10x). 4) 4° Hold. PCR samples were cleaned with Ampure XP beads with a 1:1 ratio, concentrations measured with Qubit fluorometer and bioAnalyzed with Agilent 2100 Bioanalyzer. Libraries were prepared and sequenced as paired-end, 43 base pair reads on the Illumina Nextseq 500.

Genome analysis

To calculate coverage of genome for male and female, first, *S. carpocapsae* genome was indexed with bowtie (version 1.0.0) and DNA reads were mapped to the genome with the following options: -X 1500 -a -v 2 -best-strata -S (Langmead *et al.* 2009). Bam files were used to calculate coverage using DeepTools2 (version 3.1.1) with default settings for 100 kb (Ramirez *et al.* 2016). Next, gene density was calculated and graphed with GenomicFeatures (version 3.8) and KaryoploteR (version 3.8) packages for R/Bioconductor (Gel and Serra 2017; Lawrence *et al.*, 2013). GC content was calculated with perl script GC-content-in-sliding-window using default settings for 100-kb windows (Richard 2018). Number of repeats were calculated with RepeatModeler (version 1.0.8) and RepeatMasker (version 4.0.7) using default settings (Smit, Hubley, & Green, n.d.).

Orthology analysis

Orthologs were determined across *S. carpocapsae* and *C. elegans*, and *S. carpocapsae* and *P. pacificus* by blasting the longest protein sequence for each gene using OrthoMCL (version 1.4) with the default settings (Li *et al.* 2003). Protein sequences were downloaded from WormBase Parasite for *C. elegans* (version PRJNA13758) and *P. pacificus* (version PRJNA12644).

Protease identification, location, and phylogeny

Protein sequences for the new *S. carpocapsae* genome were aligned to the MEROPS (Rawlings *et al.* 2018) database using command-line BLAST

(Camacho *et al.* 2009). An additional analysis was done by Neil Rawlings using a custom pipeline to search our genome against the MEROPS collection (Rawlings *et al.* 2018). Matches with e-value greater than 1×10^{-10} were kept. Locations for serine and metalloprotease genes were obtained from GTF files. Plots for gene locations and clusters were done using ggplot2. Sequences for orthologs in *C. elegans* were obtained as mentioned above to generate phylogenetic trees for serine and metalloproteases. Genes were aligned in MEGA7 using MUSCLE and a maximum likelihood tree was done using bootstrap support of 100 replicates.

RNA-sequencing of *S. carpocapsae* time-course, male, and female

To collect RNA for the time-course, *S. carpocapsae* IJs (~10,000) were cultured with *X. nematophila* in a lipid plate for each allotted time: 3, 6, 9, 12, 15, 24, 36, 48 hr, female and male. The nematodes were washed from the plate into a 1.5mL eppendorf tube and subsequently washed 2 times with Ringer's and 3 times with HyPure water. Then, *S. carpocapsae* single nematode transcriptome sequencing was performed according to Serra *et al.* (2018). Briefly, a single nematode was picked, transferred to a PCR tube and cut with a 25-gauge needle. Then, 2 μ l of lysis buffer (for recipe see Serra *et al.*, 2018) which has RNase inhibitor and proteinase K was added to the tube and incubated at 65° for 10min, 85° for 1min. Immediately after, 1 μ l oligo-dT VN primer and 1 μ l dNTP was added to the tube and incubated at 72° for 3min. After the incubation period samples were reverse transcribed and PCR amplified according to Serra *et al.* After PCR, samples were cleaned with Ampure XP beads, concentrations measured with Qubit fluorometer and bio-analyzed with Agilent 2100 Bioanalyzer to check for cDNA quality. If samples had a good BioAnalyzer profile, they were tagged with Nextera DNA Library Prep Kit, amplified with adapters and sequenced.

RNA-seq analysis of time course, embryos, female and male

S. carpocapsae genome was indexed with a GTF file via RSEM (version 1.2.25) using the command `rsem-prepare-reference` (Li and Dewey 2011). Embryo data were downloaded from NCBI GEO datasets according to Macchietto *et al.* (2017) (GSE86381). Unstranded, paired-end 43 bp RNA-seq reads for each worm were mapped using bowtie (version 1.0.0) with the following options: `bowtie -X 1500 -a -m 200 -S -p 8 -seedlen 25 -n 2 -v 3` (Langmead *et al.* 2009). Gene expression was quantified with RSEM (version 1.2.25) with the command `rsem-calculate-expression`. For all analyses, gene expression was reported in Transcripts Per Million (TPM).

The TPM generated by `rsem-calculate-expression` for *S. carpocapsae* samples were normalized according to groups using the R package `limma` because samples were collected, processed and sequenced in different batches (Ritchie *et al.* 2015). For all RNA-seq analysis genes were kept if reached a minimum of 1TPM for three replicates. For heatmap, gene expression was mean-centered, normalized, and hierarchically clustered with Cluster 3.0 and visualized using Java Treeview (de Hoon *et al.* 2004; Saldanha 2004). Differential gene expression was determined using edgeR (Robinson *et al.* 2010). Genes were considered differentially expressed if the FDR < 0.05 and $\log_2(\text{fold change}) > 2$. Gene ontology enrichment analyses was calculated using Blast2GO fisher's exact test and considered statistically significant if FDR < 0.01 (Conesa *et al.* 2005). List of genes used in Blast2GO were differentially expressed according to edgeR or dynamically expressed according to maSigPro.

Data availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession AZBU00000000. The version described

in this paper is version AZBU02000000. RNA-seq reads and gene expression tables were submitted to Gene Expression Omnibus (GEO) under the accession number [GSE127823]. Supplemental material available at FigShare: <https://doi.org/10.25387/g3.8069438>.

RESULTS

S. carpocapsae largest scaffold identified as chromosome X

We used PacBio sequencing along with BioNano optical mapping to generate a *S. carpocapsae* genome assembly consisting of 17 scaffolds with a N50 of 7.36 Mb and GC content of 45.7%. Comparison of our genome to the Breton strain from Rougon-Cardoso *et al.* (2016) allowed us to merge our smallest contig of 150 kb with scaffold 7 for a total of 16 scaffolds. *S. carpocapsae* is known to have 5 chromosomes, and the largest scaffold of 20.9 Mb, could correspond to an assembled chromosome (Table S1). We performed Illumina sequencing of the DNA of males and females, being careful to collect females that had not mated to identify scaffolds that would encompass the chromosome X in our new assembly. The largest scaffold displayed a characteristic twofold difference in coverage between females (orange) and males (blue), which we therefore renamed as Chromosome X (Figure 1A). Scaffold 15, which is much smaller (207 kb or 1% of the size of chromosome X) also showed a similar difference in coverage and is likely a small part of the X chromosome. The remaining scaffolds of ≥ 1 Mb suggest that we have the remaining four chromosomes in 14 pieces at most.

