

Hookworm genes encoding intestinal excreted-secreted proteins are transcriptionally upregulated in response to the host's immune system

Erich M. Schwarz^{1,*}, Jason B. Noon², Jeffrey D. Chicca^{2,8}, Carli Garceau^{2,9}, Hanchen Li², Igor Antoshechkin³, Vladislav Ilík^{4,5}, Barbora Pafčo⁵, Amy M. Weeks⁶, E. Jane Homan⁷, Gary R. Ostroff², and Raffi V. Aroian^{2,*}

¹Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, 14853, USA.

²Program in Molecular Medicine, University of Massachusetts Chan Medical School, Worcester, MA, 01605, USA.

³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, 91125, USA.

⁴Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 267/2, 611 37 Brno, Czech Republic.

⁵Institute of Vertebrate Biology, Czech Academy of Sciences, Květná 8, 603 65 Brno, Czech Republic.

⁶Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, 53706, USA.

⁷ioGenetics LLC, 301 South Bedford Street, Ste.1, Madison, WI, 53703, USA.

⁸Current address: Cellular and Molecular Biology Graduate Program, University of Wisconsin, 413 Bock Labs, 1525 Linden Drive, Madison, WI, 53706, USA.

⁹Current address: Leveragen Inc., 17 Briden Street, Worcester, MA, 01605, USA.

*Correspondence: ems394@cornell.edu; raffi.aroian@umassmed.edu.

Abstract

Hookworms are intestinal parasitic nematodes that chronically infect ~500 million people, with reinfection common even after clearance by drugs. How infecting hookworms successfully overcome host protective mechanisms is unclear, but it may involve hookworm proteins that digest host tissues, or counteract the host's immune system, or both. To find such proteins in the zoonotic hookworm *Ancylostoma ceylanicum*, we identified hookworm genes encoding excreted-secreted (ES) proteins, hookworm genes preferentially expressed in the hookworm intestine, and hookworm genes whose transcription is stimulated by the host immune system. We collected ES proteins from adult hookworms harvested from hamsters; mass spectrometry identified 565 *A. ceylanicum* genes encoding ES proteins. We also used RNA-seq to identify *A. ceylanicum* genes expressed both in young adults (12 days post-infection) and in intestinal and non-intestinal tissues dissected from mature adults (19 days post-infection), with hamster hosts that either had normal immune systems or were immunosuppressed by dexamethasone. In adult *A. ceylanicum*, we observed 1,670 and 1,196 genes with intestine- and non-intestine-biased expression, respectively. Comparing hookworm gene activity in normal versus immunosuppressed hosts, we observed almost no changes of gene activity in 12-day young adults or non-intestinal 19-day adult tissues. However, in intestinal 19-day adult tissues, we observed 1,951 positively immunoregulated genes (upregulated at least two-fold in normal hosts versus immunosuppressed hosts), and 137 genes that were negatively immunoregulated. Thus, immunoregulation was observed primarily in mature adult hookworm intestine directly exposed to host blood; it may include hookworm genes activated in response to the host immune system in order to neutralize the host immune system. We observed 153 ES genes showing positive immunoregulation in 19-day adult intestine; of these genes, 69 had ES gene homologs in the closely related hookworm *Ancylostoma caninum*, 24 in the human hookworm *Necator americanus*, and 24 in the more distantly related strongylid parasite *Haemonchus contortus*. Such a mixture of rapidly evolving and conserved genes could comprise virulence factors enabling infection, provide new targets for drugs or vaccines against hookworm, and aid in developing therapies for autoimmune diseases.

Introduction

The hookworms *Necator americanus*, *Ancylostoma duodenale*, and *Ancylostoma ceylanicum* are parasitic nematodes that infect ~500 million human beings, sickening them and lowering their economic productivity (BETHONY *et al.* 2006; TRAUB 2013; PULLAN *et al.* 2014). Hookworms are remarkably long-lived parasites: an adult hookworm can feed off its human host for up to 18 years (BEAVER 1988; GEMS 2000), and infection times of 1-5 years are common (HOAGLAND AND SCHAD 1978). Hookworms achieve

this partly by feeding on the host with enzymes optimized to digest host proteins (RANJIT *et al.* 2009), and partly by suppressing the immune systems of their hosts, which would otherwise kill or expel them quickly (MAIZELS *et al.* 2018). Only one drug, albendazole, is commonly used and effective enough to be useful in mass drug administration against hookworm infections (KEISER AND UTZINGER 2008; LOUKAS *et al.* 2016). However, human hookworms may be developing genetic resistance to this drug (ORR *et al.* 2019; VLAMINCK *et al.* 2019; WALKER *et al.* 2021), as has already happened with dog hookworms (VENKATESAN *et al.* 2023). Moreover, even effective use of albendazole does not prevent endemic hookworms from reinfecting people (JIA *et al.* 2012). An anti-hookworm vaccine would be an ideal way to suppress hookworm infections (COHEN 2016), but no such vaccine exists, in part because we do not know which gene products mediating host-parasite interactions should be targeted as antigens (ZAWAWI AND ELSE 2020).

It is not yet certain how hookworms (or any other parasitic nematodes) dampen the host immune system, but many possible effectors of immunomodulation exist: secreted proteases (RANJIT *et al.* 2009; KNOX 2012; PEARSON *et al.* 2012); secreted protease inhibitors (MANOURY *et al.* 2001; HARTMANN AND LUCIUS 2003; KLOTZ *et al.* 2011); large gene families encoding diverse secreted proteins as antigenic decoys (CANTACESSI *et al.* 2009; TRIBOLET *et al.* 2015); mimics of mammalian immune proteins (LOUKAS *et al.* 2000; YOSHIDA *et al.* 2012); lipid immunomodulators (HARNETT *et al.* 2010; JEX *et al.* 2014; SHINODA *et al.* 2022); and secreted exosomes with anti-host miRNAs (BUCK *et al.* 2014; COAKLEY *et al.* 2017). These mechanisms are by no means exclusive; some or all of them could be working at once. Such complexity may have evolved over 350 million years, when vertebrates first colonized the land and became vulnerable to parasitic nematodes (ANDERSON 1984; CLARK 1994; DURETTE-DESSET *et al.* 1994). Ignorance of how hookworms immunosuppress their hosts makes it challenging to devise vaccines against hookworm disease. It also hinders the possible use of hookworms as sources of new biological reagents against autoimmune diseases (SMALLWOOD *et al.* 2021; RYAN *et al.* 2022), which appear to become more frequent with lower rates of helminth infection (MAIZELS 2020).

Hookworms and other parasitic helminths have long been known to excrete or secrete proteins into their hosts, and these excreted/secreted (ES) proteins have long been hypothesized, and in some cases demonstrated, to enable immunosuppression and virulence (STIREWALT 1963; PRITCHARD 1986; LIGHTOWLERS AND RICKARD 1988; MAIZELS *et al.* 2018; ABUZEID *et al.* 2020). ES proteins can be exported from cephalic/pharyngeal glands, intestine, or cuticle (PRITCHARD 1986; ZHAN *et al.* 2003; HUANG *et al.* 2020). Genes encoding ES proteins have been identified in the major human hookworm *N. americanus* (LOGAN *et al.* 2020), the dog hookworm *Ancylostoma caninum* (MORANTE *et al.* 2017),

and five strongylid parasitic nematodes closely related to hookworm: *Angiostrongylus vasorum* (GILLIS-GERMITSCH *et al.* 2021), *Haemonchus contortus* (WANG *et al.* 2019), *Heligmosomoides bakeri* (HEWITSON *et al.* 2011; MORENO *et al.* 2011; HEWITSON *et al.* 2013), *Nippostrongylus brasiliensis* (SOTILLO *et al.* 2014), and *Ostertagia ostertagi* (PRICE *et al.* 2024) (Figure 1). Until recently (UZOECHE *et al.* 2023), ES proteins have not been identified for the zoonotic human hookworm *A. ceylanicum*, which is an important zoonotic pathogen of humans and companion animals and which can complete its lifecycle in hamsters, making it a key hookworm species for laboratory studies (GARSIDE AND BEHNKE 1989; TRAUB 2013; STRACKE *et al.* 2020; COLELLA *et al.* 2021).

To find *A. ceylanicum* genes whose products may enable parasitism, we have identified *A. ceylanicum* ES proteins, while also using RNA-seq to identify *A. ceylanicum* genes whose products interact with the host either through intestinally-biased expression or through upregulation in response to a functioning host immune system. For the latter set of genes, we hypothesized that the parasite transcriptionally activates genes in response to the host immune system in order to neutralize the host immune response. Previously, we found that changes of *A. ceylanicum* gene activity during infection *in vivo* are much more extensive than changes seen during simulated infection *in vitro* (SCHWARZ *et al.* 2015). Here we build on that observation by correlating ES genes with intestinal and immunoregulated genes. We identify genes encoding immunoregulated intestinal ES proteins that may be important for virulence or immunosuppression, that are new targets for drugs or vaccines against hookworm, and that may aid in the development of therapies against autoimmune diseases.

Methods

Sample procurement, preparation and storage. Cultures, infections, and collections of *A. ceylanicum* followed published methods (HU *et al.* 2012; SCHWARZ *et al.* 2015). For each *A. ceylanicum* infection in support of RNA-seq, we purchased Syrian golden hamsters (*Mesocricetus auratus*) of the HsdHan:AURA strain at 3-4 weeks of age. Hamsters were provided with food and water *ad libitum*. For immunosuppression experiments, we immunosuppressed half of each set of hamsters by injecting them with dexamethasone (3 mg/kg) twice per week throughout the duration of the experiment (TRITTEN *et al.* 2012); the other half were given mock injections. After the first two injections (one week), both immunocompetent and immunocompromised hamsters were infected at approximately four to five weeks of age with *A. ceylanicum*. Infections were allowed to progress for 12 or 19 days, after which we euthanized the hamsters and collected hookworms from small intestines dissected from the hamsters. Dissected hamster intestines were put into Hank's Balanced Salt Solution (pre-warmed to 37°C); worms

were picked quickly from the dissected tissues by hand. Collected worms were snap-frozen in liquid nitrogen and stored at -80°C until use. For *A. ceylanicum* infections in support of ES protein mass spectrometry, we followed a similar protocol but without injections, with two collections of *A. ceylanicum* at 20 days after infection, and without snap-freezing of dissected hookworms. Stages of *A. ceylanicum* selected for RNA-seq or proteomics are based on previously described stages of growth in golden hamsters (RAY *et al.* 1972). For hookworm intestinal-specific studies, from each triplicate set of 19-day post-infection hookworms isolated from hamsters as above, we dissected both hookworm intestinal and hookworm non-intestinal tissue, and extracted RNA from the dissected tissues; for the triplicates of smaller 12-day post-infection young adult worms, such dissection was not possible, so we extracted RNA from whole 12-day young adult worms.

All animal experiments were carried out under protocols approved by the University of Massachusetts Chan Medical School. All housing and care of laboratory animals used in this study conformed to the NIH Guide for the Care and Use of Laboratory Animals in Research (18-F22) and all requirements and all regulations issued by the USDA, including regulations implementing the Animal Welfare Act (P.L. 89-544) as amended (18-F23).

RNA harvesting and sequencing. Total *A. ceylanicum* RNA was extracted from either whole worms or from dissected tissues as in Romeo and Lemaitre (ROMEO AND LEMAITRE 2008; SCHWARZ *et al.* 2015). RNA-seq was done largely as in Srinivasan *et al.* (SRINIVASAN *et al.* 2013). RNA-seq libraries were built with Illumina's TruSeq RNA Sample Prep Kit v2 executed according to manufacturer's instructions, using 1 μg of total RNA for each sample. RNA-seq libraries were sequenced in single-end mode with read lengths of 50 nt (for all 19-day data) or 100 nt (for all 12-day data). Newly generated *A. ceylanicum* RNA-seq libraries are listed in Supplementary Table S1.

Protein harvesting. After 20 days of infection, *A. ceylanicum* hookworms were isolated from their hamster hosts and cultured in liquid culture medium for up to 3 days at 37°C in 5% CO_2 . Liquid culture medium consisted of RPMI Medium 1640 (Gibco, Cat#11835-030) supplemented with 25 mM pH 7.2 HEPES buffer, 10 $\mu\text{g}/\text{ml}$ amphotericin B (Gibco, Cat# 15290-026), and 100 U/ml penicillin/streptomycin (Gibco, Cat# 1570-063), with medium sterilized with a 0.22 μm filter. Fetal bovine serum was omitted from the medium to eliminate any added protein when quantifying ES proteins by BCA assay and to eliminate contamination for proteomics. After the first 24 hours, proteins excreted or secreted from the hookworms were collected by aspirating media only. The aspirated media were centrifuged to remove debris (1500 gs for 20 minutes, at 4°C); the resulting supernatant was concentrated 10-fold using a 3kD Amicon ultra-centrifugal filter (4000 gs for 45 minutes, at 4°C). After this first 10-fold concentration, PBS

was added to match the initial volume, after which the media were reconcentrated by again being centrifuged (4000 gs for 45 minutes, at 4°C); we repeated this twice, for a total of three PBS rinses and reconcentrations. After the third PBS rinse/reconcentration, total protein was quantified using a Pierce BCA Protein Assay kit. Purified ES proteins were then frozen and stored at -80°C.

General computation. Where possible, we used mamba to install and run version-controlled software environments from bioconda (GRUNING *et al.* 2018). For reformatting or parsing of computational results, we used Perl scripts either developed for general use or custom-coded for a given analysis. All such Perl scripts (named below with italics and the suffix ".pl") were archived on GitHub (https://github.com/schwarzem/ems_perl). Internet sources (URLs) for other software are listed in Supplementary Table S2.

Nematode genomes, transcriptomes, coding sequences, and proteomes. For transcriptomic or proteomic analyses, published genome sequences, coding sequences, proteomes, and gene annotations of relevant nematodes were downloaded from WormBase (DAVIS *et al.* 2022) or ParaSite (HOWE *et al.* 2017; LEE *et al.* 2017) (Supplementary Table S3). Published RNA-seq data of *A. ceylanicum* (SCHWARZ *et al.* 2015; WEI *et al.* 2016; BERNOT *et al.* 2020) and *H. contortus* (LAING *et al.* 2013) were downloaded from the European Nucleotide Archive (Supplementary Table S4). Alternative gene predictions for *A. ceylanicum* recently published by Uzoечи *et al.* (UZOECHI *et al.* 2023) were obtained as a Generic Feature Format Version 3 (GFF3) annotation file (<https://github.com/The-Sequence-Ontology/Specifications/blob/master/gff3.md>) from Young-Jun Choi <choi.y@wustl.edu> that we have archived at <https://osf.io/dxfsb>. We extracted protein sequences from this GFF3 via gffread 0.12.7 (PERTEA AND PERTEA 2020) with the arguments '-g [input genome sequence FASTA file] -o /dev/null -C --sort-alpha --keep-genes -P -V -H -l 93 -y [output protein FASTA file] [input gene prediction GFF3 file]'.

Heligmosomoides species nomenclature. Recent genomic analysis has shown that the gastrointestinal parasitic nematode *Heligmosomoides* has two distinct species: *H. bakeri* and *H. polygyrus* (STEVENS *et al.* 2023). Although laboratory strains of *Heligmosomoides* have often been described as *H. polygyrus* (MAIZELS *et al.* 2011), it now appears likely that many or all of these strains have actually been *H. bakeri*. Thus, following previous suggestions for revised nomenclature (BEHNKE AND HARRIS 2010), we refer exclusively to *H. bakeri* even when citing published work that was nominally done with *H. polygyrus*.

RNA-seq subsampling. We found that published *A. ceylanicum* RNA-seq data (Supplementary Table S4) were too extensive for gene predictions by BRAKER2 because of memory limitations. To make these data usable by BRAKER2, we selected a representative subset of them with khmer (ZHANG *et al.*

2014; CRUSOE *et al.* 2015). Because khmer requires "#/1" and "#/2" suffixes for paired-end reads, we retrofitted paired-end RNA-seq reads lacking such suffixes with *retroname_fastq_reads.pl*. We ran *normalize-by-median.py* from khmer on all data (paired and unpaired) twice, with the arguments '-k 31 -C 100 -M 100G', and '-k 31 -C 30 -M 100G'; we then ran *filter-abund.py* from khmer on paired-end data with the arguments '--variable-coverage -C 2 [k-mer hash] [paired-end reads]'. We sorted khmer-filtered data into interleaved paired- and unpaired-end read files with *paired_vs_unpaired_fastq_or_a.pl* with the arguments '--r1 "#0\1" --r2 "#0\2"'.