We used our RNA-seq transcriptomes from published studies (Marissa Macchietto *et al.*, 2017) and this study to reannotate the genes on this new assembly and identified 30,957 genes using Augustus (see methods). We then calculated the gene density along the scaffolds and observed a uniform distribution of genes on chromosome X, which is similar to observations in *Caenorhabditis elegans* and *P. pacificus* (Andersen *et al.*, 2012; Rödelsperger *et al.*, 2017). Gene density varied little along and among all other scaffolds (Figure 1B). Chromosome X has a rich %GC content in its center when compared to its arms. All other scaffolds either have uniform %GC content or show a distinct increase or decrease (Figure 1C). Lastly, we analyzed the repeat content of chromosome X and scaffolds. As expected from other assembled nematode chromosomes, chromosome X repeats are more frequent on the arms than in its central region (Hillier *et al.*, 2007; Rödelsperger *et al.*, 2017; Yin *et al.*, 2018). All other scaffolds that are part of autosomal chromosomes either have uniform repetitive sequences (such as scaffold 3) or scaffolds in which repetitive sequence either decrease or increase (such as scaffolds 2 and 4) (Figure 1D). In summary, we assembled the *S. carpocapsae* genome into 16 scaffolds, 14 of which are greater than 1Mb, including a nearly complete chromosome X that shows similar characteristics to other assembled nematode chromosomes.

S. carpocapsae chromosome X is primarily syntenic to *C. elegans* chromosomes II and IV

We performed a macrosyntenic analysis between *S. carpocapsae* and *C. elegans* genomes by identifying the one-to-one orthologs between both species and plotting these genes according to their position in *S. carpocapsae* scaffolds (Figure 2). We found that *S. carpocapsae* chromosome X is primarily homologous to sections of *C. elegans* chromosomes II, X, and IV. *S. carpocapsae* chromosome X has approximately 10 Mb homologous to *C. elegans* chromosome II, 7.5 Mb homologous to *C. elegans* chromosome IV, and 2.5 Mb homologous to *C. elegans* chromosome X. The other scaffolds are largely homologous to only one

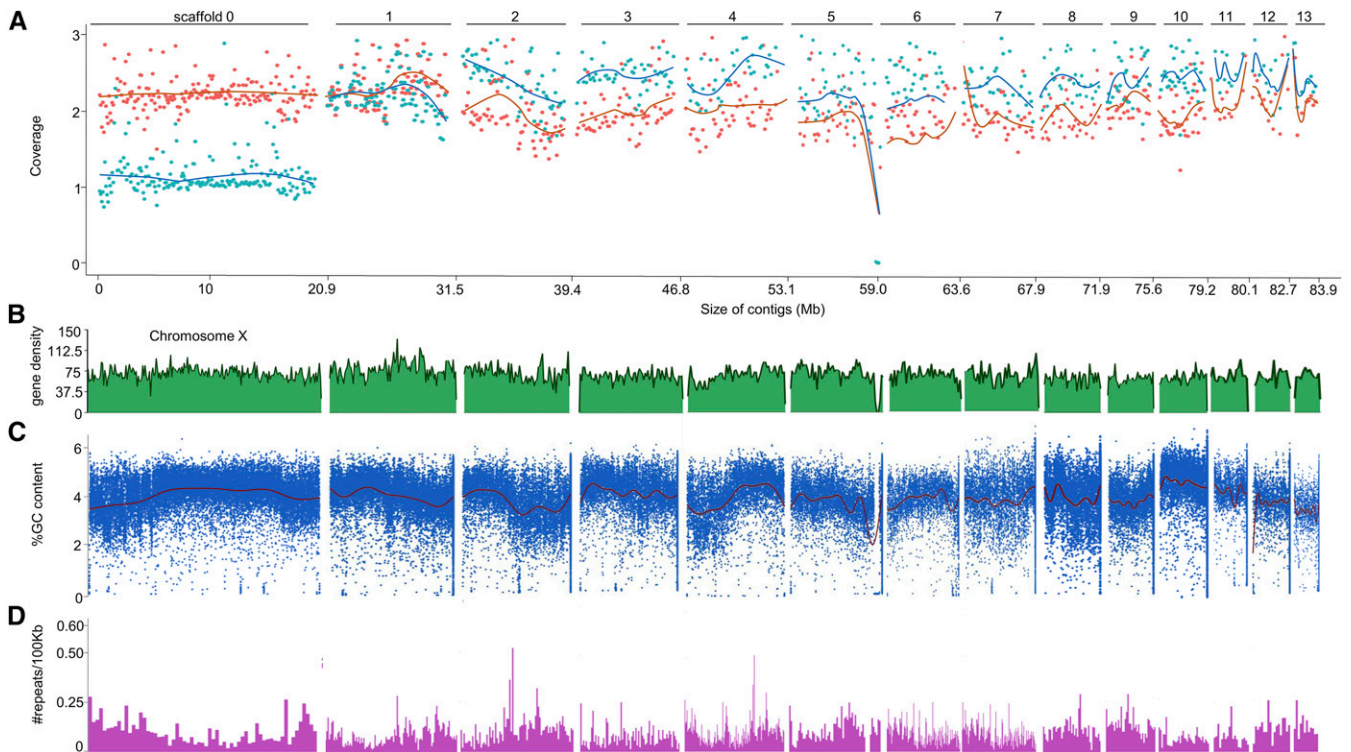


Figure 1 Identifying the X chromosome in *S. carpocapsae*. (A) Illumina sequencing of females (orange) and males (blue) shows coverage in order of longest to shortest for scaffolds >1 Mb, in which scaffold 0 is chromosome X. (B) Gene density calculated for 100 kb intervals. (C) GC content for every 100 kb with a sliding window of 1 kb. (D) Number of repeats per 100 kb.

of the *C. elegans* chromosomes. Scaffolds 1, 9, and 11 are homologous to chromosome I. Scaffold 15 is homologous to chromosome II. Scaffolds 4, 7, 8, 13, and 16 are homologous to chromosome III. Scaffolds 2, 6, 10, and 14 are homologous to chromosome IV. Lastly, scaffolds 3, 5, and 12 are homologous to chromosome X.