Reprediction of protein-coding genes. To repredict protein-coding genes in *A. ceylanicum*, we ran *braker.pl* in BRAKER2 2.1.6 (BRŮNA *et al.* 2021) on our published repeat-softmasked *A. ceylanicum* genome assembly (SCHWARZ *et al.* 2015), with the arguments '--genome [genome assembly FASTA] --prot_seq [collected proteomes FASTA] --bam [sorted mapped khmer-filtered paired-end read BAM alignment] --etpmode --softmasking --cores 48 --gff3'. BRAKER2 requires an input genome sequence to have its FASTA header lines previously stripped of comments; we did this with *uncomment_FASTA_headers.pl*. Running BRAKER2 also required us to install: AUGUSTUS 3.4.0 (STANKE *et al.* 2008); BamTools 2.5.2 (BARNETT *et al.* 2011); cdbfasta 0.99 (PERTEA *et al.* 2003); DIAMOND 2.0.9 (BUCHFINK *et al.* 2021); GeneMark-ES/ET 4.68 (LOMSADZE *et al.* 2014); and SAMtools 1.12 (DANECEK *et al.* 2021). To guide BRAKER2 gene predictions, we used our khmer-subsampled paired-end subset of published *A. ceylanicum* RNA-seq data, along with predicted proteomes from the related nematodes *Caenorhabditis elegans* (DAVIS *et al.* 2022), *H. contortus* (DOYLE *et al.* 2020), and *N. americanus* (LOGAN *et al.* 2020). After running BRAKER2, we renamed gene, transcript, and exon names in the GFF3 prediction file with a modified version of *updateBRAKERGff.py* (<https://github.com/Gaius-Augustus/BRAKER/issues/416>) and Perl one-line commands. To extract coding DNAs (CDS DNAs) and protein sequences from the renamed GFF3, we used *gffread* 0.12.7 (PERTEA AND PERTEA 2020) as above, with the additional argument '-x [output CDS DNA file]'.

This gave us a new gene set ("v2.0") that was generally superior (in completeness as assayed by BUSCO, and nonfragmentation as assayed by *count_fasta_residues.pl*) to our original 2015 gene set ("v1.0"). However, we observed that some v1.0 genes encoded ES proteins (as determined by mass spectrophotometric mapping) but had no genomic overlap with v2.0 genes. To ensure no valid ES genes would be overlooked, we used BEDtools 2.30.0 (QUINLAN AND HALL 2010) to identify v1.0 genes whose protein-coding exon sequences (CDSes) had no genomic overlap with v2.0 CDSes, as follows. We used *get_gff3_gene_subset.pl* to remove GFF3 v2.0 annotations that could not be translated by *gffread*, and then used Perl to extract CDS annotation lines from the GFF3 files of published v1.0 and *gffread*-

translatable v2.0 gene predictions. We reformatted the CDS-subset files from GFF3 to BED with *gff2bed* in BEDOPS 2.4.41 (NEPH *et al.* 2012). We identified overlapping coordinates of the CDS-only BED annotation files for v1.0 and v2.0 genes with *intersect* from BEDtools 2.30.0 (QUINLAN AND HALL 2010), using the argument '*-loj*', and used Perl to extract a list of v1.0 genes having no CDS overlaps with v2.0 genes. Given this list of non-overlapping v1.0 genes, we used *extract_fasta_subset.pl* to extract their encoded CDS DNAs and proteins as subsets from the full published v1.0 CDS DNA and protein sets, and added these CDS DNA and protein subsets to our previous generated v2.0 CDS DNA and protein sequences. This gave us our final hybrid prediction set ("v2.1") of CDS DNAs and proteins on which all further analyses in this paper were performed. The v2.1 CDS DNAs and proteome have been archived at <https://osf.io/dxfsb>.

We generated a v2.1 GFF3 as follows. From the published v1.0 GFF3, we extracted GFF3 annotations for these non-overlapping v1.0 genes first by running *extract_parasite_GFF3_subset.pl* and then by selecting '*WormBase_imported*' annotation lines with Perl. We reformatted the resulting v1.0 non-overlapping GFF3 with AGAT 1.1.0 (DAINAT 2023). Likewise, we reformatted the gffread-translatable v2.0 GFF3 with AGAT. We then merged the two AGAT-reformatted GFF3s to produce a single v2.1 GFF3 with uniform formatting, which we have archived at <https://osf.io/dxfsb>.

Assessing quality of protein-coding genes. We determined general statistics of protein-coding gene products with *count_fasta_residues.pl* using the arguments '*-e -t prot*'; we obtained gene counts from maximum-isoform proteome subsets generated with *get_largest_isoforms.pl* using the arguments '*-t parasite*' or '*-t maker*'. We determined and compared the completeness of predicted *A. ceylanicum* protein-coding gene sets with BUSCO 5.2.2 using the arguments '*--lineage_dataset nematoda_odb10 --mode proteins*' (WATERHOUSE *et al.* 2018), which tested a given proteome against 3,131 highly conserved single-gene orthologs in nematodes.

Gene reannotation. For the protein products of our v2.1 *A. ceylanicum* gene set, we predicted both N-terminal signal sequences and transmembrane alpha-helical anchors with Phobius 1.01 (KÄLL *et al.* 2004), reformatting results with *tabulate_phobius_hits.pl*. We predicted coiled-coil domains with Ncoils 2002.08.22 (LUPAS 1996), reformatting results with *tabulate_ncoils_x.fa.pl*. We predicted low-complexity regions with PSEG 1999.06.10 (WOOTTON 1994) using the argument '*-l*' and reformatting results with *summarize_psegs.pl*. We identified protein domains with Pfam 35.0 (MISTRY *et al.* 2021) database with *hmmscan* in HMMER 3.3.2 (EDDY 2009; FINN *et al.* 2016), using the arguments '*--cut_ga*' to impose family-specific significance thresholds, '*-o /dev/null*' to discard text outputs, and '*--tblout*' to export tabular outputs; Pfam results were reformatted with *pfam_hmmscan2annot.pl*. We also identified protein domains

with *interproscan.sh* in InterProScan 5.57-90.0 (PAYSAN-LAFOSSE *et al.* 2023) using the arguments '*-dp -iprlookup -goterms*', and reformatting results with *tabulate_iprscan_tsv.pl*. We generated Gene Ontology (GO) terms (ASHBURNER *et al.* 2000; CARBON *et al.* 2021), EggNOG descriptions (HERNANDEZ-PLAZA *et al.* 2023), and KEGG codes (KANEHISA *et al.* 2023) with EnTAP 0.10.7-beta (HART *et al.* 2020) using the argument '*--runP*', and using selected UniProt (UNIPROT 2023) and RefSeq (O'LEARY *et al.* 2016) proteome databases from highly GO-annotated model organisms (Supplementary Table S3); proteome databases were generated with *madekb* from DIAMOND 0.9.9 (BUCHFINK *et al.* 2021), itself bundled with EnTAP; annotations from EnTAP were reformatted from protein to gene annotation tables with *cds2gene_EnTAP_annot.pl* and *cds2gene_annot.pl*. We identified genes encoding possible antimicrobial peptides (AMP) by mapping predictions by Irvine *et al.* (IRVINE *et al.* 2023) from v1.0 to v2.1 of our gene predictions. We identified orthologies between *A. ceylanicum* and other nematodes with OrthoFinder 2.5.4 (EMMS AND KELLY 2019) using the arguments '*-S diamond_ultra_sens -og*'; results were reformatted with *prot2gene_ofind.pl* and *genes2omcls.pl*. For all analyses except OrthoFinder, full proteomes were used; for OrthoFinder, we used maximum-isoform proteome subsets generated with *get_largest_isoforms.pl*. An overall annotation table was constructed from both these and RNA-seq annotations (below) with *add_tab_annot.pl*.

Gene annotation for related parasitic nematodes. To make consistent comparisons of ES genes (or other categories of genes) between *A. ceylanicum* and other species, we annotated the protein-coding genes for three related parasites (*A. caninum*, *H. contortus*, and *N. americanus*) using the same methods as for *A. ceylanicum* above; the analyzed proteomes are listed in Supplementary Table S3.

Selection of ES gene sets from related parasitic nematodes for comparative analysis. We selected three published ES gene sets from *A. caninum*, *H. contortus*, and *N. americanus* for comparative analysis by the following criteria. They had to have either gene identification numbers (IDs) from public genome assemblies, or, if they did not have gene IDs from valid genomes, they needed to at least have expressed sequence tag (EST) IDs for which the original sequence data were publicly available so that they could be mapped onto genomes (and thus to modern genes with proper IDs). This requirement disqualified previously published ES gene sets for *H. bakeri* and *N. brasiliensis*. Second, the gene IDs needed to be correctly mapped from protein mass spectrometric data of one species onto genes of the same species. Remarkably, this requirement disqualified the ES genes from *Angiostrongylus*, which were generated from *A. vasorum* protein mass spectrometry data but were mapped onto genes of *A. cantonensis* and *A. costaricensis*. Third, the ES genes needed to be from parasitic nematodes that had evolved parasitism in common with *A. ceylanicum* hookworms. This criterion disqualified ES genes from the giant

roundworm *Ascaris suum* or the whipworm *Trichuris muris*, which evolved parasitism independently of hookworms and other strongylids.

RNA-seq data. For *A. ceylanicum*, we generated biologically triplicated RNA-seq data for each biological condition. We also analyzed published RNA-seq data for *A. ceylanicum* (SCHWARZ *et al.* 2015; WEI *et al.* 2016; BERNOT *et al.* 2020) and *H. contortus* (LAING *et al.* 2013), listed in Supplementary Table S4. Before analysis, we Chastity-filtered new *A. ceylanicum* RNA-seq reads with *quality_trim_fastq.pl* using the arguments '-q 33 -m 50'. We then quality-filtered and adaptor-trimmed new *A. ceylanicum* RNA-seq reads with *fastp* 0.20.0 (CHEN *et al.* 2018) using the arguments '--dont_overwrite --detect_adapter_for_pe --n_base_limit 0 --length_required X', with X = 50 for new *A. ceylanicum* data. We quality-filtered and adaptor-trimmed previously published *A. ceylanicum* and *H. contortus* RNA-seq reads with *fastp* 0.20.0 using the arguments '--dont_overwrite --n_base_limit 0 --length_required X' with X = 50 for *A. ceylanicum*.

RNA-seq expression values and significances. We quantified gene expression from RNA-seq data sets with Salmon 1.9.0, generating expression values in transcripts per million (TPM) and estimating mapped read counts per gene (PATRO *et al.* 2017). To prevent spurious mappings of RNA-seq reads, we used full selective alignment to a "gentrome" (a CDS DNA set, treated as a target for real mappings, combined with its genome, treated as a decoy for spurious mappings), followed by quantification using Salmon in non-alignment mode (SRIVASTAVA *et al.* 2020; SRIVASTAVA *et al.* 2021). For Salmon's *index* program, we used the arguments '--no-version-check --keepDuplicates -t [gentrome sequence] -d [decoy list]'; for Salmon's *quant* program, we used the arguments '--no-version-check --seqBias --gcBias --posBias --libType A --geneMap [transcript-to-gene table]', with '--unmatedReads' used for single-end data. Results from *quant.genes.sf* output files were reformatted with *make_salmon_TPM_slices.pl*.

Heatmapping of RNA-seq data. To visualize and cluster gene expression values for RNA-seq replicates, we converted expression values of 0 TPM to empirical pseudozeros with *assort_tpms.pl*, with each replicate's empirical pseudozero being defined as the lowest non-zero TPM expression value observed for all genes within that replicate. This allowed logarithmic transformation of zero expression values while defining these values in a replicate-specific way. We then converted the expression values of replicates to $\log_{10}(\text{TPM})$ scores with *log10_tsv_numbers.pl*. Finally, we heatmapped and dendrogram-clustered all replicates with ComplexHeatmap 2.18.0 (GU 2022) using default parameters. Heatmapping showed individual non-dexamethasone (nonDEX) and dexamethasone (DEX) intestinal RNA-seq replicates that anomalously clustered with two DEX and nonDEX intestinal RNA-seq replicates,

respectively (Supplementary Figure 1); these anomalous replicates were thus not used for differential gene expression analysis.

Differential gene expression analysis. Having generated RNA-seq readcounts, we used the exactTest function of edgeR 3.36.0 (ROBINSON *et al.* 2010) to compute log₂ fold-changes (log₂FC) and false discovery rate (FDR) significance values (NOBLE 2009), to identify statistically significant changes of gene expression between pairs of biological conditions in our RNA-seq data. Dispersions were computed with edgeR's estimateDisp function using the argument '*robust=TRUE*'. For this analysis, we used all non-anomalous replicates of our *A. ceylanicum* RNA-seq data (Supplementary Table S1) along with published *A. ceylanicum* RNA-seq data from whole fourth-stage larval (L4) males, whole adult males, whole adult females, and adult male intestine (Supplementary Table S4). To increase the statistical signal for differentially expressed genes, we used *filter_minimum_readcounts.pl* with the argument '*10*' to remove genes which failed to achieve 10 mapped reads in any biological replicate before submitting readcounts to edgeR. An edgeR R script (Supplementary File S1) was generated with *make_edgeR_classic_scripts.pl*, and run in batch mode with R 4.1.3. In summarizing edgeR results, we defined a gene as being differentially expressed (e.g., with tissue-, sex-, or condition-biased expression) if edgeR scored it as being ≥ 2 -fold more strongly or weakly expressed in one condition than another (either $\log_2\text{FC} \geq 1$, or $\log_2\text{FC} \leq -1$) with an FDR of ≤ 0.01 . All other genes with intermediate expression ratios between two conditions ($-1 < \log_2\text{FC} < 1$), or with expression ratios that were determined with lower significance (FDR > 0.01) were not classified as having differential gene expression. We used edgeR's exactTest with these criteria because it is well-adapted for analyzing small numbers of biological replicates (SCHURCH *et al.* 2016).

RNA-seq and differential gene expression analysis for H. contortus. These were performed with published RNA-seq data for dissected adult female intestine and whole adult female bodies (Supplementary Table S4) (LAING *et al.* 2013) and with predicted genes from the published chromosomal reference genome (Supplementary Table S3) (DOYLE *et al.* 2020), by the same methods as used for *A. ceylanicum*.

Constructing a protein set for mass spectrophotometric analysis. To combine our newly predicted *A. ceylanicum* proteins with alternative predictions by Coghlan *et al.* (COGHLAN *et al.* 2019), we ran cd-hit-2d from CD-HIT 4.8.1 (LI AND GODZIK 2006) with the arguments '*-i [v2.1 predicted proteome] -i2 [alternative published proteome by Coghlan et al.] -o [distinct Coghlan protein sequences] -d 100 -c 1.0 -M 0 -T 1 -l 5 -s 0.0 -aL 0.0 -aS 1.0*'. We then joined the v2.1 proteome with distinct Coghlan sequences to yield a nonredundant *A. ceylanicum* protein set for LC-MS/MS analysis.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Peptides were analyzed with an Orbitrap Exploris 480 hybrid quadrupole-Orbitrap mass spectrometer coupled to an UltiMate 3000 RSLCnano liquid chromatography system (ThermoFisher Scientific). Peptides were loaded onto an Acclaim PepMap RSLC column (75 $\mu\text{m} \times 15\text{ cm}$, 2 μm particle size, 100 \AA pore size, ThermoFisher Scientific) over 15 min in 97% mobile phase A (0.1% formic acid) and 3% mobile phase B (0.1% formic acid, 80% acetonitrile) at 0.5 $\mu\text{L}/\text{min}$. Peptides were eluted at 0.3 $\mu\text{L}/\text{min}$ using a linear gradient from 3% mobile phase B to 50% mobile phase B over 120 min. Peptides were electrosprayed through a nanospray emitter tip connected to the column by applying 2000 V through the ion source's DirectJunction adapter. Full MS scans were performed at a resolution of 60,000 at 200 m/z over a range of 300-1,200 m/z with an AGC target of 300% and the maximum injection time set to 'auto'. The top 20 most abundant precursors with a charge state of 2-6 were selected for MS/MS analysis with an isolation window of 1.4 m/z and a precursor intensity threshold of 5×10^3 . A dynamic exclusion window of 20 s with a precursor mass tolerance of $\pm 10\text{ ppm}$ was used. MS/MS scans were performed using HCD fragmentation using a normalized collision energy of 30% and a resolution of 15,000 with a fixed first mass of 110 m/z. The AGC target was set to 'standard' with a maximum injection time of 22 ms.

Mass spectrometry data analysis. Thermo RAW files were searched against our nonredundant *A. ceylanicum* protein set using the SEQUEST algorithm in Proteome Discoverer 2.4 (Thermo). Data were searched for peptides with tryptic specificity with a maximum of two missed cleavages. The precursor mass tolerance was set at 10 ppm and the fragment mass tolerance was set at 0.02 Da. Search parameters included carbamidomethylation at cysteine (+57.021 Da) as a constant modification and the following dynamic modifications: oxidation at Met (+15.995 Da), acetylation at protein N-termini (+42.011 Da), Met loss at protein N-termini (-131.040 Da), Met loss+acetylation at protein N-termini (-89.030 Da). The Percolator node of Proteome Discoverer was used for PSM validation at a false discovery rate of 1%. Final results for both LC-MS/MS analyses are given in Supplementary Table S5.

Mapping ES mass spectrometry hits onto our gene predictions. Our protein database for analyzing ES data included both our new v2.1 *A. ceylanicum* predictions and any predictions by Coghlan *et al.* that differed by even one amino acid residue from ours. We selected ES protein hits in Supplementary Table S5 whose experimental combined q-value was ≤ 0.01 ; for each ES sample, we extracted their gene names as separate v2.1 and Coghlan genes, yielding four ES gene lists in all.

Because gene prediction in complex eukaryotes remains challenging (HATJE *et al.* 2019) and because the Coghlan predictions were made on a different strain of *A. ceylanicum* (Indian strain, US National Parasite Collection Number 102954) than our laboratory strain used for genome sequencing and

gene prediction (HY135), ES proteins whose mass spectrophotometric data fit a Coghlan prediction exclusively might do so only through slight differences in predicted gene structures or through strain-specific amino acid polymorphisms. To identify such cases where a Coghlan gene was equivalent to one of our v2.1 genes, we mapped the protein-coding exons (CDSes) of Coghlan genes onto our v2.1 gene set, as follows. We downloaded Coghlan gene annotations in GFF3 format and their corresponding *A. ceylanicum* genome assembly from ParaSite (Supplementary Table S3). We mapped the coordinates of Coghlan gene annotations onto our *A. ceylanicum* genome assembly sequence (i.e., lifted them over) with Liftoff 1.6.3 (SHUMATE AND SALZBERG 2020) using the arguments '-copies -polish -cds'. We extracted CDS coordinate lines from the resulting liftover GFF3 file with Perl, and did likewise for the CDS coordinates from the GFF3 of our v2.1 gene prediction set. We converted Coghlan and v2.1 CDS coordinate files from GFF3 to BED format with *gff2bed* in BEDOPS 2.4.41 (NEPH *et al.* 2012). We identified overlapping Coghlan and v2.1 CDSes with *intersect* in BEDtools 2.30.0 (QUINLAN AND HALL 2010), using the argument '-loj'. We extracted columns 10 and 20 from the resulting intersection file with *cut* (<https://www.gnu.org/software/coreutils/cut>) using the argument '-f 10,20', and processed the results with *washu.cds_loj_umass.cds_genemap.pl* using the positional arguments '[CDS-to-gene table] [BEDtools intersect columns 10 and 20 table]'; the CDS-to-gene map file was extracted from the FASTA headers of Coghlan and v2.1 proteomes via Perl.

We then used *annot_es_v01.pl* with our Coghlan-to-v2.1 gene map and our four ES gene lists to produce a unified list of ES-encoding v2.1 genes.

Mapping ES genes of Uzoechi et al. onto our gene predictions. Uzoechi *et al.* have recently identified ES-encoding genes in *A. ceylanicum*, using their own reprediction of our original 2015 gene set (UZOECHI *et al.* 2023). To compare their results to ours, we first mapped their results from ES transcripts to ES genes with *map_tx_to_genes.pl*. We then mapped their predicted gene set onto ours using methods similar to those for the Coghlan gene set. Because their repredictions were performed on our *A. ceylanicum* HY135 genome sequence rather than their *A. ceylanicum* Indian genome (as Coghlan predictions had been), we did not need to use Liftoff to map coordinates from one genome to the other. Subsequent mapping steps (CDS extraction with Perl; GFF to BED conversion with BEDOPS *gff2bed*; overlap determination with BEDtools *intersect*; extraction and summarizing of ES gene results with *annot_es_v02.pl*) were as above.

Statistical significance of overlapping gene sets. To identify significant overlaps between sets of *A. ceylanicum* genes, we used the Perl script *motif_group_fisher.pl*, which computed p-values from two-tailed Fisher tests with the Perl module *Text::NSP::Measures::2D::Fisher::twotailed*; for multiple

hypothesis testing (e.g., comparisons to sets of protein motifs) this Perl script also computed q-values via the *qvalue* program of the MEME 5.4.1 software suite (BAILEY *et al.* 2009).

Results

Improved *A. ceylanicum* gene predictions. Analyzing gene function depends crucially on the quality of gene predictions (HATJE *et al.* 2019; GUIGÓ 2023). Our published gene set for *A. ceylanicum* ("version 1.0" or "v1.0"; Table 1) used the best resources we had at the time (SCHWARZ *et al.* 2015), but both software and RNA-seq/protein data for gene prediction have improved since 2015. We thus repredicted *A. ceylanicum* protein-coding genes with BRAKER2 (BRÚNA *et al.* 2021), whose predictions we guided with both greatly expanded RNA-seq data sets across the *A. ceylanicum* life cycle (SCHWARZ *et al.* 2015; WEI *et al.* 2016; BERNOT *et al.* 2020) and protein sequences from related nematodes (DOYLE *et al.* 2020; LOGAN *et al.* 2020; DAVIS *et al.* 2022). We assessed our gene predictions and compared them to others with BUSCO (WATERHOUSE *et al.* 2018), checking for the presence of 3,131 highly conserved single-copy index genes in nematodes and defining completeness by the percentage of index genes found in a given gene set (Table 1). This gave us an initial repredicted gene set ("version 2.0" or "v2.0") that was more complete and less fragmented than previous *A. ceylanicum* gene predictions by either us or others. Whereas our v1.0 predictions had detected 87.8% of BUSCO index genes, our v2.0 predictions detected 95.1%. The v2.0 set also had fewer protein-coding genes than v1.0 (19,419 versus 36,687) encoding longer proteins (medians of 300 versus 199 residues), implying that gene predictions in v2.0 were less fragmented than in v1.0. However, we observed that v1.0 included genes that encoded ES proteins (see below) yet were absent from the v2.0 gene set. To detect as many valid ES genes as possible, we identified all v1.0 genes whose exons were completely nonoverlapping with exons of v2.0 genes, and combined them with v2.0 to make a hybrid v2.1 set of 33,190 protein-coding genes. This v2.1 set showed a small but detectable increase in BUSCO completeness over v2.0 (95.1% to 95.3%). Recently, Uzoечи *et al.* have also repredicted *A. ceylanicum* genes using methods similar to ours (UZOECHI *et al.* 2023); their new gene set is also substantially better than earlier predictions (94.5% completeness), though less complete than our v2.0 or v2.1 gene sets (Table 1). Having predicted the v2.1 gene set, we annotated its predicted protein products with predicted N-terminal signal sequences (KÄLL *et al.* 2004), conserved protein domains (MISTRY *et al.* 2021; PAYSAN-LAFOSSE *et al.* 2023), orthologies to genes in related nematode species such as *N. americanus* (EMMS AND KELLY 2019), and Gene Ontology (GO) terms describing biological and molecular functions (ASHBURNER *et al.* 2000; CARBON *et al.* 2021); annotations are listed in Supplementary Table S6.

Identifying ES proteins. To collect ES proteins from adult *A. ceylanicum* hookworms, we isolated hookworms from hamster hosts 20 days after infection in two independent experiments, incubated the hookworms in protein-free culture medium for three days, concentrated and purified their supernatants, subjected their proteins to mass spectrometry, and mapped the resulting protein spectra to a nonredundant set of protein sequences. By this means, we identified 565 genes encoding ES proteins observed in at least one ES collection (1.7% of all genes), and 350 genes encoding ES proteins in both independent collections (1.05% of all genes; Table 2; Supplementary Table S6). Uzoечи *et al.* have also identified genes encoding ES proteins from male and female *A. ceylanicum* (UZOECHEI *et al.* 2023); mapping their ES genes onto our v2.1 gene set, we find 860 ES genes from their analysis (2.6% of all genes), of which 430 are identical to ES genes from ours (1.3% of all genes; 76% of our ES genes). This overlap is 29-fold greater than chance (two-tailed Fisher test, p-value = 0), and shows high reproducibility of ES genes in *A. ceylanicum*. Conversely, each ES gene set has unique members; they collectively have 995 genes (Table 2).

Out of our 565 ES genes, we consider two subsets for detailed analysis here: either categories of ES genes for which we can show statistically significantly enhanced overlaps with other gene categories of interest, or individual instances of ES genes for which we have found specific published information in the literature which indicates that an individual ES gene may have a biologically informative function or effect.

A. ceylanicum ES genes encode possible host-parasite interaction proteins. Two-thirds of our ES genes (380) were predicted to encode classically secreted proteins with N-terminal signal sequences, 4.6-fold above genome-wide background ($q = 4.5 \cdot 10^{-181}$); the remaining one-third of ES genes (185) encode proteins that are presumably non-classically secreted (Table 3; Supplementary Table S7) (DIMOU AND NICKEL 2018). This mix of predominant but not universal N-terminal signals also exists in ES proteins from *A. caninum*, *N. americanus*, and *H. contortus* (Supplementary Tables S8-S10) (MORANTE *et al.* 2017; WANG *et al.* 2019; LOGAN *et al.* 2020), and may reflect nonclassical secretion of ES proteins through extracellular vesicles (BALMER AND FASO 2021; WANG *et al.* 2024).

Compared to the whole *A. ceylanicum* proteome, ES genes disproportionately encoded several protein families with plausible roles in host-parasite interactions (Table 3; Supplementary Table S7) (ABUZEID *et al.* 2020). These included five types of proteases (aspartyl, astacin, cysteine, serine, and metallopeptidase) that may digest host tissues and blood (WILLIAMSON *et al.* 2006; RANJIT *et al.* 2009; KNOX 2012; YANG *et al.* 2015; CAFFREY *et al.* 2018), along with three types of protease inhibitors (Kunitz, TIL, and TIMP) that may protect *A. ceylanicum* from native or host proteases; both proteases and protease inhibitors may also counteract host immune responses (CHU *et al.* 2004; KNOX 2007). Five ES genes

encoded glutathione S-transferase, an enzyme thought to detoxify free heme generated during hemoglobin proteolysis and other toxins (MATOUSKOVA *et al.* 2016; ABUZEID *et al.* 2020). Sixteen ES genes encoded *Stichodactyla helianthus* toxin (ShK)-related proteins, which might suppress host T cells (CHHABRA *et al.* 2014; MCNEILLY *et al.* 2017); and nine ES genes encoded C-type lectin proteins, which might mimic mammalian immune proteins (enabling immunosuppression) or dissociate cells (enabling tissue invasion) (LOUKAS *et al.* 2000; HARCUS *et al.* 2009; YOSHIDA *et al.* 2012; SCHWARZ *et al.* 2015; WANG *et al.* 2023).

One quarter of ES genes (140) encoded secreted venom-like allergen proteins with CAP or SCP/TAPS domains (GIBBS *et al.* 2008; CANTACESSI *et al.* 2009); in hookworms, these are called *Ancylostoma* secreted proteins, activation-associated secreted proteins, or ASPs (HAWDON *et al.* 1996; YATSUDA *et al.* 2002; WILBERS *et al.* 2018). One ASP, neutrophil inhibitory factor in *A. caninum* (*Acan-NIF*), has been experimentally shown to block neutrophil adhesion (MOYLE *et al.* 1994; LO *et al.* 1999); another ASP, hookworm platelet inhibitor in *A. caninum* (*Acan-HPI*), has been shown to prevent blood clotting (DEL VALLE *et al.* 2003). We observed ES genes encoding orthologs of both these ASPs (*Acey-NIF-B* and *Acey-HPI*) which may themselves be immunosuppressant and antithrombotic proteins (Supplementary Table S6). However, this leaves the other 138 ES ASP genes with only conjectural functions.

ES genes also encode three other multigene families of secreted proteins that are suspected to be involved in infection, although their molecular roles remain largely uncharacterized: ASP-related (ASPR), transthyretin-like (TTL), and secreted clade V proteins (SCVP). ASPRs were first observed in *A. ceylanicum* as a divergent subfamily of ASPs transcriptionally upregulated immediately after host infection (SCHWARZ *et al.* 2015), and were later observed in ES proteins from *N. americanus* (LOGAN *et al.* 2020). TTL proteins were previously observed as a major component of ES proteins from *H. contortus* (WANG *et al.* 2019) and as multigene families in several parasitic nematodes (HUNT *et al.* 2017; TRITTEN *et al.* 2021; STEVENS *et al.* 2023); one TTL protein of *H. contortus* (*HcTTR*) antagonizes goat IL-4 *in vitro*, and thus may be immunosuppressive *in vivo* (TIAN *et al.* 2019). SCVPs were observed in *A. ceylanicum* as a novel family of secreted proteins upregulated in young hookworm adults that is conserved in *C. elegans* and expanded in hookworms and other strongylids (SCHWARZ *et al.* 2015).

Other ES genes were potentially relevant to infection despite not being from overrepresented families (Table 4). One encodes a macin homolog that may protect hookworms from being infected by bacteria, and thus enable their long-term residence in the human gut (JUNG *et al.* 2012; IRVINE *et al.* 2023). Three ES homologs of *A. caninum* anticoagulant protein (STASSENS *et al.* 1996) could act with *Acey-HPI* to prevent clotting during blood feeding. Seven other ES genes encode possible immunosuppressants: two

ES apyrases and one ES adenosine deaminase could suppress both blood clotting and inflammation by hydrolyzing extracellular ATP (GOUNARIS AND SELKIRK 2005; GUIGUET *et al.* 2016; PALA *et al.* 2023); phospholipase A2, phospholipase D, and fatty acid- and retinol-binding (FAR) ES proteins could suppress lipid signals in immune responses (PARKS *et al.* 2021; PARKS *et al.* 2023); the ES macrophage migration inhibitory factor *Acey*-MIF-1, previously observed to bind the human MIF receptor CD74 and promote monocyte migration, could alter cellular immune responses (CHO *et al.* 2007; SPARKES *et al.* 2017); and an ES deoxyribonuclease II could hydrolyze extracellular DNA nets secreted by neutrophils to trap and kill large pathogens (BOUCHERY *et al.* 2020).

Because ES proteins are exposed to the human host's immune system and are likely to promote infection, ES proteins are promising targets for an anti-hookworm vaccine. Twelve ES genes encoded proteins that have already been shown to give partial protection as vaccines in laboratory mammals, or to be homologous to such proteins (Table 4): the cysteine protease *Acey*-CP-1 (NOON *et al.* 2019); the M13 peptidases *Acey*-MEP-6 and *Acey*-MEP-7 (WISNIEWSKI *et al.* 2013; WISNIEWSKI *et al.* 2016); the glutathione S-transferase *Acey*-GST-1, homologous to *A. caninum*'s *Acan*-GST-1 (XIAO *et al.* 2008) and *N. americanus*' *Na*-GST-1 (ZHAN *et al.* 2010); a homolog of the immunodominant hypodermal *Ac*-16 antigen of *A. caninum* (FUJIWARA *et al.* 2007); two homologs of apyrases in *Teledorsagia circumcincta* and *H. bakeri* (NISBET *et al.* 2019; BERKACHY *et al.* 2021); two homologs of fatty acid- and retinol-binding proteins in *A. ceylanicum* (FAIRFAX *et al.* 2009) and *Onchocerca volvulus* (JENKINS *et al.* 1996); one homolog of *Brugia malayi* calreticulin (YADAV *et al.* 2021); one homolog of *Onchocerca volvulus* fructose-1,6-bisphosphate aldolase (MCCARTHY *et al.* 2002); and one homolog to *A. suum* enolase (CHEN *et al.* 2012). In addition to these strict orthologs, three instances of astacin, ASP, and TTL proteins (all overrepresented among ES proteins) have given partial protection as vaccines: *Ac*-MTP-1 of *A. caninum*, *Acan*-ASP-2 of *A. caninum*, and *Hc*TTR of *H. contortus* (ZHAN *et al.* 2002; BETHONY *et al.* 2005; TIAN *et al.* 2020). These results show the vaccine potential of ES proteins, with hundreds of other ES proteins remaining untested as vaccine subunits.