Next, we investigated whether *S. carpocapsae* represents the ancestral state by repeating the macrosyntentic analysis with *Pristionchus pacificus* (Figure S2). The *P. pacificus* chromosome configuration has been reported to represent an ancestral state while *C. elegans* and stronglylid nematodes represent the derived state (Rödelsperger *et al.*, 2017). *S. carpocapsae* chromosome X is primarily syntenic to *P. pacificus* chromosomes II, X and IV. The other scaffolds are largely homologous to only one of *P. pacificus*' chromosomes. Scaffold 15 is homologous to chromosome II. Scaffolds 2, 3, 5, 6, 7, 10, 12, and 14 are homologous to chromosome I. Scaffolds 4, 8, and 13 are homologous to chromosome III. Lastly, scaffolds 1, 9 and 11 are homologous to chromosome V. *P. pacificus* chromosomes II, III, IV, X are homologous to the same chromosomes in *C. elegans* (Rödelsperger *et al.*, 2017). However, *P. pacificus* chromosome I is partially homologous to *C. elegans* chromosome V and X. The similar pattern of homologies between chromosome X of *S. carpocapsae* and chromosomes II, X, and V of both *C. elegans* and *P. pacificus* is unsurprising, given that these species' chromosomes are themselves homologous. *S. carpocapsae* chromosome X configuration is likely similar between *C. elegans* and *P. pacificus* because of the homology between these species for chromosomes II, IV, X. *S. carpocapsae* scaffolds do not share homologous regions with chromosome X of *P. pacificus*. However, several scaffolds are homologous with chromosome I (Figure S2). In summary, *S. carpocapsae* chromosome X is homologous to *C. elegans* and *P. pacificus* chromosomes II, IV, and X. Therefore, *S. carpocapsae* have

a derived state and translocation has occurred in the branch leading to the ancestors of *Steinernema*.

Rapid metalloprotease expansion in *S. carpocapsae*

A previous study comparing nematode genomes found an expansion of proteases and protease inhibitors in *Steinernema* (Dillman *et al.*, 2015). *S. carpocapsae* in particular had 654 peptidase genes, approximately one-third of them encoding metalloproteases and one-third being serine proteases (Dillman *et al.*, 2015). It was suggested that the potential function of these genes may be to aid in parasitism (Dillman *et al.*, 2015). Further studies on the mechanisms of *S. carpocapsae* infection identified 472 venom proteins that included proteases and protease inhibitors (Lu *et al.*, 2017). Our new genome allows us to investigate potential molecular mechanisms by which these gene families may be expanding. In addition, an improved assembly allows us to determine the location of these genes and these genes are in fact numerous individual unique genes or misassembled genes.

In our assembly, we identified 228 unique metalloprotease genes and 254 serine protease genes. Mapping the genes to their chromosomal locations revealed aggregation of gene clusters (Figure 3A-B). In order to visualize the location of venom genes, we colored them in red. Venomous metalloprotease genes form a cluster on scaffold 7, while venomous serine protein genes cluster on scaffolds 5, 8 and 14 (Figure 3A-3B). The location of these expanded genes suggest that they are evolving by tandem duplications. One-to-one orthologs to *C. elegans* shows that the 228 metalloprotease genes correspond to 90 *C. elegans* gene models and the 284 serine proteases correspond to 50 genes in *C. elegans*. The phylogenetic trees for these gene families inform us about their potential molecular history and sequence similarity. While many of the metalloproteins have orthologs in *C. elegans* with a few

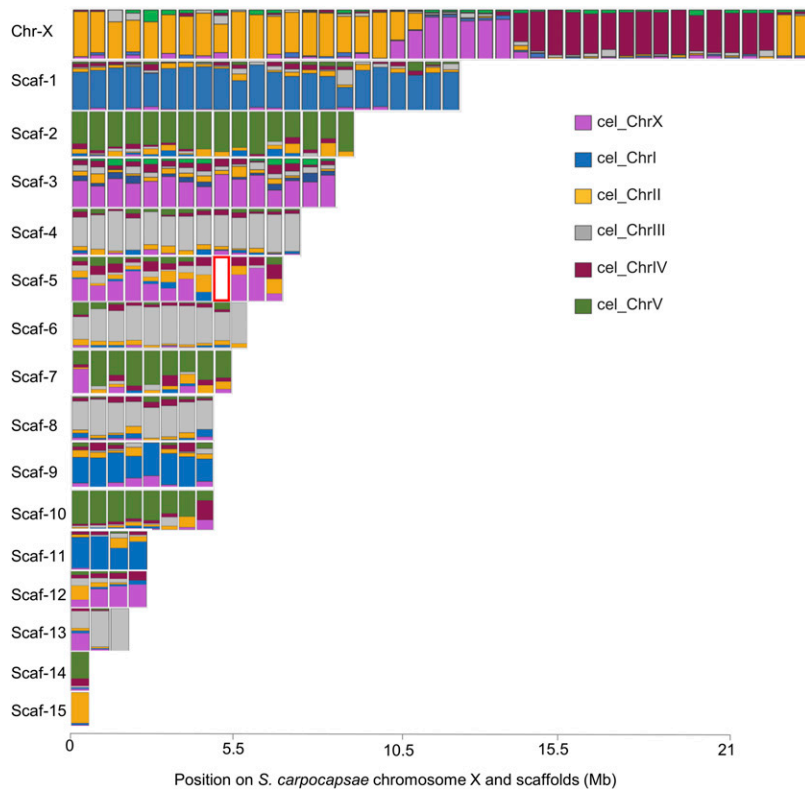


Figure 2 Macrosynteny between *S. carpocapsae* and *C. elegans*. *C. elegans* one-to-one orthologous genes had their position predicted in the *S. carpocapsae* assembly. Each rectangle represents the fraction of *C. elegans* genes present per 500 kb window in *S. carpocapsae*. Red rectangle indicates no synteny.

duplications, serine proteases seem to have undergone an extensive and rapid expansion in *S. carpocapsae*. Interestingly, there is a small grouping of metalloprotease duplicates that belong to the cluster on chromosome 7; these venomous genes all encode an M14A domain (Figure 2C). Similarly, many of the venomous serine proteases group within what we refer to as the “serine expansion”; these genes encode the S01A domain (Figure 3D).

Gene expression profiling of eleven post-activation developmental stages reveals sex-specific differences

We collected early-IJ activation developmental stages of *S. carpocapsae* at 3, 6, 9, 12, and 15 hr as well as later-IJ post-activation time points that are thought to correspond to Juvenile 4 such as 24 (Juvenile 4), 36, and 48 hr, adult females and males using single-nematode RNA-seq analysis (Strauch *et al.* 2010, Serra *et al.*, 2018). We next compared the expression profiles of 11 embryonic stages from Macchietto *et al.* (2017) (zygote, 2-cells, 4-cells, 8-cells, 24-44-cells, 64-78-cells, comma, 1.fivefold, twofold, threefold, L1) and 11 post-IJ developmental stages. We first generated a heatmap of all 22 developmental stages of all genes expressed with a minimum of 1 Transcript Per Million (TPM), where all the genes in red are upregulated and blue are downregulated (Figure 4A). All eleven embryonic stages and five early post-activation stages have distinct gene expression profiles from three late post-activation stages which has a similar gene expression profile to female and male adults. Interestingly, we found that two of the nematodes collected at 24 hr have a transcriptional profile similar to adult females, whereas the third nematode collected at the same hour is more similar to an adult male. Similarly, we found that two of the 36 hr nematodes have a male-like profile and one nematode has a female-like profile. At 48 hr all nematodes have a female transcriptome profile, which we assume simply represents our chance of collecting three out of three female samples randomly.