Over half of *A. ceylanicum* ES genes are functionally conserved in related blood-feeding parasitic nematodes: they do not merely have homologs, but ES gene homologs (Table 5; Supplementary Tables S6 and S11). Out of 565 *A. ceylanicum* ES genes, 269 (47.6%) had ES gene homologs in the closely related hookworm *A. caninum* (MORANTE *et al.* 2017), 21-fold more than expected by chance ($p = 6.82 \cdot 10^{-301}$). For the more distantly related hookworm *N. americanus* (LOGAN *et al.* 2020), 141/565 ES genes (25.0%) had ES gene homologs (18-fold over chance; $p = 1.78 \cdot 10^{-137}$); for the still more distantly related blood-feeding *H. contortus* (WANG *et al.* 2019), 180/565 ES genes (31.9%) had ES gene homologs

(8.9-fold over chance, $p = 4.10 \cdot 10^{-120}$). Considering all three species at once, 350/565 ES genes (61.9%) had some ES homolog (12-fold over chance, $p = 0$). Similarly high conservation was observed for the 860 ES genes described by Uzoechi *et al.* (Supplementary Table S11) (UZOECHI *et al.* 2023). These homologies of *A. ceylanicum* ES genes in other species are predominantly to ES genes themselves (i.e., genes experimentally shown to encode ES proteins in these other species). When ES genes in other species are set aside, the remaining homologies of *A. ceylanicum* ES genes fall to, or even below, background frequencies. For *A. ceylanicum* ES genes without known ES gene homologs, homology to *A. caninum* dropped to 1.6-fold over chance ($p = 1.31 \cdot 10^{-29}$); in *N. americanus*, to 1.1-fold below chance ($p = 0.22$); and in *H. contortus*, to 2.4-fold below chance ($p = 2.05 \cdot 10^{-30}$). Thus, the *A. ceylanicum* ES gene set predominantly encodes ES proteins not only in *A. ceylanicum* itself but also in related parasitic nematodes.

Identification of intestinal and immunoregulated genes. Because the hookworm intestine is an important source of ES proteins, we set out to characterize intestine- and non-intestine-biased expression of hookworm genes. We infected hamsters with *A. ceylanicum* hookworms, allowed infections to proceed for either 12 or 19 days, collected hookworms from hamster small intestines, extracted hookworm RNAs, and performed RNA-seq with 36.9 to 55.8 million reads per biological replicate (Supplementary Table S1). Hookworms at the 12-day infection stage were young adults that were too small for us to dissect further, so RNA-seq was done on whole 12-day hookworms. By 19 days, hookworms had grown into mature adults large enough to dissect, so we separated them into intestinal and non-intestinal tissues before RNA-seq. In addition, we hypothesized that hookworms would respond to the mammalian host's immune system by increasing their expression of proteins that counteract the mammalian immune system. To detect hookworm genes regulated by the state of the host's immune system, 12-day and 19-day infections were performed both with normal hamster hosts (nonDEX) and with hamsters whose immune systems had been suppressed by dexamethasone (DEX). We analyzed levels and changes of gene expression both for our RNA-seq data and for published *A. ceylanicum* RNA-seq data from adult male intestine, male adults, and female adults (Table 6; Figure 2; Supplementary Table S6). We defined a gene as significantly changing its activity between two biological conditions (e.g., intestinal versus non-intestinal 19-day tissues, or intestinal nonDEX versus intestinal DEX) if the gene changed its expression by at least 2-fold ($\log_2FC \leq -1$ or ≥ 1) with a false discovery rate (p-value corrected for multiple hypothesis testing) of no more than 0.01 ($FDR \leq 0.01$).

By these criteria, in 19-day adults we observed 1,670 genes with intestine-biased expression and 1,196 with non-intestine-biased expression, the remaining 30,324 genes being intermediate (Table 6; Figure 2). Of the intestine-biased genes, 723 (43.3%) were homologous to *H. contortus* genes with

intestine-biased expression (4.4-fold above background; $p = 8.18 \cdot 10^{-306}$; Supplemental Table S15). For comparison, when we reanalyzed published RNA-seq data for male intestine versus whole male adults (WEI *et al.* 2016; BERNOT *et al.* 2020) as part of the same computation, we identified 635 genes with intestine-biased expression, of which only 180 were also found in our intestine-biased gene set (Supplementary Table S13). By chance alone, one would expect to observe 32 genes shared by both intestinal sets; the observed overlap was 5.6-fold above background ($p = 9.93 \cdot 10^{-85}$) but limited (only 28% of the 635 genes were shared). The observation of over 700 genes with intestine-biased expression conserved between strongylids supports our larger gene set. The disagreement of these intestine-biased gene sets could have one or more causes: different protocols for dissection and RNA extraction; different background tissues for comparison to intestinal RNA-seq (non-intestinal here, versus whole adults in the previous data); different sexes (mixed-sex here, versus males previously); and different strains of *A. ceylanicum* (HY135 here, versus Indian previously), which could cause different rates of RNA-seq read mapping onto our HY135 reference genome (Supplementary Table S12) (DEGNER *et al.* 2009; REZANSOFF *et al.* 2019; BELL *et al.* 2023).

When comparing *A. ceylanicum* in normal (nonDEX) versus immunosuppressed (DEX) hosts, we detected substantial changes of gene expression in 19-day hookworm intestinal tissues, but almost no changes of gene expression either in 12-day young adults or in 19-day non-intestinal tissues. We observed 1,951 genes upregulated in 19-day nonDEX intestine versus 19-day DEX intestine (5.9% of all genes), along with 137 downregulated genes (0.41%). In contrast, we observed only 32 genes positively or negatively immunoregulated either in 12-day young adults or in 19-day non-intestinal tissues (Table 6; Figure 2). One explanation for this pattern is that only 19-day intestinal tissues are directly exposed to the host's immune system through blood feeding. Although 12-day young adult hookworms inhabit the small intestine, they have only just started feeding on blood; up to that point, they instead probably eat mucosal cells (BANSEMIR AND SUKHDEO 1994; BANSEMIR AND SUKHDEO 2001). By 19 days, hookworm adults have been feeding on blood for up to a week; however, most of their bodily contact with this blood (and, thus, with the host's immune system) is through the lumen of their intestine (although cephalic or pharyngeal glands may interact with their host as well).

Intestine-biased genes encode likely components of food digestion, detoxification, and absorption. Whereas ES genes overwhelmingly encoded proteins with N-terminal signal sequences alone, intestine-biased genes were modestly but significantly enriched for several predicted types of secreted or transmembrane proteins (Table 7; Supplementary Table S13). These included secreted proteins, proteins predicted to have one transmembrane anchor sequence after the N-terminal signal peptide, and multi-

transmembrane proteins. Intestinal genes disproportionately encoded protein families relevant to digesting, detoxifying, and absorbing food (Table 7): aspartyl, cysteine, aminopeptidase, and metallopeptidase proteases (WILLIAMSON *et al.* 2006; RANJIT *et al.* 2009; KNOX 2012; YANG *et al.* 2015; CAFFREY *et al.* 2018); UDP-glucuronosyl or UDP-glucosyl transferases, ecdysteroid kinase-like enzymes, and ABC transporters (MATOUSKOVA *et al.* 2016; SCANLAN *et al.* 2022; RAZA *et al.* 2023); and major facilitator and sugar transporters (CHEN *et al.* 2015; DREW *et al.* 2021). Intestinal genes also disproportionately encoded histones; this raises the question of whether mitosis persists in the adult hookworm intestine, as has been observed or inferred for *A. suum* and *H. bakeri* (ANISIMOV AND TOKMAKOVA 1974; ANISIMOV AND USHEVA 1974; POLLO *et al.* 2024). In addition to these well-studied protein families, another notably overrepresented family was Strongylid L4 proteins (SL4Ps), first observed in *A. ceylanicum* as a novel family of non-classically secreted proteins upregulated in fourth-stage (L4) larvae (SCHWARZ *et al.* 2015). Analyzing previously published *H. contortus* intestinal RNA-seq data (Supplementary Table S4), we observed that SL4P genes were also disproportionately represented among intestine-biased *H. contortus* genes, along with several of the better-characterized protein families implicated in digestion, detoxification, or absorption (e.g., cysteine proteases, UDP-glucuronosyl/glucosyl transferases, and major facilitator transporters; Supplementary Table S14). Moreover, the *C. elegans* genes *numr-1* and *numr-2* encode SL4P proteins that are intestinally expressed (TVERMOES *et al.* 2010). We conclude that SL4Ps have conserved intestinal expression in hookworms and other nematodes, and that (as in *C. elegans*) their function might be to counteract toxic environmental stresses (WU *et al.* 2019; HONG *et al.* 2024).

Immunoregulated genes encode signal transduction, male-associated, host-parasite, and ES proteins. The 1,951 positively immunoregulated intestinal genes were modestly, but significantly, enriched for encoding secreted proteins, proteins with a single transmembrane anchor, or both: in other words, possible secreted or cell-surface proteins (Table 8). Overrepresented protein families included possible signal transduction components: protein kinases, protein phosphatases, and SH2 scaffold proteins (TAYLOR AND KORNEV 2011; DIOP *et al.* 2022; KOKOT AND KOHN 2022). Other overrepresented families included homologs of nematode motile sperm proteins (MSPs) (SMITH 2014), along with two other classes of proteins associated with MSPs: MSP fiber 2 proteins (MFP2s) (GRANT *et al.* 2005) and DUF236 proteins (RÖDELSPERGER *et al.* 2021). Although MSPs are indeed hallmark proteins of nematode sperm cells and are usually assumed to be entirely specific to sperm, there are in fact MSPs expressed in somatic cells and required for their mobility. Two different MSP homologs in *Caenorhabditis elegans* have cell and axonal migration RNAi phenotypes in male linker cells and hermaphroditic neurons, with one of them

being expressed in linker cells (SCHMITZ *et al.* 2007; SCHWARZ *et al.* 2012). Moreover, *msp* genes are expressed in *C. elegans* ADL chemosensory neurons (OW *et al.* 2024), and are evolutionarily retained even in parthenogenetic nematode species lacking sperm (HEGER *et al.* 2010). Thus, nematodes express some *msp* genes somatically, which fits our observation here of *msp*, *mfp2*, and *duf236* gene expression in dissected intestines of adult *A. ceylanicum*. In addition, positively immunoregulated intestinal genes disproportionately encoded protein families with possible roles in host-parasite interactions. These families included: astacin, leishmanolysin, metallopeptidase, and trypsin-like proteases; TIMP protease inhibitors; ShK-related proteins; ASPs and ASPRs; SCVPs; and immunoglobulin domain-containing proteins (Table 8). While most of these families were also enriched among ES genes, leishmanolysins, trypsins, and immunoglobulin domains were uniquely enriched here. Despite being expressed intestinally, only one of these immunoregulated genes had intestine-biased expression; instead, 258 were non-intestine-biased (3.67-fold over chance; $p = 5.51 \cdot 10^{-78}$; Supplementary Table S15).

Genes with types of immunoregulation other than positive intestinal had fewer notable protein families (Supplementary Table S15). The 137 negatively immunoregulated intestinal genes disproportionately encoded cysteine proteases and saposins, with saposins being uniquely enriched in this gene set (36-fold over chance; $q = 9.94 \cdot 10^{-3}$). At least four *C. elegans* saposins have antimicrobial activity *in vitro* (BANYAI AND PATTHY 1998; ROEDER *et al.* 2010; HOECKENDORF AND LEIPPE 2012; HOECKENDORF *et al.* 2012); conversely, one saposin in *A. ceylanicum* has no antibacterial activity but does lyse blood cells *in vitro* (HE *et al.* 2021). Negatively immunoregulated hookworm saposins might have either or both functions. Of the 26 positively immunoregulated non-intestinal genes, 15 encoded ASP proteins with CAP domains (46-fold over chance; $q = 7.61 \cdot 10^{-19}$); 10 had non-intestine-biased expression (10.7-fold over chance; $p = 1.12 \cdot 10^{-8}$). Both gene sets disproportionately encoded secreted proteins; neither set had prominently sex-biased expression.

Immunoregulated genes have significantly male-biased expression. Multicellular parasites such as hookworms are generally studied to understand mechanisms of infection common to both sexes, and indeed we ourselves collected both ES proteins and RNA from mixed-sex populations. However, males of the African tick *Rhipicephalus appendiculatus* excrete immunoglobulin-binding proteins that the males themselves do not seem to need, and that are instead required for coinfecting female ticks to feed on host blood efficiently (WANG *et al.* 1998). This instance of sexual cooperation in a parasite made us wonder whether there existed male-biased or female-biased genes among our immunoregulated gene set. Using published RNA-seq data from *A. ceylanicum* adult males and females (BERNOT *et al.* 2020), we observed that 3,135 and 1,547 of our 33,190 predicted *A. ceylanicum* genes had male- or female-biased expression,

respectively (Table 6). Going on to check for overlaps of these gene sets with our immunoregulated gene set, we found that 50.1% of our positively immunoregulated intestinal genes (977/1,951; 5.3-fold over background; $p = 0$) were also male-biased, while only 11.6% of them (227/1,951; 2.5-fold over background; $p = 3.34 \cdot 10^{-38}$) were female-biased (Table 9; Supplementary Table S15).

Such predominantly male-biased expression, along with the enrichment of three male-associated gene families (MSP, MFP2, and DUF236) in this gene set, raises the question of whether the changes in hookworm intestinal gene expression we observed with different host immunological backgrounds were actually due to our having harvested a higher proportion of male hookworms from normal than immunosuppressed hosts, leading to a male skew in putatively immunoregulated genes. We see two reasons why such bias is not sufficient to explain our observations. First, we only observed significant changes of gene activity in dissected intestines from 19-day adults, while observing no such changes in non-intestinal tissues from those same adults (Table 6). If apparent immunoregulation of genes had actually been due to overlooked male bias in the hookworms we collected from normal hosts, one would expect to see at least as much (if not more) changed gene expression in non-intestinal tissues (which contained most, if not all, of the gonadal tissue from adults) as we saw from dissected intestines. The absence of significant immunoregulation in non-intestinal tissues (or in our 12-day whole animals) is inconsistent with such an artifact. Second, positively immunoregulated intestinal genes are not a simple subset of male-biased genes, as would be expected if male selection accounted for immunological changes of gene expression: 747 of these 1,951 immunoregulated genes had no sex bias, and 227 of them had female-biased expression. We conclude that the transcriptional immunoregulation we describe here, though dominated by male-biased genes, is nevertheless real. This suggests that male and female *A. ceylanicum* have different responses to the host immune system, and that such differences might account for the observed male-bias from our mixed-sex *A. ceylanicum* 19-day intestinal RNA-seq data.

The host immune system affects ES gene expression. Protein products of ES genes are thought to affect the host immune system, but whether the host immune system affects ES genes has been unclear. Comparing the ES gene set with immunomodulated gene sets, we observed significant overlaps (Table 9). Out of 1,951 positively immunoregulated intestinal genes in 19-day hookworms, 153 also encoded ES proteins (27.1% of ES genes; 4.6-fold over chance; $p = 1.19 \cdot 10^{-59}$). Of 26 positively immunoregulated non-intestinal genes, 13 were also ES genes (29-fold above chance; $p = 7.47 \cdot 10^{-17}$); such expression might reflect synthesis and secretion of ES proteins by cephalic/pharyngeal glands (HUANG *et al.* 2020).

The 153 positively immunoregulated intestinal ES genes disproportionately encoded astacin proteases, TIMP and TIL protease inhibitors, ShK-like proteins, ASPs, ASPRs, and SCVPs (Table 10).

Of these genes, 69 had ES gene homologs in *A. caninum* (20-fold enrichment; $p = 6.00 \cdot 10^{-72}$), 24 in *N. americanus* (11-fold enrichment; $p = 2.35 \cdot 10^{-18}$), and 24 in *H. contortus* (4.4-fold enrichment; $p = 1.23 \cdot 10^{-9}$; Supplementary Table S16). As noted above, all of the above protein families may affect immunomodulation in hookworm hosts; in addition, astacin has been shown to enable tissue invasion *in vitro* by *A. caninum* (WILLIAMSON *et al.* 2006). Other immunoregulated ES genes also encoded proteins with possible functions in immunoregulation, antithrombosis, or digesting host tissue (Supplementary Table S6): the ASPs *Acey-NIF-B* (MOYLE *et al.* 1994; LO *et al.* 1999) and *Acey-HPI* (DEL VALLE *et al.* 2003), two apyrases (GUIGUET *et al.* 2016; PALA *et al.* 2023), deoxyribonuclease II (BOUCHERY *et al.* 2020), hyaluronidase (HOTEZ *et al.* 1992; YANG *et al.* 2020), and superoxide dismutase (KNOX AND JONES 1992; BROPHY *et al.* 1995).

Discussion

The ability of hookworms to suppress or evade their hosts' immune systems and feed on their hosts' blood and other tissues, for years, has motivated identifying hookworm genes whose products interact with the host, either by encoding ES proteins or by directly contacting host blood and tissues in the hookworm intestine (WEI *et al.* 2016; ABUZEID *et al.* 2020). Here we have identified ES proteins, intestine-biased genes, and immunoregulated genes in the hookworm *A. ceylanicum*, the last of which is unique to this study. Indeed, we hypothesized that hookworms regulate genes in response to the host immune system in order to suppress it, which motivated our immunosuppression/transcriptomic study here. Strongylid parasitic nematodes interact with the immune system despite generally not being blood-feeding (BANSEMIR AND SUKHDEO 1994; MCNEILLY *et al.* 2017; PRICE *et al.* 2019), so blood feeding is not necessary for such interactions. However, in the case of hookworms, our findings imply that interaction with the host immune system primarily occurs through contact of the hookworm intestinal lumen with host blood.

Zoonotic *A. ceylanicum* productively infects both humans and other mammals, making it important to humans while also enabling laboratory studies of a true hookworm that is clinically relevant (TRAUB 2013; COLELLA *et al.* 2021). We have analyzed both our new gene sets and previously identified gene sets (WEI *et al.* 2016; BERNOT *et al.* 2020; UZOECHI *et al.* 2023), to identify traits encoded either by ES genes, intestinal genes, immunoregulated genes, or all three. To enable this work, we repredicted protein-coding genes in *A. ceylanicum* to the highest completeness so far achieved. We have greatly expanded the number of intestine-biased genes in *A. ceylanicum* and found them to be extensively homologous to intestine-biased genes in *H. contortus* (LAING *et al.* 2013). We identified a mixture of functionally suggestive and

poorly understood protein families in *A. ceylanicum* ES proteins and intestine-biased genes, observed genes that are immunoregulated in adult intestines (where they are exposed to host blood and circulating immune factors) but not other tissues, detected an unexpected male bias in positively immunoregulated intestinal genes, and defined a positively immunoregulated subset of ES genes that have a mosaic of conservation and species-specificity.

On examination, it was possible to infer biologically coherent functions from what might seem to be a jumble of hookworm ES proteins. Immunosuppressors have been long suspected to be part of the ES protein repertoire, and we observed several protein types that could have this role. We also observed a pattern of diverse multigene families encoding ES proteins (ASPs, ASPRs, TTLs, and SCVPs); in other parasitic nematodes, such a pattern has been repeatedly seen both in ES proteins and in genetic diversity between strains or sibling species (ABUZEID *et al.* 2020; COLE *et al.* 2023; STEVENS *et al.* 2023). It is not clear why this pattern exists; perhaps hookworms and other parasitic nematodes use diverse ES multigene families to survive unpredictably varying host immune systems by confusing their hosts with complex and varying mixtures of protein antigens, which leads parasitic nematodes to accumulate multigene families through balancing selection (THOMAS AND ROBERTSON 2008; TELLIER *et al.* 2014; EBERT AND FIELDS 2020). Another possible function that consistently emerged when examining ES proteins was inhibition of blood clotting. Despite being less often discussed and less obviously virulent, antithrombotics are at least as important for parasitism as immunosuppressants. Hookworm anticoagulants were first observed in 1904 (LOEB AND SMITH 1904), and are likely to be necessary for sustained blood-feeding (KVIST *et al.* 2020; OLIVEIRA AND GENTA 2021; PALA *et al.* 2023).

The strong overlap that we observed between immunoregulation and male-biased gene expression raises the question of whether hookworms have sex-specialized repertoires of genes that they express during infections (COSTA *et al.* 2009). There has been one striking instance (in the African tick *R. appendiculatus*) where males and females interact differently with their hosts and the males act to promote female success during parasitism (WANG *et al.* 1998). Our current data do not indicate whether such cooperation happens during hookworm infections, let alone whether the male-biased immunoregulated genes we describe are relevant to such hypothetical cooperation. However, such a bias might ensure that both sexes are present in a successful infection, which is required for reproduction. More extensive analysis to test for sexual cooperation during infections or sex-biased gene immunoregulation should be pursued for hookworms and other parasitic nematodes.

Immunoregulated genes of *A. ceylanicum* have homologs in *A. caninum*, *N. americanus*, and *H. contortus*. The homologs of immunoregulated ES genes also encode ES proteins, suggesting that other

parasitic nematodes may also have immunoregulated genes, and some findings indicate that they actually do. For the strongylid *Teladorsagia circumcincta*, being embedded in host mucosa upregulates ES genes and putative immunomodulatory genes (MCNEILLY *et al.* 2017; PRICE *et al.* 2019). For the strongylid *H. bakeri*, infecting mouse hosts with colitis induces different ES proteins than normal hosts (MARUSZEWSKA-CHERUIYOT *et al.* 2023). Finally, for the strongylid *N. brasiliensis*, recent transcriptomic and proteomic analyses shows genes that are positively or negatively regulated during infections of normal versus immunocompromised *stat6* mutant mouse hosts, a subset of which encode ES proteins (FERGUSON *et al.* 2023; FERGUSON *et al.* 2024). Thus, the immunoregulation we observe here for *A. ceylanicum* may be a general phenomenon found in many other parasitic nematodes.

Figure Legends

Figure 1. Morphology and evolution of the hookworm *Ancylostoma ceylanicum*. (A) Female of *A. ceylanicum* by 12 days after infection of a hamster host; scale bar, 500 μ m. At this stage of development, the hookworms are young adults that have only just begun blood feeding, with mature males and a few gravid females, and with little or no egg laying. (B) Male of *A. ceylanicum* by 19 days after infection of a hamster host; scale bar, 1 mm. At this stage of development, the hookworms are fully mature adults that have been blood-feeding for at least a week, have mated, and have begun extensive egg laying that can last for weeks in a hamster host. Both photographs are reproduced from Schwarz *et al.* (SCHWARZ *et al.* 2015). (C) Evolutionary relationships of the hookworm *A. ceylanicum* to some other nematodes discussed in this paper. This phylogeny links *A. ceylanicum* to other *Ancylostoma* and *Necator* hookworms, to more distantly related strongylid parasitic nematodes, and to two well-studied free-living rhabditid nematodes (*C. elegans* and *Pristionchus pacificus*). Phylogenetic data are taken from Coghlan *et al.*, van Megen *et al.*, and Xie *et al.* (VAN MEGEN *et al.* 2009; XIE *et al.* 2017; COGHLAN *et al.* 2019). *A. ceylanicum* is highlighted in gold; three parasitic species (*A. caninum*, *N. americanus*, and *H. contortus*) whose ES genes were compared to those of *A. ceylanicum* in this paper are highlighted in red. Hookworms form a single clade (marked with a yellow bar); they are a subset of a larger clade of strongylid nematodes (marked with a green bar) which, like hookworms, are parasitic. Strongylids are, in turn, a subset of rhabditid nematodes (marked with a blue bar); this clade encompasses both parasitic nematodes and the free-living nematodes *C. elegans* and *P. pacificus*. Notably, *A. ceylanicum* and other strongylid parasites are more closely related to *C. elegans* than to *P. pacificus* despite having highly divergent parasitic life cycles.

Figure 2. Gene expression in *A. ceylanicum*. Gene activity is shown for 8,130 *A. ceylanicum* genes with significant differences in expression between biological conditions, with biological replicates used for differential gene expression analysis on the x-axis and individual genes on the y-axis. Expression levels are in TPM (log10). Genes sharing similar patterns of expression are split into 10 clusters. Biological replicates are: young 12-day hookworm adults from normal hosts (YA_12dpi_noDEX); young 12-day adults from immunosuppressed hosts (YA_12dpi_DEX); dissected 19-day intestines from mature hookworm adults in normal hosts (intestines_19dpi_noDEX); dissected 19-day intestines in immunosuppressed hosts (intestines_19dpi_DEX); dissected 19-day intestines in normal hosts (non.intestines_19dpi_noDEX); dissected 19-day non-intestinal tissues in immunosuppressed hosts (non.intestines_19dpi_DEX); previously published adult male intestine (WashU_male_intestine);

published adult males (WashU_adult_male); and published adult females (WashU_adult_fem). Additional heatmaps of gene expression for all 33,190 genes and all replicates are given in Supplemental Figure 1.

Tables and Table Legends

Table 1. Protein-coding gene predictions for *A. ceylanicum*.

Predictions	Acey v1.0	Coghlan 2019	Uzoechi 2023	Acey v2.0	Acey v2.1
Genes	36,687	17,332	18,773	19,419	33,190
Proteins	65,583	17,332	22,925	21,309	37,343
Max. residues	8,404	5,888	8,389	7,897	7,897
Mean residues	306.7	297.8	437.8	401.5	299
Median residues	199	224.0	320	300	170
Min. residues	18	26	21	30	30
Completeness	87.8%	62.0%	94.5%	95.1%	95.3%
Reference	(SCHWARZ <i>et al.</i> 2015)	(COGHLAN <i>et al.</i> 2019)	(UZOECHI <i>et al.</i> 2023)	This paper	This paper

This gives both general statistics and BUSCO completeness scores for three previously predicted gene sets (including our original predictions, v1.0), our first repredictions (v2.0), and our final repredictions used in this paper (v2.1). Maximum, mean, median, and minimum amino acid numbers are given for each set of predicted proteins.

Table 2. Numbers of genes encoding ES proteins of *A. ceylanicum*.

Category	Genes
Total v2.1 gene set	33,190
Our ES, at least one collection (≥ 1)	565
Our ES, both collections (2)	350
Uzoechi ES, either sex	860
Uzoechi ES, both sexes	608
Uzoechi ES, male only	135
Uzoechi ES, female only	117
Our ES (≥ 1) and U. (either)	430
Our ES (2) and U. (either)	308
Our ES (≥ 1), no U. ES	135
Our ES (2), no U. ES	42
Our ES (≥ 1), no U. gene	14
Our ES (2), no U. gene	7
Either our ES (≥ 1) or U. (either)	995
Either our ES (2) or U. (either)	902

Numbers are given for our observations, for recently published observations by Uzoechi *et al.* (UZOECHI *et al.* 2023) indicated by "U.", and for their subsets and overlaps. Uzoechi *et al.* independently observed ES proteins from males and females, so a gene in their set could encode ES proteins from either or both sexes. ES genes by Uzoechi *et al.* were mapped from their predicted gene set (Table 1) to ours before analysis. Of our observed ES genes, 14 had no corresponding gene in the Uzoechi gene set, and thus might have actually been present in the Uzoechi ES mass spectrometry data yet not mapped onto a gene.

Table 3. Protein motifs overrepresented in *A. ceylanicum* ES genes.

Motif	Motif genes	Overlap	Enrichment	q-value
N-terminal signal sequence (SigP)	4,825	380	4.63	4.51e-181
Cysteine-rich secretory protein (CAP) [PF00188.29]	418	140	19.67	5.64e-140
ASP-related (ASPR) [PF17641.5]	200	51	14.98	1.88e-41
Papain family cysteine protease [PF00112.26]	67	24	21.04	1.60e-22
Trypsin inhibitor-like cysteine-rich domain (TIL) [PF01826.20]	71	19	15.72	7.83e-15
Astacin (Peptidase family M12A) [PF01400.27]	122	22	10.59	1.41e-13
Kunitz/bovine pancreatic trypsin inhibitor domain [PF00014.26]	111	19	10.06	3.74e-11
Transthyretin-like family (TTL, e.g., TTR-52) [PF01060.26]	87	16	10.8	1.01e-09
Potassium channel inhibitor ShK domain-like [PF01549.27]	111	16	8.47	4.18e-08
Tissue inhibitor of metalloproteinase (TIMP) [PF00965.20]	31	8	15.16	0.0000169
Secreted clade V proteins (SCVP) [PF17619.5]	59	10	9.96	0.000023
Glutathione S-transferase, C-terminal domain [PF00043.28]	22	5	13.35	0.0109
Peptidase family M13/nepriylsin metallopeptidase [PF01431.24]	39	6	9.04	0.0163
Aspartyl protease [PF00026.26]	57	7	7.21	0.0163
Lectin C-type domain [PF00059.24]	102	9	5.18	0.0191
Serine carboxypeptidase S28 [PF05577.15]	14	4	16.78	0.0196
Peptidase family M13/nepriylsin metallopeptidase, N-terminal domain [PF05649.16]	42	6	8.39	0.0196

This table shows 16 Pfam protein motifs overrepresented in the ES gene set with q-values of ≤ 0.02 , along with one overrepresented signal or transmembrane domain configuration (predicted by Phobius) treated as a motif. "Motif" denotes the protein motif-encoding gene set being tested for nonrandomly high overlap with the ES gene set, with its Pfam accession number; "Motif genes" gives the number of protein motif-encoding genes in the genome being tested for overlap; "Overlap" gives the observed number of genes falling into both motif and ES categories; "Enrichment" gives the ratio of observed overlaps to those expected randomly with 565 out of 33,190 genes encoding ES proteins; "q-value" gives a statistical significance for the observed overlap, computed by a two-tailed Fisher test and corrected for multiple hypothesis testing (BAILEY *et al.* 2009; NOBLE 2009). A statistical value of "Xe-Y" denotes $X \cdot 10^{-Y}$. A full set of all protein motif statistics computed by Pfam, InterPro, or Phobius for both our ES gene set and that of Uzoechi *et al.* (UZOECHI *et al.* 2023) is given in Supplementary Table S7.

Table 4. Individual *A. ceylanicum* ES genes of interest.

Predicted product	ES protein-encoding gene	Possible importance
<i>A. caninum</i> anticoagulant protein (AcAP) homolog	Acey_s0112.v2.g13097	Anticoagulant
<i>A. caninum</i> anticoagulant protein (AcAP) homolog	Acey_s0112.v2.g13104	Anticoagulant
<i>A. caninum</i> anticoagulant protein (AcAP) homolog	Acey_s0112.v2.g13105	Anticoagulant
Adenosine deaminase	Acey_s0243.v2.g530	Anti-coagulation/-inflammation
Apyrase	Acey_s0038.v2.g9677	Anti-coagulation/-inflammation; vaccine subunit
Apyrase	Acey_s0038.v2.g9678	Anti-coagulation/-inflammation; vaccine subunit
Calreticulin	Acey_s0340.v2.g13758	Vaccine subunit
Cysteine protease	Acey-CP-1	Digestion; vaccine subunit
Deoxyribonuclease II	Acey_s0473.v2.g18261	Hydrolyzing neutrophil DNA nets
Enolase	Acey_s0389.v2.g8512	Vaccine subunit
Fatty acid- and retinol-binding (FAR) protein	Acey_s0005.v2.g14583	Lipid immunosuppression; vaccine subunit
Fatty acid- and retinol-binding (FAR) protein	Acey_s0064.v2.g6025	Lipid immunosuppression; vaccine subunit
Fructose-1,6-bisphosphate aldolase	Acey_s0324.v2.g12031	Vaccine subunit
Glutathione S-transferase	Acey-GST-1	Detoxification; vaccine subunit
Hypodermal Ac-16 antigen homolog	Acey_s0009.v2.g6653	Vaccine subunit
M13 peptidase	Acey-MEP-6	Digestion; vaccine subunit
M13 peptidase	Acey-MEP-7	Digestion; vaccine subunit
Macin	Acey_s0022.v2.g18553	Antimicrobial peptide
Macrophage migration inhibitory factor	Acey-MIF-1	Suppressing cellular immunity
Phospholipase A2	Acey_s0002.v2.g4275	Lipid immunosuppression
Phospholipase D	Acey_s0097.v2.g996	Lipid immunosuppression

Table 5. Homologies of ES genes of *A. ceylanicum* to those of related parasitic nematodes.

Category 1	Category 2	Category 1 genes	Category 2 genes	Overlap	Enrichment	p-value
Homology to <i>A. caninum</i> ES gene	All ES	748	565	269	21.13	6.82e-301
Homology to any <i>A. caninum</i> gene	All ES	17,742	565	519	1.72	4.44e-90
Homology to any <i>A. caninum</i> gene	ES w/o homol. to <i>A. caninum</i> ES	17,742	296	250	1.58	1.31e-29
Homology to <i>A. caninum</i> ES gene	Pos. immun. int. ES	748	153	69	20.01	6.00e-72
Homology to any <i>A. caninum</i> gene	Pos. immun. int. ES	17,742	153	130	1.59	2.64e-16
Homology to <i>N. americanus</i> ES gene	All ES	466	565	141	17.77	1.78e-137
Homology to any <i>N. americanus</i> gene	All ES	16,103	565	334	1.22	4.01e-07
Homology to any <i>N. americanus</i> gene	ES w/o homol. to <i>N. americanus</i> ES	16,103	424	193	0.94	0.222
Homology to <i>N. americanus</i> ES gene	Pos. immun. int. ES	466	153	24	11.17	2.35e-18
Homology to any <i>N. americanus</i> gene	Pos. immun. int. ES	16,103	153	59	0.79	0.0148
Homology to <i>H. contortus</i> ES gene	All ES	1,188	565	180	8.90	4.10e-120
Homology to any <i>H. contortus</i> gene	All ES	15,959	565	257	0.95	0.218
Homology to any <i>H. contortus</i> gene	ES w/o homol. to <i>H. contortus</i> ES	15,959	385	77	0.42	2.05e-30
Homology to <i>H. contortus</i> ES gene	Pos. immun. int. ES	1,188	153	24	4.38	1.23e-09
Homology to any <i>H. contortus</i> gene	Pos. immun. int. ES	15,959	153	39	0.53	1.27e-08
Any ES homology	All ES	1,684	565	350	12.21	0
Any homology	ES w/o homol. to any ES	20,368	215	173	1.31	2.16e-09

This table shows overlaps between sets of *A. ceylanicum* genes homologous either to ES protein-encoding genes of related parasitic nematodes, or to any genes of those related parasitic nematodes, versus *A. ceylanicum* genes that themselves encode ES proteins. "Category 1" and "Category 2" denote which classes of genes are being compared for nonrandomly high (or low) degrees of overlap. "All ES" denotes the full set of *A. ceylanicum* ES genes; "Pos. immun. int. ES" denotes ES genes with positive immunoregulation in 19-day adult intestines; and other ES categories show subsets of ES genes that lack homology to ES protein-coding genes of related species. "Category [1/2] genes" gives the number of genes

in each category being compared; "Overlap" the number of genes in both categories; "Enrichment" the ratio of observed to randomly expected overlaps, against the genome-wide background of 33,190 genes; "p-value" gives a statistical significance for the observed overlap, computed by a two-tailed Fisher test (BAILEY *et al.* 2009). A full set of all ES gene homology statistics for both our ES gene set and that of Uzoechi *et al.* (UZOECHI *et al.* 2023) is given in Supplementary Table S11.