We then conducted a comparative gene expression analysis between adult males and females. We found 2,328 genes downregulated in males and 1,026 upregulated in females (Figure 4B). GO terms for the 2,328 downregulated genes showed strong enrichment for lipid metabolic process ($P = 1.6 \times 10^{-8}$), catalytic activity ($P = 2.3 \times 10^{-11}$), immune response ($P = 3.2 \times 10^{-4}$) and embryo development ($P = 3.0 \times 10^{-2}$). In contrast, the 1,026 upregulated genes are related to male gamete generation ($P = 8.7 \times 10^{-3}$), sperm motility ($P = 2.0 \times 10^{-2}$), biological regulation ($P = 1.6 \times 10^{-9}$) and catalytic activity ($P = 4.4 \times 10^{-4}$). We generally recovered distinct major GO term categories between female and male except for the category “catalytic activity”, which led us to investigate whether the GO terms for females and males are enriched for distinct enzymes. Such distinct enzymes have been extensively studied in *S. carpocapsae* IJ activation but not in adults (Lu *et al.*, 2017). Male GO terms were enriched for phosphotransferase activity ($P = 5.93 \times 10^{-94}$), kinase activity ($P = 1.14 \times 10^{-23}$), phosphatase activity ($P = 5.69 \times 10^{-9}$) and phosphoric ester hydrolase activity ($P = 1.38 \times 10^{-7}$). In contrast, female GO terms are enriched for peptidase activity ($P = 2.38 \times 10^{-6}$), serine-type peptidase activity ($P = 5.22 \times 10^{-6}$), serine hydrolase activity ($P = 5.22 \times 10^{-6}$) and amine-lyase activity ($P = 8.87 \times 10^{-5}$). This indicates that females and males express different enzymes at higher levels in adulthood. In a separate analysis, we analyzed the genes from Figure 3B with highest expression in male and highest expression in female (Figure S3A). GO terms for highly expressed genes in females and males confirmed that females’ enzymatic activity is related to hydrolase, peptidase and serine activity while males are enriched for kinase activity, and phosphorous processes (Figure S3B).

We then used a Principal Component Analysis (PCA) to assess how gene expression of embryos, late post-developmental stages and adults varies across stages (Figure 4C). Principal component 1 (PC1) accounts for 19.7% of the variance across stages and PC2 accounts for 14.2%.

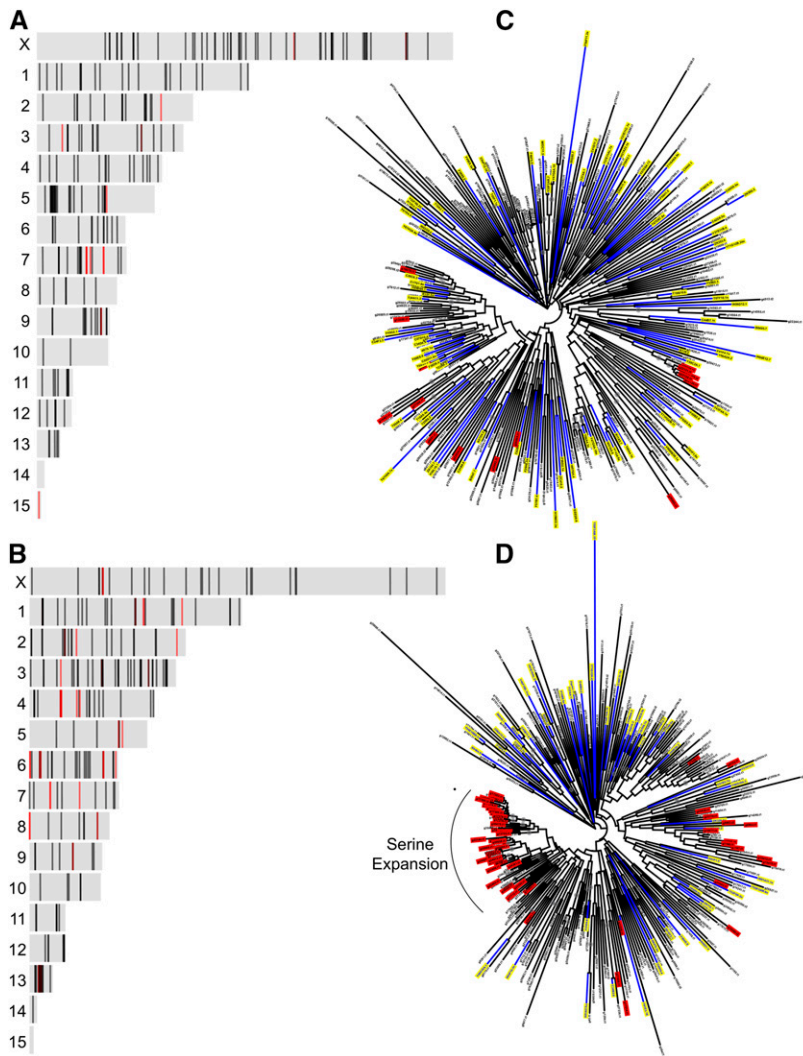


Figure 3 Protease expansion in *S. carpocapsae*. A) Location of metalloprotease genes on *S. carpocapsae* scaffolds. Darker bands represent gene overlap or gene clusters. Red bands represent venom component-encoding genes. B) Location of serine proteases on *S. carpocapsae* scaffolds. Darker bands represent gene overlap or gene clusters. Red bands represent venom component-encoding genes. C) Maximum-likelihood phylogenetic tree of metalloproteases in *S. carpocapsae* and their orthologs in *C. elegans* (highlighted yellow) using a bootstrap support of 100 replicates. Gene names highlighted in red represent genes encoding metalloprotease venom components. D) Maximum-likelihood phylogenetic tree of genes encoding serine protease venom components in *S. carpocapsae* and their orthologs in *C. elegans* (highlighted yellow) using a bootstrap support of 100 replicates. Gene names highlighted in red represent venomous serine protease.

Interestingly, PC3 (10.3%) separated all stages into two clusters. One cluster (blue circle) has all the stages that are developing while the other cluster (pink circle) has all the stages that are fully developed or developmentally arrested (namely infective juveniles (IJs)). GO analysis of PC3 positive values were enriched for catalytic activity, embryo development, testosterone dehydrogenase, female sex differentiation, steroid dehydrogenase, regulation of hormone process and female genitalia development. Meanwhile, PC3 negative values were enriched for kinase activity, ligand-gated ion channel, calcium-release channel activity, nicotinate nucleotide metabolic process, NAD metabolic process and regulation of autophagy (Figure S4). In summary, profiling gene expression in 11 post-activation developmental stages of *S. carpocapsae* shows that starting at 24 hr nematodes are differentiating their gonads and are transcriptionally different from early post-developmental stages. We also found that a gene expression comparison between female and male reveals catalytic activity as differentially expressed with females enriched for peptidase and serine activity while males are enriched for phosphatases. Lastly, PCA separated embryos, late post-developmental stages and adults into two clusters based on their developmental activity.