Table 6. Numbers of *A. ceylanicum* genes showing differential RNA expression (significant changes of gene expression between biological conditions).

Category	Genes
Total v2.1 gene set	33,190
Intestine-biased (19-day)	1,670
Non-intestine-biased (19-day)	1,196
Intestine-biased (prev. pub. data)	635
Non-intestine-biased (prev. pub. data)	954
19-day positively immunoregulated intestinal	1,951
19-day negatively immunoregulated intestinal	137
19-day positively non-intestinal immunoreg.	26
19-day negatively non-intestinal immunoreg.	4
Any 19-day immunoregulated	2,112
12-day positively immunoregulated	0
12-day negatively immunoregulated	2
Male-biased (prev. pub. data)	3,135
Female-biased (prev. pub. data)	1,547
Any change of gene expression above	8,130

Differentially expressed genes were computed from both our RNA-seq data and previously published *A. ceylanicum* RNA-seq data (Supplementary Table S4) (WEI *et al.* 2016; BERNOT *et al.* 2020).

Table 7. Protein motifs overrepresented in *A. ceylanicum* genes with intestine-biased expression.

Motif	Motif genes	Overlap	Enrichment	q-value
Papain family cysteine protease [PF00112.26]	67	48	14.24	1.08e-43
Eukaryotic aspartyl protease [PF00026.26]	57	30	10.46	7.40e-21
Xylanase inhibitor N-terminal [PF14543.9]	43	22	10.17	1.37e-14
Uncharacterized oxidoreductase <i>dhs-27</i> (DUF1679) [PF07914.14]	19	15	15.69	1.14e-13
Ecdysteroid kinase-like family [PF02958.23]	18	14	15.46	1.44e-12
UDP-glucuronosyl and UDP-glucosyl transferase [PF00201.21]	48	21	8.70	2.18e-12
Major facilitator superfamily [PF07690.19]	149	32	4.27	1.63e-09
Phosphotransferase enzyme family [PF01636.26]	34	15	8.77	1.33e-08
Core histone H2A/H2B/H3/H4 [PF00125.27]	53	18	6.75	2.35e-08
SigP+TM(1x)	673	72	2.13	7.83e-08
Glycosyltransferase family 28 C-terminal domain [PF04101.19]	39	15	7.64	1.14e-07
TM(11x)	111	23	4.12	1.22e-07
Histone-like transcription factor (CBF/NF-Y) and archaeal histone [PF00808.26]	54	16	5.8	2.21e-06
SigP	4,825	318	1.31	3.02e-06
Sugar (and other) transporter [PF00083.27]	71	18	5.04	3.56e-06
ABC transporter [PF00005.30]	73	18	4.90	5.29e-06
TM(12x)	140	23	3.27	5.34e-06
TM(10x)	99	18	3.61	1.60e-05
Strongyloid L4 protein (SL4P) [PF17618.5]	25	10	7.95	5.33e-05
TM(6x)	237	29	2.43	9.55e-05
C-terminus of histone H2A [PF16211.8]	18	8	8.83	0.000338
Peptidase family M1/aminopeptidase [PF01433.23]	24	9	7.45	0.000375
Peptidase family M13/neprilysin metallopeptidase [PF01431.24]	39	11	5.61	0.000611

This table shows 17 Pfam protein motifs and six overrepresented signal or transmembrane domain configurations predicted by Phobius that are overrepresented in the intestine-biased gene set (1,670 out of 33,190 genes) with q-values of ≤ 0.001 . Phobius domains are abbreviated with 'SigP' for an N-terminal signal sequence, 'TM(Nx)' for N transmembrane domains, and 'SigP+TM(Nx)' for both. Data columns are labeled as in Table 3. A full set of all protein motif statistics computed by Pfam, InterPro, or Phobius for intestine- and non-intestine-biased genes is given in Supplementary Table S13.

Table 8. Protein motifs overrepresented in positively immunoregulated intestinal *A. ceylanicum* genes.

Motif	Motif genes	Overlap	Enrichment	q-value
SigP	4,825	469	1.65	7.85e-29
Protein kinase domain [PF00069.28]	341	84	4.19	3.08e-26
Protein-tyrosine phosphatase [PF00102.30]	64	37	9.83	8.41e-26
Protein tyrosine and serine/threonine kinase [PF07714.20]	339	82	4.11	1.92e-25
SigP+TM(1x)	673	101	2.55	1.04e-16
Major sperm protein (MSP) domain [PF00635.29]	50	23	7.83	9.07e-13
TM(1x)	2,662	237	1.51	9.49e-10
Src homology 2 (SH2) [PF00017.27]	50	19	6.46	1.49e-08
Trypsin [PF00089.29]	37	16	7.36	4.32e-08
Astacin (Peptidase family M12A) [PF01400.27]	122	29	4.04	4.32e-08
Receptor L domain [PF01030.27]	43	17	6.73	5.44e-08
Cysteine-rich secretory protein family (CAP) [PF00188.29]	418	60	2.44	7.53e-08
Nematode trypsin-6-like family (DUF316) [PF03761.18]	29	13	7.63	1.10e-06
Immunoglobulin domain (ig) [PF00047.28]	58	17	4.99	8.22e-06
Immunoglobulin V-set domain [PF07686.20]	44	14	5.41	4.21e-05
Immunoglobulin domain (Ig 3) [PF13927.9]	59	16	4.61	5.64e-05
Secreted clade V proteins (SCVP) [PF17619.5]	59	16	4.61	5.64e-05
Immunoglobulin I-set domain [PF07679.19]	67	17	4.32	6.13e-05
ASP-related (ASPR) [PF17641.5]	200	32	2.72	6.13e-05
Immunoglobulin domain (Ig 2) [PF13895.9]	53	15	4.81	6.13e-05
DUF236 repeat [PF03057.17]	7	6	14.58	6.47e-05
Immunoglobulin domain (Ig 6) [PF18452.4]	11	7	10.83	0.000145
Trypsin-like peptidase domain [PF13365.9]	20	9	7.66	0.000163
Tissue inhibitor of metalloproteinase (TIMP) [PF00965.20]	31	11	6.04	0.000163
C17orf99 Ig domain [PF17736.4]	8	6	12.76	0.000203
von Willebrand factor type A domain (VWA) [PF00092.31]	46	13	4.81	0.000295
SigP+TM(2x)	53	12	3.85	0.000460
Myosin tail [PF01576.22]	9	6	11.34	0.000522
Dopamine beta-monoxygenase N-terminal (DOMON) domain [PF03351.20]	13	7	9.16	0.000522
Peptidase family M13/nepriylsin metallopeptidase, N-terminal domain [PF05649.16]	42	12	4.86	0.000582
MFP2b cytosolic motility protein [PF12150.11]	6	5	14.18	0.000621
Leishmanolysin (peptidase family M8) [PF01457.19]	6	5	14.18	0.000621
TM(4x)	415	45	1.84	0.000664
Potassium channel inhibitor ShK domain-like [PF01549.27]	111	19	2.91	0.00367
Immunoglobulin C1-set domain [PF07654.18]	12	6	8.51	0.00397
CD80-like C2-set immunoglobulin domain [PF08205.15]	29	9	5.28	0.00397
Myosin N-terminal SH3-like domain [PF02736.22]	8	5	10.63	0.00461
Sodium:neurotransmitter symporter family (SNF) [PF00209.21]	24	8	5.67	0.00587
Peptidase family M13/nepriylsin metallopeptidase [PF01431.24]	39	10	4.36	0.00819

This table shows 10 Pfam protein motifs and four overrepresented signal or transmembrane domain configurations predicted by Phobius that are overrepresented in the positively immunoregulated intestinal gene set (1,951 out of 33,190 genes) with q-values of ≤ 0.01 . Phobius domains are abbreviated with 'SigP'

for an N-terminal signal sequence, 'TM(Nx)' for N transmembrane domains, and 'SigP+TM(Nx)' for both. Data columns are labeled as in Table 3. A full set of all protein motif statistics computed by Pfam, InterPro, or Phobius for immunoregulated genes is given in Supplementary Table S15.

Table 9. Gene categories overrepresented in *A. ceylanicum* immunoregulated genes.

Immunoreg. category	Other category	Immunoreg. genes	Other genes	Overlap	Enrichment	p-value
Pos. intestinal	Male-biased	1,951	3,135	977	5.30	0
Pos. intestinal	Female-biased	1,951	1,547	227	2.50	3.34e-38
Pos. intestinal	Intestine-biased	1,951	1,670	1	0.0102	1.54e-43
Pos. intestinal	Non-intestine-biased	1,951	1,196	258	3.67	5.51e-78
Pos. intestinal	All ES	1,951	565	153	4.61	1.19e-59
Neg. intestinal	Male-biased	137	3,135	20	1.55	0.0548
Neg. intestinal	Female-biased	137	1,547	2	0.31	0.0988
Neg. intestinal	Intestine-biased	137	1,670	50	7.25	6.56e-30
Neg. intestinal	Non-intestine-biased	137	1,196	4	0.810	1
Neg. intestinal	All ES	137	565	8	3.43	0.00247
Pos.non-intestinal	Male-biased	26	3,135	6	2.44	0.0312
Pos.non-intestinal	Female-biased	26	1,547	0	0	0.632
Pos.non-intestinal	Intestine-biased	26	1,670	0	0	0.640
Pos.non-intestinal	Non-intestine-biased	26	1,196	10	10.67	1.12e-08
Pos.non-intestinal	All ES	26	565	13	29.37	7.47e-17
Pos. int. or non-int.	All ES	1,975	565	164	4.88	5.53e-68

All categories are from 19-day hookworms, and are fully described in Results. Overlaps, enrichments, and statistics are computed as in Table 5. Full statistics are given in Supplementary Table S15.

Table 10. Protein motifs overrepresented in positively immunoregulated intestinal *A. ceylanicum* ES genes.

Motif	Motif genes	Overlap	Enrichment	q-value
SigP	4825	116	5.22	1.32e-63
Cysteine-rich secretory protein (CAP) [PF00188.29]	418	49	25.43	2.07e-51
ASP-related (ASPR) [PF17641.5]	200	25	27.12	2.63e-25
Astacin (Peptidase family M12A) [PF01400.27]	122	15	26.67	2.72e-14
Tissue inhibitor of metalloproteinase (TIMP) [PF00965.20]	31	7	48.98	1.04e-07
Trypsin inhibitor-like cysteine-rich domain (TIL) [PF01826.20]	71	7	21.39	3.61e-05
Potassium channel inhibitor ShK domain-like [PF01549.27]	111	8	15.63	3.77e-05
Secreted clade V proteins (SCVP) [PF17619.5]	59	5	18.38	0.00515

This table shows seven Pfam protein motifs and one overrepresented signal or transmembrane domain configuration predicted by Phobius that are overrepresented in the positively immunoregulated intestinal ES gene set (153 out of 33,190 genes) with q-values of ≤ 0.01 . Phobius domains are abbreviated with 'SigP' for an N-terminal signal sequence. Data columns are labeled as in Table 3. A full set of all protein motif statistics computed by Pfam, InterPro, or Phobius for positively immunoregulated intestinal ES genes is given in Supplementary Table S16.

Supplementary Table, Figure, and File Legends

Supplementary Table S1. *A. ceylanicum* RNA-seq libraries newly generated in this study. "RNA-seq library ID" provides a short abbreviation for each library; these abbreviations have been used for expression levels (TPMs), mapped read counts (reads), and differential gene expression analyses in later supplementary data tables. For each library, "SRA accession", "BioProject accession" and "BioSample accession" provide accession numbers; "Reads", "Total nt", and "Read length in nt" give the number of reads, total sequence in nt, and read length; and "RNA-seq description" summarizes biological contents.

Supplementary Table S2. Software used in this study. "Software" provides the name of each computer program or suite of computer programs; "Purpose" describes why this software was used here; "Main web site (URL) or code location" gives the primary Web site for this software; "Bioconda source (if used)" gives the bioconda web site for programs that were installed as bioconda environments; "Online documentation" gives the web site for detailed online manuals, if a given program has one. Publications and arguments for each program are cited and described in Methods.

Supplementary Table S3. Published genomic data used in this study. Genome sequences, coding sequences, proteomes, and gene annotations of relevant nematodes were downloaded from WormBase (DAVIS *et al.* 2022) or ParaSite (HOWE *et al.* 2017; LEE *et al.* 2017) and used for transcriptomic or proteomic analyses; they include Coghlan gene predictions and their corresponding *A. ceylanicum* genome assembly, along with UniProt (UNIPROT 2023) and RefSeq (O'LEARY *et al.* 2016) proteome databases from highly GO-annotated model organisms. Four separate spreadsheets are given for data files of genomes, proteomes, CDS DNA sets, and gene annotations in GFF format. For each data file, "Species" gives the biological species described by the file, "Comments" describes the particular use(s) to which the data file was put in this study, and "URL" gives the Web source of the file.

Supplementary Table S4. Published RNA-seq data of *A. ceylanicum* (SCHWARZ *et al.* 2015; WEI *et al.* 2016; BERNOT *et al.* 2020) and *H. contortus* (LAING *et al.* 2013) used in this study. All RNA-seq files were downloaded from the European Nucleotide Archive (ENA). "Species" gives the file's origin species. "Abbreviation" provides a short abbreviation for each library; these abbreviations have been used for expression levels (TPMs), mapped read counts (reads), and differential gene expression analyses in later supplementary data tables. "Biological condition (sex, developmental stage, tissue) and replicate number" describes biological content and replication. "Notes" describes any ambiguities that had to be resolved

during analysis of the data. "Database" is uniformly 'ENA', but is included for completeness in the table. "Accession" and "URL" give SRA accession numbers and ENA Web sources for each file.

Supplementary Table S5. LC-MS/MS analyses of two *A. ceylanicum* ES protein sets, E20201108-05 and E20201108-07. Data columns were generated by Proteome Discoverer 2.4; other details of the analysis are given in Methods.

Supplementary Table S6. Annotations, RNA-seq expression, and differential gene expression for the *A. ceylanicum* v2.1 protein-coding gene set. Its data columns are as follows.

Gene identifications and equivalencies

Gene: A given predicted protein-coding gene in the *A. ceylanicum* genome assembly. All further data columns are pertinent to that particular gene.

Gene_name: A human-readable gene name, where it exists (e.g., "Acey-CP-1" instead of simply "Acey_s0154.v2.g3234"). These names are used in the main text to discuss genes of particular interest.

Mapped_v1.0_gene: Any gene or genes from our previous v1.0 predictions which overlap the v2.1 gene here by least 1 nt of coding exon sequence in the *A. ceylanicum* genome. Although this criterion errs on the side of sensitivity and we have made no effort to filter out short overlaps, we expect that in practice this will identify extensive exon overlaps between an earlier v1.0 gene and its reprediction in v2.1.

Mapped_Coghlan_gene: Any gene or genes from the previous *A. ceylanicum* gene predictions by Coghlan *et al.* (COGHLAN *et al.* 2019) which, when lifted over onto our *A. ceylanicum* genome assembly, overlap the v2.1 gene here by least 1 nt of coding exon sequence.

Mapped_Uzoechi_gene: Any gene or genes from the recent *A. ceylanicum* gene repredictions by Uzoechi *et al.* (UZOECHI *et al.* 2023) which overlap the v2.1 gene here by least 1 nt of coding exon sequence.

General traits of protein products

Max_prot_size: The size of the largest predicted protein product.