We investigated conservation of the *C. elegans* sex determination pathway in *S. carpocapsae*. We performed an orthology analysis of the

genes described in Haag (2005) to identify which sex determination genes are conserved in *S. carpocapsae*, including genes with more than one paralog (Table S2). Out of the 27 *C. elegans* sex determination pathway genes, we found 10 that have one to one orthologs such as *sex-1* and *fox-1*, which are responsible for X dosage counting elements and are female-promoting (Hodgkin *et al.*, 1994). Another seven genes have one to many paralogs such as the case of *sdc-1*, the *her-1* transcriptional repressor, which have three homologs; however, as expected, *sdc-2* and *sdc-3* do not have orthologs in *S. carpocapsae* (Klein and Meyer 1993; Lieb *et al.*, 1996; Nonet and Meyer 1991). Interestingly, *mog-4* and *mog-5*, which are global repressors of *fem-3*, are orthologous to the same five genes in *S. carpocapsae* with the best hit for gene g2524 (Puoti & Kimble 2000). Genes *fbf-1* and *fbf-2*, which are germline translational repressors of *fem-3*, are orthologous to one gene g11152 which suggests that this duplication happened along the *C. elegans* lineage (Zhang *et al.*, 1997). Lastly, *fog-1*, which is a promoter of spermatogenesis, has one orthologous gene, contrary to *fog-3* which has orthologs in multiple nematodes species but not in *S. carpocapsae* (Chen *et al.*, 2000; Jin *et al.*, 2001). In summary, we found 19 out of the 27 genes to *C. elegans* in the sex-determination pathway, which might be the core genes important in the nematode sex determination pathway.

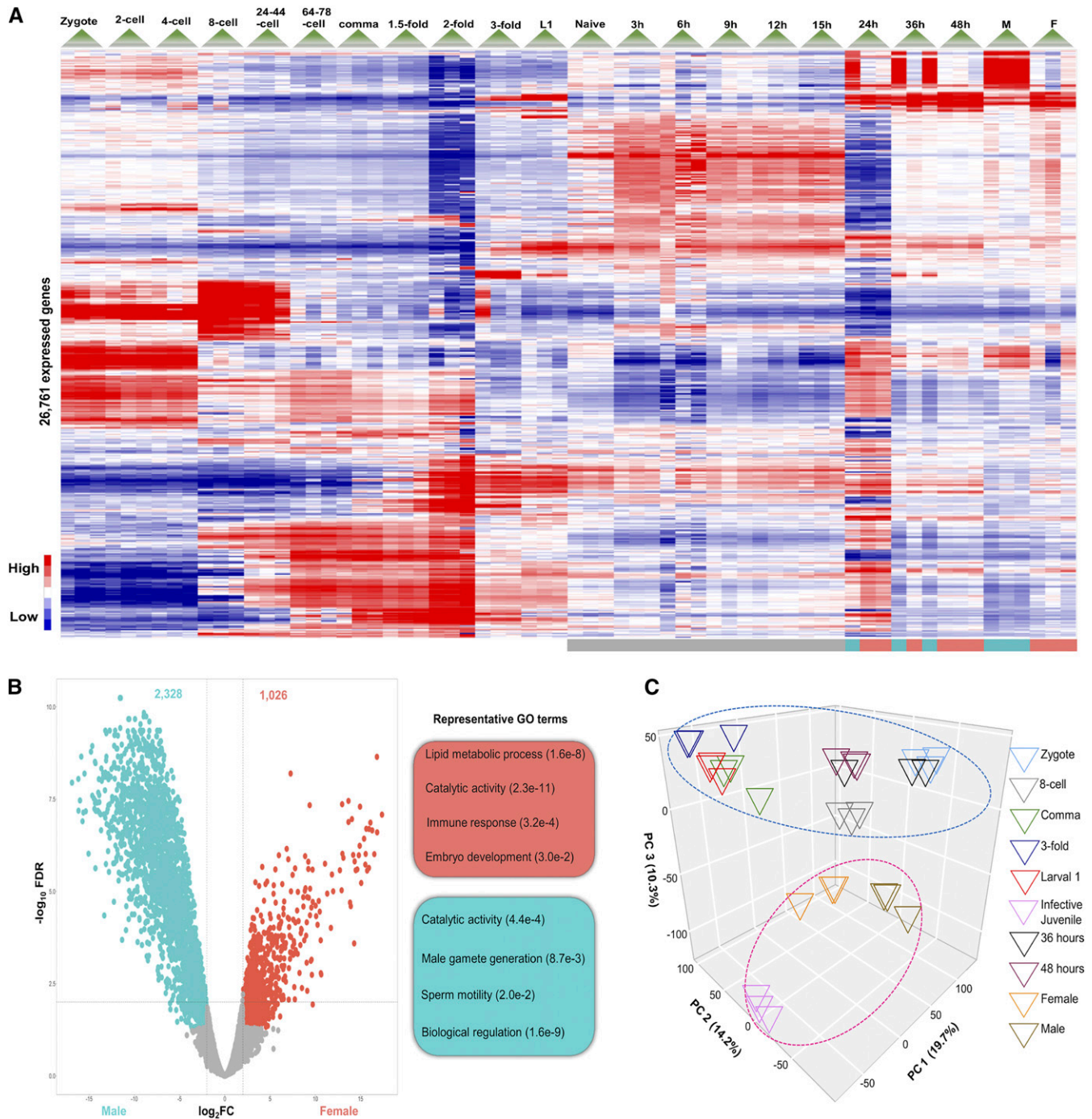


Figure 4 Gene expression profiling of *S. carpocapsae* development in 11 embryonic stages and 11 post-IJ development. (A) Heatmap showing 26,761 genes with minimum of 1 transcript per million (TPM). (B) Volcano plot of 12,461 differentially expressed genes with minimum of 1 TPM, between male (blue) and female (orange) and representative GO terms. (C) Principal Component Analysis of all genes expressed in embryos and during an activation time course. Developmental stages are separated from major life stages (IJ, adults) by PC3.

DISCUSSION

In this study we improved the genome of *S. carpocapsae* ALL strain with PacBio technology and BioNano Iris system and identified chromosome X. The *S. carpocapsae* genome is in 16 scaffolds with an N50 of 7.36 Mb. The sum of the 10 largest scaffolds' lengths achieves ~90% of the genome size, while the top 4 scaffolds cover 50% of the genome. In addition, we used stage-specific developmental transcriptomes to

re-annotate our new assembly and predicted 30,957 genes, which we used to infer patterns of rearrangements between scaffolds based on macrosynteny analysis compared to *C. elegans*. This also allowed us to map the expansion of metalloproteases clusters and to identify the set of genes differentially expressed between males and females.

S. carpocapsae has four autosomes and one sex chromosome (Curran 1989). Chromosome-level macrosynteny revealed that *S. carpocapsae*

chromosome X is orthologous to both *C. elegans* and *P. pacificus* chromosome II and IV. In addition, *S. carpocapsae* chromosome X has a small 3 Mb section orthologous to *P. pacificus* and *C. elegans* chromosome X. The small 3 Mb of *S. carpocapsae* chromosome X section is the only section orthologous to *P. pacificus* chromosome X, while most of the scaffolds are orthologous to *P. pacificus* chromosome I. This small segment of chromosome X orthologous among *S. carpocapsae*, *C. elegans* and *P. pacificus* might represent a Nigon unit, which are proposed to be deeply conserved linkage groups found conserved between nematodes (Tandonnet *et al.*, 2018). A comparison of *A. rhodensis* chromosomes to four *Caenorhabditis* species including *C. elegans* pinpointed these Nigon units, one of which (NX) was found as a sole component of chromosome X or found combined with other Nigon units (Tandonnet *et al.*, 2018). Interestingly, *S. carpocapsae* scaffolds are orthologous to single *C. elegans* chromosomes, which is reminiscent of a macrosyteny analysis in other nematode clades that also found conservation of orthologous chromosomal parts (Fradin *et al.*, 2017). If the last common ancestor among clades had 6 chromosomes, then it is likely that *S. carpocapsae* had a fusion of *C. elegans* chromosomes II and IV, which also occurred in unichromosomal nematode *Diploscapter pachys* in which the order of chromosomes fused were I, X, III, II, IV, and V (Fradin *et al.*, 2017). The conservation of synteny of *S. carpocapsae* to both *C. elegans* and *P. pacificus* highlights the large-scale chromosomal rearrangements in nematode genome evolution. High quality genomes of other *Steinernema* species will allow for comparisons of evolution of parasitism in this genus. Thus, the nematode field would continue to benefit from assembling high quality genomes which would further the study of parasitology.

Differential expression analysis between males and females uncovered an interesting divergence in catalytic activity. Females have high levels of hydrolase activity, serine-peptidase activity, and proteolysis. Similar catalytic activity is found in adult parasitic females (PF) of Strongyloides species (Hunt *et al.*, 2016). Strongyloides species have a free-living generation with gonochoristic females and males, and a parasitic female generation that lives inside the mammalian host (Hunt *et al.*, 2016). Although, in *Steinernema*, there are female sex-bias in emerging IJ populations, little is known about the role of *Steinernema* females in infecting a host populations (Alsaiyah *et al.*, 2009). On the other hand, the catalytic activity for males was enriched for phosphorylation, phosphorous activity and kinase activity. Protein phosphorylation and dephosphorylation regulates protein function with phosphatases important in spermatogenesis and regulation of sperm motility (Cottee *et al.*, 2004; Guillermet-Guibert *et al.*, 2015). Kinases are important in nematode male fertility and a target of drugs to treat human parasites (Guillermet-Guibert *et al.*, 2015; Nolan *et al.*, 2004). The role of females in infecting a host and the molecular biology of reproduction processes in EPNs are largely unexplored. The study of the reproductive and developmental processes in parasitic nematodes will be important because it could lead to better methods for biocontrol and for mammalian parasite control through the disruption or interruption of the reproductive cycle.

In conclusion, we have improved the *S. carpocapsae* genome from 1,578 to 16 scaffolds and identified its X chromosome. A macrosyteny analysis found that chromosome X is orthologous to *C. elegans* and *P. pacificus* chromosomes II and IV. Our results point to a conserved region of chromosome X among the three species. We also found catalytic activity differences between adult females and males. Further analysis will be required to assess the role of female adults in EPN infection and regulatory relationships of male development. Lastly, we believe the improved genome of *S. carpocapsae* will advance the field of comparative nematode genomics and allow for the mining of new insights in the evolution of nematode parasitism.

ACKNOWLEDGMENTS

The authors thank Neil Rawlings for helping with the MEROPS analysis.

Initiative for Maximizing Student Development (IMSD) NIH Grant R25GM05524.

LITERATURE CITED

- Alsaiyah, M. A. M., L. Ebssa, A. Zenner, K. M. O'Callaghan, and C. T. Griffin, 2009 Sex ratios and sex-biased infection behaviour in the entomopathogenic nematode genus *Steinernema*. *Int. J. Parasitol.* 39: 725–734. <https://doi.org/10.1016/j.ijpara.2008.11.003>
- Andersen, E. C., J. P. Gerke, J. A. Shapiro, J. R. Crissman, and R. Ghosh, 2012 Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.* 44: 285–290. <https://doi.org/10.1038/ng.1050>
- Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman *et al.*, 1998 A molecular evolutionary framework for the phylum Nematoda. *Nature* 392: 71–75. <https://doi.org/10.1038/32160>
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
- C. elegans Sequencing Consortium, T. C. elegans S., 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science (New York, N.Y.)*, 282(5396), 2012–2018. <https://doi.org/10.1126/SCIENCE.282.5396.2012>
- Chang, D. Z., L. Serra, D. Lu, A. Mortazavi, and A. R. Dillman, 2019 A core set of venom proteins is released by entomopathogenic nematodes in the genus *Steinernema*. *PLoS Pathog.* 15: e1007626. <https://doi.org/10.1371/journal.ppat.1007626>
- Chakraborty, M., J. G. Baldwin-Brown, A. D. Long, and J. J. Emerson, 2015 A practical guide to de novo genome assembly using long reads. <https://doi.org/10.1101/029306>
- Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon *et al.*, 2005 Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Chen, P.-J., A. Singal, J. Kimble, and R. E. Ellis, 2000 A Novel Member of the Tob Family of Proteins Controls Sexual Fate in *Caenorhabditis elegans* Germ Cells. *Dev. Biol.* 217: 77–90. <https://doi.org/10.1006/dbio.1999.9521>
- Cottee, P. A., A. J. Nisbet, P. R. Boag, M. Larsen, and R. B. Gasser, 2004 Characterization of major sperm protein genes and their expression in *Oesophagostomum dentatum* (Nematoda: Strongyloida). *Parasitology* 129: 179–190. <https://doi.org/10.1017/S003118200400561X>
- Curran, J., 1989 Chromosome numbers of *Steinernema* and *Heterorhabditis* species. *Revue Nématologie* 12: 145–148. Retrieved from <https://core.ac.uk/download/pdf/39865835.pdf>.
- de Hoon, M. J. L., S. Imoto, J. Nolan, and S. Miyano, 2004 Open source clustering software. *Bioinformatics* 20: 1453–1454. <https://doi.org/10.1093/bioinformatics/bth078>
- Deppe, U., E. Schierenberg, T. Cole, C. Krieg, D. Schmitt *et al.*, 1978 Cell lineages of the embryo of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 75(1), 376 LP – 380. Retrieved from <http://www.pnas.org/content/75/1/376.abstract>; <https://doi.org/10.1073/pnas.75.1.376>
- Dillman, A. R., J. M. Chaston, B. J. Adams, T. A. Ciche, H. Goodrich-Blair *et al.*, 2012 An Entomopathogenic Nematode by Any Other Name. *PLoS Pathog.* 8: e1002527. <https://doi.org/10.1371/journal.ppat.1002527>
- Dillman, A. R., M. Macchietto, C. F. Porter, A. Rogers, B. Williams *et al.*, 2015 Comparative genomics of *Steinernema* reveals deeply conserved gene regulatory networks. *Genome Biol.* 16: 200. <https://doi.org/10.1186/s13059-015-0746-6>
- Dito, D. F., D. I. Shapiro-Ilan, C. A. Dunlap, R. W. Behle, and E. E. Lewis, 2016 Enhanced biological control potential of the entomopathogenic nematode, *Steinernema carpocapsae*, applied with a protective gel