Prot_size: This shows the full range of sizes for all protein products from a gene's predicted isoforms.

Phobius: This denotes predictions of signal and transmembrane sequences made with Phobius 1.01 (KÄLL *et al.* 2004). 'SigP' indicates a predicted signal sequence, and 'TM' indicates one or more transmembrane-spanning helices, with N helices indicated with '(Nx)'. Varying predictions from different isoforms are listed.

NCoils: This shows coiled-coil domains, predicted by Ncoils 2002.08.22 (LUPAS 1996). Both the proportion of such sequence (ranging from 0.01 to 1.00) and the exact ratio of coiled residues to total residues are given. Proteins with no predicted coiled residues are blank.

Psegs: This shows what fraction of a protein is low-complexity sequence, as detected by PSEG 1999.06.10 (WOOTTON 1994). As with Ncoils, relative and absolute fractions of low-complexity residues are shown.

Pfam: Predicted protein domains from Pfam 35.0 (MISTRY *et al.* 2021), with family-specific significance thresholds.

InterPro: Predicted protein domains from InterProScan 5.57-90.0 (PAYSAN-LAFOSSE *et al.* 2023).

AMP: An antimicrobial peptide (AMP) gene annotation for v1.0 of our *A. ceylanicum* gene predictions, predicted by Irvine *et al.* (IRVINE *et al.* 2023), and mapped onto our v2.1 gene predictions here.

GO_Biological: Annotations from the biological subset of Gene Ontology (GO) terms (ASHBURNER *et al.* 2000; CARBON *et al.* 2021), generated with EnTAP 0.10.7-beta (HART *et al.* 2020).

GO_Molecular: Annotations from the molecular subset of Gene Ontology (GO) terms (ASHBURNER *et al.* 2000; CARBON *et al.* 2021), generated with EnTAP 0.10.7-beta (HART *et al.* 2020).

GO_Cellular: Annotations from the cellular subset of Gene Ontology (GO) terms (ASHBURNER *et al.* 2000; CARBON *et al.* 2021), generated with EnTAP 0.10.7-beta (HART *et al.* 2020).

EggNOG_description: EggNOG descriptions (HERNANDEZ-PLAZA *et al.* 2023), generated with EnTAP 0.10.7-beta (HART *et al.* 2020).

EggNOG_KEGG: KEGG codes (KANEHISA *et al.* 2023), generated with EnTAP 0.10.7-beta (HART *et al.* 2020).

Orthologies of protein products

OFind_Summary and **OFind_Full:** The results for our OrthoFinder analysis (EMMS AND KELLY 2019) of orthologies between *A. ceylanicum* and the related nematodes *A. caninum*, *C. elegans*, *H. contortus*, *H. bakeri*, *N. americanus*, *N. brasiliensis*, *Pristionchus pacificus*, and *T. circumcincta*. For one of these species (*N. americanus*) two different proteomes were included: the original predicted proteome by Tang *et al.* (TANG *et al.* 2014) labeled 'necator_orig', and the repredicted proteome by Logan *et al.* (LOGAN *et al.* 2020) labeled 'necator_rev'. Two different views of these results are given: the summary lists taxa and gene counts, while the full results give individual gene names.

ES- and gene-expression-related traits

Intest_haemonchus: Homologous *H. contortus* genes (taken from **OFind_Full**) that have intestine-biased gene expression, as computed by our analysis of previously published *H. contortus* RNA-seq data (Supplementary Table S4).

Non-intest_haemonchus: Homologous *H. contortus* genes (taken from **OFind_Full**) that have non-intestine-biased gene expression, as computed by our analysis of previously published *H. contortus* RNA-seq data (Supplementary Table S4).

ES_component: A gene whose protein product was detected in at least one of our two ES mass spectrometry experiments, E20201108-05 and E20201108-07. Note that this was used as a binary (Boolean) classification for statistical analyses of gene set overlaps; see below for other such binary classifications.

ES_observations: Observations of a gene's protein product being present in either E20201108-05, or E20201108-07, or both.

Coghlan-spec_ES: In our mass spectrometry analysis, our observation of a peptide mapping specifically to a Coghlan gene-encoded protein that was distinct from our v2.1 proteins by at least one amino acid residue, but whose Coghlan gene's coding exons could then be lifted over to the coding exons of the v2.1 gene here (**Mapped_Coghlan_gene**). For each such mapping, the ES observations behind it are noted (either E20201108-05, or E20201108-07, or both).

Uzoechi_ES: An ES-encoding gene detected by Uzoechi *et al.* (UZOECHI *et al.* 2023) whose coding exons mapped onto the coding exons of this v2.1 gene (in **Mapped_Uzoechi_gene**). Note that Uzoechi *et al.* separately observed both male and female ES proteins, and those specific observations are given here.

ES_a_caninum: Homology to an *A. caninum* gene encoding an *A. caninum* ES protein observed by Morante *et al.* (MORANTE *et al.* 2017), extracted from **OFind_Full**.

ES_necator_rev: Homology to an *N. americanus* gene encoding an *N. americanus* ES protein observed by Logan *et al.* (LOGAN *et al.* 2020), extracted from **OFind_Full**.

ES_haemonchus: Homology to an *H. contortus* gene encoding an *H. contortus* ES protein observed by Wang *et al.* (WANG *et al.* 2019), extracted from **OFind_Full**.

Binary (Boolean) classifications of genes

Male-biased: Annotation here indicates a gene with male-biased gene expression, as defined by ≥ 2 -fold higher expression in males versus females, with a false discovery rate (FDR) of ≤ 0.01 .

Female-biased: Annotation here indicates a gene with female-biased gene expression, as defined by ≥ 2 -fold higher expression in females versus males, with a false discovery rate (FDR) of ≤ 0.01 .

Intestine-biased: Annotation here indicates a gene with intestine-biased gene expression, as defined by ≥ 2 -fold higher expression in 19-day intestinal tissue from normal hosts versus 19-day non-intestinal tissue from normal hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Non-intestine-biased: Annotation here indicates a gene with non-intestine-biased gene expression, as defined by ≥ 2 -fold higher expression in 19-day non-intestinal tissue from normal hosts versus 19-day intestinal tissue from normal hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Any.19d.immunoreg: Annotation here indicates a gene with either intestinal or non-intestinal and either positively or negatively immunoregulated gene expression, as defined by ≥ 2 -fold higher or lower expression in 19-day intestinal or non-intestinal tissue from normal hosts versus 19-day corresponding tissue from dexamethasone-immunosuppressed hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Any.19d.pos.immunoreg: Annotation here indicates a gene with either intestinal or non-intestinal positively immunoregulated gene expression, as defined by ≥ 2 -fold higher in 19-day intestinal or non-intestinal tissue from normal hosts versus 19-day corresponding tissue from dexamethasone-immunosuppressed hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Pos.intest.immunoreg: Annotation here indicates a gene with intestinal positively immunoregulated gene expression, as defined by ≥ 2 -fold higher expression in 19-day intestinal tissue from normal hosts versus 19-day intestinal tissue from dexamethasone-immunosuppressed hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Neg.intest.immunoreg: Annotation here indicates a gene with intestinal negatively immunoregulated gene expression, as defined by ≥ 2 -fold higher expression in 19-day intestinal tissue from dexamethasone-immunosuppressed hosts versus 19-day intestinal tissue from normal hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Pos.non-intest.immunoreg: Annotation here indicates a gene with non-intestinal positively immunoregulated gene expression, as defined by ≥ 2 -fold higher expression in 19-day non-intestinal tissue from normal hosts versus 19-day non-intestinal tissue from dexamethasone-immunosuppressed hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Neg.non-intest.immunoreg: Annotation here indicates a gene with non-intestinal negatively immunoregulated gene expression, as defined by ≥ 2 -fold higher expression in 19-day non-intestinal tissue from dexamethasone-immunosuppressed hosts versus 19-day non-intestinal tissue from normal hosts, with a false discovery rate (FDR) of ≤ 0.01 .

Neg.12d.immunoreg: Annotation here indicates a gene with negatively immunoregulated gene expression in 12-day young adults, as defined by ≥ 2 -fold higher expression in 12-day young adults from dexamethasone-immunosuppressed hosts versus 12-day young adults from normal hosts, with a false discovery rate (FDR) of ≤ 0.01 . (Note that no genes were observed with positively immunoregulated gene expression in 12-day young adults, so no '**Pos.12d.immunoreg**' data column was needed.)

Uzoechi_any_ES: Annotation here indicates a gene which was detected as encoding some sort of ES protein by Uzoechi *et al.* (UZOECHI *et al.* 2023). Practically, this means a gene with any annotation in **Uzoechi_ES** as encoding either a male ES ('Uzoechi_male_ES'), or a female ES ('Uzoechi_female_ES'), or both.

Uzoechi_male.only_ES: Annotation here indicates a gene which was detected as encoding a male-specific ES protein by Uzoechi *et al.* (UZOECHI *et al.* 2023), annotated as solely 'Uzoechi_male_ES' in **Uzoechi_ES**.

Uzoechi_female.only_ES: Annotation here indicates a gene which was detected as encoding a female-specific ES protein by Uzoechi *et al.* (UZOECHI *et al.* 2023), annotated as solely 'Uzoechi_female_ES' in **Uzoechi_ES**.

Uzoechi_both_ES: Annotation here indicates a gene which was detected as encoding an ES protein in both males and females by Uzoechi *et al.* (UZOECHI *et al.* 2023), annotated as both 'Uzoechi_male_ES' and 'Uzoechi_female_ES' in **Uzoechi_ES**.

Gene expression

[X].TPM: For each individual RNA-seq data set (with 'X' denoting the data set's abbreviation), this gives gene expression levels in TPM, computed by Salmon 1.9.0 (PATRO *et al.* 2017). Keys to all abbreviations are given in Supplementary Tables Sx and Sy. Biological replicates of RNA-seq samples are denoted by suffixes such as '_1', '_2', or '_3'.

[X].reads: For each individual RNA-seq data set (with 'X' denoting the data set's abbreviation), this gives numbers of mapped RNA-seq reads per gene, computed for individual RNA-seq data sets by Salmon 1.9.0 (PATRO *et al.* 2017), with fractional values rounded down to integers.

Differential gene expression between biological conditions

[X].vs.[Y].logFC: The fold-changes of gene expression between biological condition X and biological condition Y, expressed as log₂ values, and with positive values representing greater expression in condition X. The values listed here are only those were computed to be significant using edgeR 3.36.0 (ROBINSON *et al.* 2010), with multiple biological RNA-seq replicates for most conditions being compared, with all biological replicates being analyzed in a single edgeR run, and with significant results annotated for individual genes. Biological conditions of RNA-seq samples are abbreviated as shown in **[X].TPM** or **[X].reads** but without the replicate suffixes.

[X].vs.[Y].FDR: The false discovery rate (FDR) for gene expression changes between biological condition X and biological condition Y, annotated for individual genes. The FDR for a given set of positive results is defined as that significance threshold which, if accepted, will lead to the entire set of positives

having a collective false-positive rate no greater than the FDR; it therefore provides a way to correct for testing multiple hypotheses without rejecting excessive numbers of true positives. As with [X].vs.[Y].logFC, only changes that were computed to be significant by edgeR are listed.

Supplementary Table S7. Statistical analyses of the overlaps between *A. ceylanicum* genes encoding protein motifs (or other traits) and *A. ceylanicum* genes encoding ES proteins. Each category of genes (e.g., genes encoding a particular protein motif or having a particular trait) was compared for its degree of overlap to a set of ES genes, and statistically analyzed for the non-randomness of this overlap by a two-tailed Fisher test. In situations where many categories were compared at once (for instance, an entire set of protein motifs from Pfam, InterPro, or Phobius), p-values initially generated by Fisher testing were used to compute q-value significance scores which corrected for multiple hypothesis testing. Statistics are provided both for the set of 565 ES genes described in this paper (labeled "UMass_ES") and for the set of 860 ES genes described by Uzoечи *et al.* (UZOECHEI *et al.* 2023) (labeled "WashU_ES"). For each ES set, spreadsheets are given for analyses of overlaps with the following traits; signal or transmembrane domain organizations predicted by Phobius; protein motifs in Pfam; protein motifs in InterPro; and various binary comparisons with other gene traits ("Various"). Each analysis provides the following data: "Motif" denotes the specific protein motif (or other gene category) for which genes encoding it are being tested for nonrandomly high (or low) overlap with the ES gene set; "All_genes" gives the total number of v2.1 protein-coding genes within which overlaps were tested; "Motif_genes" gives the number of genes annotated with the protein motif (or other trait) being tested for overlap; "Class_genes" gives the number of ES genes being tested for overlap; "Motif.Class_overlap" gives the observed number of genes falling into both categories; "Exp_rand_overlap" gives the number of genes that would be expected to overlap purely randomly; "Enrichment" gives the ratio of observed versus random overlaps (note that this ratio can be lower than 1, and in fact can be as low as 0); "p-value" gives an initial statistical significance for the observed overlap, computed by a two-tailed Fisher test; "q-value" gives, for cases of many comparisons at once (e.g., testing for all Pfam motifs simultaneously), a significance score that conservatively corrects for multiple hypothesis testing (BAILEY *et al.* 2009; NOBLE 2009). Note that q-values were not computed for simple binary comparisons of gene traits listed in "Various", which are annotated for *A. ceylanicum* genes in **Supplementary Table S6**, and defined in its table legend: Intestine-biased; Non-intestine-biased; Pos.intest.immunoreg; Neg.intest.immunoreg; Pos.non-intest.immunoreg; Neg.non-intest.immunoreg; Male-biased; Female-biased; Uzoечи_any_ES; Uzoечи_male.only_ES;

Uzoechi_female.only_ES; and Uzoechi_both_ES. Also, again note that highly significant overlaps can be either higher or lower than the randomly expected genome-wide background rate.

Supplementary Table S8. Annotations, RNA-seq expression, and differential gene expression for the *H. contortus* protein-coding gene set. We annotated predicted protein products with N-terminal signal sequences, conserved protein domains, orthologies to genes in related nematode species, and Gene Ontology (GO) terms describing biological and molecular functions. Protein-coding genes were predicted by Doyle *et al.* (DOYLE *et al.* 2020). Since the annotation methods we used for this *H. contortus* proteome were identical to those we used for our *A. ceylanicum* v2.1 proteome, almost all the data columns used here are equivalent to those in **Supplementary Table S6**. One data column unique to this table is "Hco_ES", which denotes *H. contortus* genes previously demonstrated by Wang *et al.* (WANG *et al.* 2019) to encode ES proteins.

Supplementary Table S9. Annotations for protein-coding gene sets of *A. caninum* and *N. americanus*. We annotated predicted protein products with N-terminal signal sequences, conserved protein domains, orthologies to genes in related nematode species, and Gene Ontology (GO) terms describing biological and molecular functions. Protein-coding genes were predicted for *A. caninum* by Coghlan *et al.* (COGHLAN *et al.* 2019) and for *N. americanus* by Logan *et al.* (LOGAN *et al.* 2020). Since the annotation methods we used for these proteomes were identical to those we used for our *A. ceylanicum* v2.1 proteome, almost all the data columns used here are equivalent to those in **Supplementary Table S6**. Two data columns unique to these tables are **Acan_ES** and **Necator_ES**, which respectively denote *A. caninum* or *N. americanus* genes previously demonstrated by Morante *et al.* (MORANTE *et al.* 2017) or Logan *et al.* (LOGAN *et al.* 2020) to encode ES proteins.

Supplementary Table S10. Statistical analyses of the overlaps between *A. caninum*, *N. americanus*, or *H. contortus* genes encoding protein motifs (or other traits) and *A. caninum*, *N. americanus*, or *H. contortus* genes encoding ES proteins. ES gene annotations are taken from Supplementary Tables S8 and S9. Statistical significances of overlaps between motif/trait genes and ES genes were computed and described as in **Supplementary Table S7**.