- formulation. *Biocontrol Sci. Technol.* 26: 835–848. <https://doi.org/10.1080/09583157.2016.1159659>
- Fradin, H., K. Kiontke, C. Zegar, M. Gutwein, J. Lucas *et al.*, 2017 Genome Architecture and Evolution of a Unichromosomal Asexual Nematode. *Curr. Biol.* 27: 2928–2939.e6. <https://doi.org/10.1016/j.cub.2017.08.038>
- Gang, S. S., and E. A. Hallem, 2016 Mechanisms of host seeking by parasitic nematodes. *Mol. Biochem. Parasitol.* 208: 23–32. <https://doi.org/10.1016/j.molbiopara.2016.05.007>
- Gaugler, R., and H. Kaya, 1990 *Entomopathogenic nematodes in biological control*, edited by Gaugler, R., and H. Kaya. CRC Press, Boca Raton.
- Gel, B., and E. Serra, 2017 karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33: 3088–3090. <https://doi.org/10.1093/bioinformatics/btx346>
- Guillemet-Guibert, J., L. B. Smith, G. Halet, M. A. Whitehead, W. Pearce *et al.*, 2015 Novel Role for p110 β PI 3-Kinase in Male Fertility through Regulation of Androgen Receptor Activity in Sertoli Cells. *PLoS Genet.* 11: e1005304. <https://doi.org/10.1371/journal.pgen.1005304>
- Haag, E., 2005 The evolution of nematode sex determination: *C. elegans* as a reference point for comparative biology. *WormBook*. <https://doi.org/10.1895/wormbook.1.120.1>
- Hillier, L. W., R. D. Miller, S. E. Baird, A. Chinwalla, L. A. Fulton *et al.*, 2007 Comparison of *C. elegans* and *C. briggsae* Genome Sequences Reveals Extensive Conservation of Chromosome Organization and Synteny. *PLoS Biol.* 5: e167. <https://doi.org/10.1371/journal.pbio.0050167>
- Hirao, A., R. U. Ehlers, and O. Strauch, 1999 *Nematology: international journal of fundamental and applied nematological research*. *Nematology* (Vol. 12). Brill. Retrieved from <https://www.cabdirect.org/cabdirect/abstract/20103116226>
- Hodgkin, J., J. D. Zellan, and D. G. Albertson, 1994 Identification of a candidate primary sex determination locus, fox-1, on the X chromosome of *Caenorhabditis elegans*. *Development*, 120(12), 3681 LP – 3689. Retrieved from <http://dev.biologists.org/content/120/12/3681.abstract>
- Huang, S., Z. Chen, G. Huang, T. Yu, P. Yang *et al.*, 2012 HaploMerger: Reconstructing Allelic Relationships for Polymorphic Diploid Genome Assemblies. *Genome Research* 22: 1581–88. <https://doi.org/10.1101/gr.133652.111>
- Hunt, V. L., I. J. Tsai, A. Coghlan, A. J. Reid, N. Holroyd *et al.*, 2016 The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nat. Genet.* 48: 299–307. <https://doi.org/10.1038/ng.3495>
- Jin, S.-W., J. Kimble, and R. E. Ellis, 2001 Regulation of Cell Fate in *Caenorhabditis elegans* by a Novel Cytoplasmic Polyadenylation Element Binding Protein. *Dev. Biol.* 229: 537–553. <https://doi.org/10.1006/dbio.2000.9993>
- Kajitani, R., K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura *et al.*, 2014 Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24: 1384–1395. <https://doi.org/10.1101/gr.170720.113>
- Kaletta, T., and M. O. Hengartner, 2006 Finding function in novel targets: *C. elegans* as a model organism. *Nat. Rev. Drug Discov.* 5: 387–399. <https://doi.org/10.1038/nrd2031>
- Klein, R. D., and B. J. Meyer, 1993 Independent domains of the Sdc-3 protein control sex determination and dosage compensation in *C. elegans*. *Cell* 72: 349–364. [https://doi.org/10.1016/0092-8674\(93\)90113-5](https://doi.org/10.1016/0092-8674(93)90113-5)
- Koren, S., M. C. Schatz, B. P. Walenz, J. Martin, J. T. Howard *et al.*, 2012 Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology* 30: 693–700. <https://doi.org/10.1038/nbt.2280>
- Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, Antonescu *et al.*, 2004 *Open Access Versatile and open software for comparing large genomes* (Vol. 5). Retrieved from <http://www.tigr.org/software/mummer>
- Lam, E. T., A. Hastie, C. Lin, D. Ehrlich, S. K. Das *et al.*, 2012 Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 30: 771–776. <https://doi.org/10.1038/nbt.2303>
- Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson *et al.*, 2013 Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology* 9: e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>
- Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, L., C. J. Stoeckert, and D. S. Roos, 2003 OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13: 2178–2189. <https://doi.org/10.1101/gr.1224503>
- Lieb, J. D., E. E. Capowski, P. Meneely, B. J. Meyer, T. L. Davis *et al.*, 1996 DPY-26, a Link Between Dosage Compensation and Meiotic Chromosome Segregation in the Nematode. *Science* 274: 1732–1736. <https://doi.org/10.1126/science.274.5293.1732>
- Lu, D., M. Macchietto, D. Chang, M. M. Barros, J. Baldwin *et al.*, 2017 Activated entomopathogenic nematode infective juveniles release lethal venom proteins. *PLoS Pathog.* 13: e1006302. <https://doi.org/10.1371/journal.ppat.1006302>
- Lyons, E., and M. Freeling, 2008 How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* 53: 661–673. <https://doi.org/10.1111/j.1365-313X.2007.03326.x>
- Macchietto, M., D. Angdembe, N. Heidarpour, L. Serra, B. Rodriguez *et al.*, 2017 Comparative transcriptomics of steinernema and caenorhabditis single embryos reveals orthologous gene expression convergence during late embryogenesis. *Genome Biol. Evol.* 9: 2681–2696. <https://doi.org/10.1093/gbe/evx195>
- Martens, E. C., and H. Goodrich-Blair, 2005 The Steinernema carpocapsae intestinal vesicle contains a subcellular structure with which *Xenorhabditis nematophila* associates during colonization initiation. *Cell. Microbiol.* 7: 1723–1735. <https://doi.org/10.1111/j.1462-5822.2005.00585.x>
- Morgulis, A., E. M. Gertz, A. A. Schaffer, and R. Agarwala, 2006 “WindowMasker: Window-Based Masker for Sequenced Genomes.” *Bioinformatics* 22: 134–41. <https://doi.org/10.1093/bioinformatics/bti774>
- Morris, R., L. Wilson, M. Sturrock, N. D. Warnock, D. Carrizo *et al.*, 2017 A neuropeptide modulates sensory perception in the entomopathogenic nematode *Steinernema carpocapsae*. *PLoS Pathog.* 13: e1006185. <https://doi.org/10.1371/journal.ppat.1006185>
- Nolan, M. A., D. F. Babcock, G. Wennemuth, W. Brown, K. A. Burton *et al.*, 2004 Sperm-specific protein kinase A catalytic subunit orchestrates cAMP signaling for male fertility. *Proceedings of the National Academy of Sciences of the United States of America*, 101(37), 13483 LP – 13488. <https://doi.org/10.1073/pnas.0405580101>
- Nonet, M. L., and B. J. Meyer, 1991 Early aspects of *Caenorhabditis elegans* sex determination and dosage compensation are regulated by a zinc-finger protein. *Nature* 351: 65–68. <https://doi.org/10.1038/351065a0>
- Puoti, A., and J. Kimble, 2000 The hermaphrodite sperm/oocyte switch requires the *Caenorhabditis elegans* homologs of PRP2 and PRP22. *Proc. Natl. Acad. Sci. USA* 97: 3276–3281. <https://doi.org/10.1073/pnas.97.7.3276>
- Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert *et al.*, 2016 deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* 44: W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Rawlings, N.D., A. J. Barrett, P. D. Thomas, X. Huang, A. Bateman *et al.*, 2018 The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* 46: D624–D632.
- Rhoads, A., and K. F. Au, 2015 PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 13: 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>
- Richard, D., 2018 GC-content-in-sliding-window-. Retrieved from <https://github.com/DamienFr/GC-content-in-sliding-window-git>

- Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law *et al.*, 2015 limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43: e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rödelsperger, C., J. M. Meyer, N. Prabh, C. Lanz, F. Bemm *et al.*, 2017 Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism *Pristionchus pacificus*. *Cell Reports* 21: 834–844. <https://doi.org/10.1016/j.celrep.2017.09.077>
- Rothman, J. H., A. Singson, L. Rose, P. Gönczy, 2014 Polarity establishment, asymmetric division and segregation of fate determinants in early *C. elegans* embryos *. <https://doi.org/10.1895/wormbook.1.30.2>
- Rougon-Cardoso, A., M. Flores-Ponce, H. E. Ramos-Aboites, C. E. Martínez-Guerrero, Y. J. Hao *et al.*, 2016 The genome, transcriptome, and proteome of the nematode *Steinernema carpocapsae*: evolutionary signatures of a pathogenic lifestyle. *Sci. Rep.* 6: 37536. <https://doi.org/10.1038/srep37536>
- Segref, A., J. Cabello, C. Clucas, R. Schnabel, and I. L. Johnstone, 2010 Fate Specification and Tissue-specific Cell Cycle Control of the *Caenorhabditis elegans* Intestine. *Mol. Biol. Cell* 21: 725–738. <https://doi.org/10.1091/mbc.e09-04-0268>
- Saldanha, A. J., 2004 Java Treeview--extensible visualization of microarray data. *Bioinformatics* 20: 3246–3248. <https://doi.org/10.1093/bioinformatics/bth349>
- Serra, L., D. Z. Chang, M. Macchietto, K. Williams, R. Murad *et al.*, 2018 Adapting the Smart-seq2 Protocol for Robust Single Worm RNA-seq. *Bio Protoc.* 8. <https://doi.org/10.21769/BioProtoc.2729>
- Shapiro-Ilan, D. I., I. Hiltbold, and E. E. Lewis, 2017 Nematodes, pp. 415–440 in *Ecology of Invertebrate Diseases*, John Wiley & Sons, Ltd., Chichester, UK., <https://doi.org/10.1002/9781119256106.ch11>
- Sicard, M., N. Le Brun, S. Pages, B. Godelle, N. Boemare *et al.*, 2003 Effect of native *Xenorhabdus* on the fitness of their *Steinernema* hosts: contrasting types of interaction. *Parasitol. Res.* 91: 520–524. <https://doi.org/10.1007/s00436-003-0998-z>
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Srinivasan, J., A. R. Dillman, M. G. Macchietto, L. Heikkinen, M. Lakso *et al.*, 2013 The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics* 193: 1279–1295. <https://doi.org/10.1534/genetics.112.148809>
- Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler, 2008 Using Native and Syntenically Mapped CDNA Alignments to Improve de Novo Gene Finding. *Bioinformatics* 24: 637–644. <https://doi.org/10.1093/bioinformatics/btn013>
- Strauch, O., R. U. Ehlers, and A. Hirao, 2010 Life cycle and population development of the entomopathogenic nematodes *Steinernema carpocapsae* and *S. feltiae* (Nematoda, Rhabditida) in monoxenic liquid culture. *Nematology* 12: 201–210. <https://doi.org/10.1163/156854109X463756>
- Tandonnet, S., G. D. Koutsovoulos, S. Adams, D. Cloarec, M. Parihar *et al.*, 2018 Chromosome-wide evolution and sex determination in a nematode with three sexes. <https://doi.org/10.1101/466961>
- Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 9: e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Yang, Y., Y. Wen, Y. N. Cai, I. Vallée, P. Boireau *et al.*, 2015 Serine proteases of parasitic helminths. *Korean J. Parasitol.* 53: 1–11. <https://doi.org/10.3347/kjp.2015.53.1.1>
- Yin, D., E. M. Schwarz, C. G. Thomas, R. L. Felde, I. F. Korf *et al.*, 2018 Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science* 359: 55–61. <https://doi.org/10.1126/science.aao0827>
- Zhang, B., M. Gallegos, A. Puoti, E. Durkin, S. Fields *et al.*, 1997 A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 390: 477–484. <https://doi.org/10.1038/37297>

Communicating editor: M.-A. Félix