Supplementary Table S11. Statistical analyses of overlaps between pairs of *A. ceylanicum* gene sets, with *A. ceylanicum* genes encoding ES proteins versus *A. ceylanicum* genes having various homologies

to either ES genes or any genes in *A. caninum*, *H. contortus*, or *N. americanus*. "ES_a_caninum", "ES_haemonchus", and "ES_necator_rev" denote sets of *A. ceylanicum* genes with homology to ES genes in *A. caninum*, *H. contortus*, or *N. americanus*; "ES_Acan.or.Hco.or.L2020" denotes a set of *A. ceylanicum* genes with homology to ES genes in any of these three species. "Acan_homology", "Haemonchus_homology", and "Necator_homology" denote sets of *A. ceylanicum* genes with homology to any genes (ES protein-encoding, or not) in *A. caninum*, *H. contortus*, or *N. americanus*; "Acan.Hco.Nec_homology" denotes a set of *A. ceylanicum* genes with homology to any genes in any of these three species. The ES gene sets are either from this study ("all_UMass_ES") or from Uzoечи *et al.* (UZOECHI *et al.* 2023) ("all_WashU_ES"). Subsets of these ES gene sets that lack homologies to known ES genes in related parasites are denoted with "non_AcanES_homol_[ES]" (for *A. ceylanicum* ES genes lacking homologies to *A. caninum* ES genes), "non_HcoES_homol_[ES]" (for *A. ceylanicum* ES genes lacking homologies to *H. contortus* ES genes), "non_NecES_homol_[ES]" (for *A. ceylanicum* ES genes lacking homologies to *N. americanus* ES genes), or "non_anyES_homol_[ES]" (for *A. ceylanicum* ES genes lacking homologies to ES genes in any of the three other species). Statistical significances of overlaps between homologous genes and ES genes were computed and described as in **Supplementary Table S7**.

Supplementary Table S12. The frequencies with which RNA-seq reads (either from our new RNA-seq libraries listed in **Supplementary Table S1**, or from previously published RNA-seq libraries listed in **Supplementary Table S4** were mapped to *A. ceylanicum* v2.1 genes by Salmon 1.9.0 (PATRO *et al.* 2017).

Supplementary Table S13. Statistical analyses of the overlaps between *A. ceylanicum* genes encoding protein motifs (or other traits) and *A. ceylanicum* genes with either intestine-biased or non-intestine-biased gene expression. Statistical significances of overlaps between motif/trait genes and intestine-biased or non-intestine-biased genes were computed and described as in **Supplementary Table S7**. In addition, pairwise overlaps between intestine- or non-intestine-biased gene sets and other gene categories are provided in the 'Various' spreadsheet; since these do not involve multiple comparisons of entire motif sets, only p-values rather than q-values are computed for these overlaps. *A. ceylanicum* gene categories in "Various" are: "Intest_haemonchus", genes with homology to *H. contortus* genes with intestine-biased expression; "Non-intest_haemonchus", genes with homology to *H. contortus* genes with non-intestine-biased expression; "WashU-intestine-biased", genes with intestine-biased expression computed from previously published *A. ceylanicum* RNA-seq data; "WashU-non-intestine-biased", genes with non-

intestine-biased expression computed from previously published *A. ceylanicum* RNA-seq data (**Supplementary Table S4**) (WEI *et al.* 2016; BERNOT *et al.* 2020).

Supplementary Table S14. Statistical analyses of the overlaps between *H. contortus* genes encoding protein motifs (or other traits) and *H. contortus* genes with either intestine-biased or non-intestine-biased gene expression. Statistical significances of overlaps between motif/trait genes and intestine-biased or non-intestine-biased genes were computed and described as in **Supplementary Table S7**.

Supplementary Table S15. Statistical analyses of the overlaps between *A. ceylanicum* genes encoding protein motifs (or other traits) and *A. ceylanicum* genes with positively or negatively immunoregulated intestinal or non-intestinal gene expression. Statistical significances of overlaps between motif/trait genes and positively immunoregulated intestinal genes were computed and described as in **Supplementary Table S7**. In addition, pairwise overlaps between immunoregulated gene sets and other categories (such as "Male-biased") are provided in the "Various" spreadsheet; since these do not involve multiple comparisons of entire motif sets, only p-values rather than q-values are computed for these overlaps.

Supplementary Table S16. Statistical analyses of the overlaps between *A. ceylanicum* genes encoding protein motifs (or other traits) and *A. ceylanicum* genes encoding ES proteins that also have positively immunoregulated gene expression in 19-day intestines. Statistical significances of overlaps between motif/trait genes and positively immunoregulated ES genes were computed and described as in **Supplementary Table S7**. In addition, pairwise overlaps between the positively immunoregulated intestinal ES gene set and other categories (such as "Male-biased") are provided in the "Various" spreadsheet; since these do not involve multiple comparisons of entire motif sets, only p-values rather than q-values are computed for these overlaps.

Supplementary Figure 1. Gene expression in *A. ceylanicum*. Gene activity is shown for all 33,190 *A. ceylanicum* genes, with biological replicates on the x-axis and individual genes on the y-axis. Expression levels are in TPM (log₁₀). In **Supp. Fig. 1A**, replicates are ordered as in Figure 2; in **Supp. Fig. 1B**, replicates are ordered by their similarity of expression (and a dendrogram for their similarities is shown along the top x-axis). Genes sharing similar patterns of expression are split into 10 clusters. Biological replicates are: young 12-day hookworm adults from normal hosts (YA_12dpi_noDEX); young 12-day adults from immunosuppressed hosts (YA_12dpi_DEX); dissected 19-day intestines from mature

hookworm adults in normal hosts (intestines_19dpi_noDEX); dissected 19-day intestines in immunosuppressed hosts (intestines_19dpi_DEX); dissected 19-day intestines in normal hosts (non.intestines_19dpi_noDEX); dissected 19-day non-intestinal tissues in immunosuppressed hosts (non.intestines_19dpi_DEX); previously published adult male intestine (WashU_male_intestine); published adult males (WashU_adult_male); and published adult females (WashU_adult_fem). These heatmaps include two biological replicates (19-day nonDEX intestine replicate 2 and 19-day DEX intestine replicate 2) that clustered anomalously with replicates 1 and 3 of intestinal RNA-seq samples with opposite nonDEX versus DEX status. We interpret this to mean that labeling for the nonDEX/DEX status of these samples was accidentally reversed between RNA harvesting and RNA-seq. We thus omitted these two samples from differential gene expression analysis.

Supplementary File 1. An R script used in batch mode for differential gene expression with edgeR 3.36.0 and R 4.1.3.

Data Availability

A. ceylanicum transcriptomic data have been archived in the Sequence Read Archive (KATZ *et al.* 2022) with NCBI BioProject accession PRJNA1045065, and with BioSample accessions SAMN38429478, SAMN38440155, SAMN38440168, SAMN38440169, SAMN38440219, and SAMN38440220. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PEREZ-RIVEROL *et al.* 2022) with the dataset identifiers PXD047871 and PXD047879. Protein-coding gene predictions for *A. ceylanicum* have been archived at the Open Science Framework (<https://osf.io/dxfsb>).

Funding

This work was supported by the National Institutes of Health (NIH)'s National Institute of Allergy and Infectious Diseases (NIAID; <https://www.niaid.nih.gov>) grants 1R21-AI111173 and R01-AI056189 to R.V.A., by NIH's Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD; <https://www.nichd.nih.gov>) grant 1R01-HD099072 to R.V.A., and by Cornell startup funds to E.M.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

We thank Titus Brown and the Michigan State University High-Performance Computing Center (supported by U.S. Department of Agriculture grant 2010-65205-20361 and NIFA-National Science Foundation (NSF) grant IOS-0923812) for computational support; additional computing was enabled by start-up and research allocations from NSF XSEDE (TG-MCB180039 and TG-MCB190010). We also thank the Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology for sequencing and computational support.

References

- Abuzeid, A. M. I., X. Zhou, Y. Huang and G. Li, 2020 Twenty-five-year research progress in hookworm excretory/secretory products. *Parasites & vectors* 13: 136.
- Anderson, R. C., 1984 The origins of zooparasitic nematodes. *Canadian Journal of Zoology* 62: 317-328.
- Anisimov, A. P., and N. P. Tokmakova, 1974 Proliferation and growth of intestinal epithelium in *Ascaris suum* (Nematoda) during postnatal ontogeny. I. Mitotic activity in the intestinal epithelium. *Sov J Dev Biol* 4: 341-348.
- Anisimov, A. P., and L. N. Usheva, 1974 Proliferation and growth of intestinal epithelium in *Ascaris suum* (Nematoda) during postnatal ontogeny. II. Increase in the number and size of cells as a growth factor. *Sov J Dev Biol* 4: 379-383.
- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. *Nature Genetics* 25: 25-29.
- Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant *et al.*, 2009 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37: W202-208.
- Balmer, E. A., and C. Faso, 2021 The road less traveled? Unconventional protein secretion at parasite-host interfaces. *Front Cell Dev Biol* 9: 662711.
- Bansemir, A. D., and M. V. Sukhdeo, 1994 The food resource of adult *Heligmosomoides polygyrus* in the small intestine. *The Journal of parasitology* 80: 24-28.
- Bansemir, A. D., and M. V. Sukhdeo, 2001 Intestinal distribution of worms and host ingesta in *Nippostrongylus brasiliensis*. *J Parasitol* 87: 1470-1472.
- Banyai, L., and L. Patthy, 1998 Amoebapore homologs of *Caenorhabditis elegans*. *Biochim Biophys Acta* 1429: 259-264.
- Barnett, D. W., E. K. Garrison, A. R. Quinlan, M. P. Stromberg and G. T. Marth, 2011 BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27: 1691-1692.
- Beaver, P. C., 1988 Light, long-lasting *Necator* infection in a volunteer. *The American journal of tropical medicine and hygiene* 39: 369-372.
- Behnke, J., and P. D. Harris, 2010 *Heligmosomoides bakeri*: a new name for an old worm? *Trends Parasitol* 26: 524-529.
- Bell, A. D., H. T. Chou, F. Valencia and A. B. Paaby, 2023 Beyond the reference: gene expression variation and transcriptional response to RNA interference in *Caenorhabditis elegans*. *G3 (Bethesda)* 13.
- Berkachy, R., D. J. Smyth, C. Schnoeller, Y. Harcus, R. M. Maizels *et al.*, 2021 Characterisation of the secreted apyrase family of *Heligmosomoides polygyrus*. *Int J Parasitol* 51: 39-48.
- Bernot, J. P., G. Rudy, P. T. Erickson, R. Ratnappan, M. Haile *et al.*, 2020 Transcriptomic analysis of hookworm *Ancylostoma ceylanicum* life cycle stages reveals changes in G-protein coupled receptor diversity associated with the onset of parasitism. *Int J Parasitol* 50: 603-610.
- Bethony, J., S. Brooker, M. Albonico, S. M. Geiger, A. Loukas *et al.*, 2006 Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *Lancet* 367: 1521-1532.
- Bethony, J., A. Loukas, M. Smout, S. Brooker, S. Mendez *et al.*, 2005 Antibodies against a secreted protein from hookworm larvae reduce the intensity of hookworm infection in humans and vaccinated laboratory animals. *FASEB Journal* 19: 1743-1745.
- Bouchery, T., M. Moyat, J. Sotillo, S. Silverstein, B. Volpe *et al.*, 2020 Hookworms evade host immunity by secreting a deoxyribonuclease to degrade neutrophil extracellular traps. *Cell Host Microbe* 27: 277-289 e276.
- Brophy, P. M., L. H. Patterson and D. L. Pritchard, 1995 Secretory nematode SOD--offensive or defensive? *Int J Parasitol* 25: 865-866.
- Brůna, T., K. J. Hoff, A. Lomsadze, M. Stanke and M. Borodovsky, 2021 BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* 3: lqaa108.
- Buchfink, B., K. Reuter and H. G. Drost, 2021 Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18: 366-368.

- Buck, A. H., G. Coakley, F. Simbari, H. J. McSorley, J. F. Quintana *et al.*, 2014 Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat Commun* 5: 5488.
- Caffrey, C. R., L. Goupil, K. M. Rebello, J. P. Dalton and D. Smith, 2018 Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS Negl Trop Dis* 12: e0005840.
- Cantacessi, C., B. E. Campbell, A. Visser, P. Geldhof, M. J. Nolan *et al.*, 2009 A portrait of the "SCP/TAPS" proteins of eukaryotes--developing a framework for fundamental research and biotechnological outcomes. *Biotechnol Adv* 27: 376-388.
- Carbon, S., E. Douglass, B. Good, D. Unni, N. Harris *et al.*, 2021 The Gene Ontology resource: enriching a Gold mine. *Nucleic acids research* 49: D325-D334.
- Chen, L. Q., L. S. Cheung, L. Feng, W. Tanner and W. B. Frommer, 2015 Transport of sugars. *Annu Rev Biochem* 84: 865-894.
- Chen, N., Z. G. Yuan, M. J. Xu, D. H. Zhou, X. X. Zhang *et al.*, 2012 *Ascaris suum* enolase is a potential vaccine candidate against ascariasis. *Vaccine* 30: 3478-3482.
- Chen, S., Y. Zhou, Y. Chen and J. Gu, 2018 fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884-i890.
- Chhabra, S., S. C. Chang, H. M. Nguyen, R. Huq, M. R. Tanner *et al.*, 2014 Kv1.3 channel-blocking immunomodulatory peptides from parasitic worms: implications for autoimmune diseases. *FASEB J*.
- Cho, Y., B. F. Jones, J. J. Vermeire, L. Leng, L. DiFedele *et al.*, 2007 Structural and functional characterization of a secreted hookworm Macrophage Migration Inhibitory Factor (MIF) that interacts with the human MIF receptor CD74. *Journal of Biological Chemistry* 282: 23447-23456.
- Chu, D., R. D. Bungiro, M. Ibanez, L. M. Harrison, E. Campodonico *et al.*, 2004 Molecular characterization of *Ancylostoma ceylanicum* Kunitz-type serine protease inhibitor: evidence for a role in hookworm-associated growth delay. *Infection and Immunity* 72: 2214-2221.
- Clark, W. C., 1994 Origins of the parasitic habit in the nematoda. *Int J Parasitol* 24: 1117-1129.
- Coakley, G., J. L. McCaskill, J. G. Borger, F. Simbari, E. Robertson *et al.*, 2017 Extracellular vesicles from a helminth parasite suppress macrophage activation and constitute an effective vaccine for protective immunity. *Cell reports* 19: 1545-1557.
- Coghlan, A., R. Tyagi, J. A. Cotton, N. Holroyd, B. A. Rosa *et al.*, 2019 Comparative genomics of the major parasitic worms. *Nature genetics* 51: 163-174.
- Cohen, J., 2016 Unfilled vials. *Science* 351: 16-19.
- Cole, R., N. Holroyd, A. Tracey, M. Berriman and M. Viney, 2023 The parasitic nematode *Strongyloides ratti* exists predominantly as populations of long-lived asexual lineages. *Nat Commun* 14: 6427.
- Colella, V., R. Bradbury and R. Traub, 2021 *Ancylostoma ceylanicum*. *Trends Parasitol* 37: 844-845.
- Costa, A. F., R. B. Gasser, S. R. Dias and E. M. Rabelo, 2009 Male-enriched transcription of genes encoding ASPs and Kunitz-type protease inhibitors in *Ancylostoma* species. *Molecular and Cellular Probes* 23: 298-303.
- Crusoe, M. R., H. F. Alameldin, S. Awad, E. Boucher, A. Caldwell *et al.*, 2015 The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* 4: 900.
- Dainat, J., 2023 AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format, pp., Zenodo.
- Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan *et al.*, 2021 Twelve years of SAMtools and BCFtools. *Gigascience* 10.
- Davis, P., M. Zarowiecki, V. Arnaboldi, A. Becerra, S. Cain *et al.*, 2022 WormBase in 2022--data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics* 220.
- Degner, J. F., J. C. Marioni, A. A. Pai, J. K. Pickrell, E. Nkadori *et al.*, 2009 Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* 25: 3207-3212.
- Del Valle, A., B. F. Jones, L. M. Harrison, R. C. Chadderdon and M. Cappello, 2003 Isolation and molecular cloning of a secreted hookworm platelet inhibitor from adult *Ancylostoma caninum*. *Molecular and Biochemical Parasitology* 129: 167-177.
- Dimou, E., and W. Nickel, 2018 Unconventional mechanisms of eukaryotic protein secretion. *Curr Biol* 28: R406-R410.
- Diop, A., D. Santorelli, F. Malagrino, C. Nardella, V. Pennacchietti *et al.*, 2022 SH2 domains: folding, binding and therapeutical approaches. *Int J Mol Sci* 23.

- Doyle, S. R., A. Tracey, R. Laing, N. Holroyd, D. Bartley *et al.*, 2020 Genomic and transcriptomic variation defines the chromosome-scale assembly of *Haemonchus contortus*, a model gastrointestinal worm. *Communications biology* 3: 656.
- Drew, D., R. A. North, K. Nagarathinam and M. Tanabe, 2021 Structures and general transport mechanisms by the major facilitator superfamily (MFS). *Chem Rev* 121: 5289-5335.
- Durette-Desset, M. C., I. Beveridge and D. M. Spratt, 1994 The origins and evolutionary expansion of the strongylida (Nematoda). *International Journal for Parasitology* 24: 1139-1165.
- Ebert, D., and P. D. Fields, 2020 Host-parasite co-evolution and its genomic signature. *Nat Rev Genet* 21: 754-768.
- Eddy, S. R., 2009 A new generation of homology search tools based on probabilistic inference. *Genome Inform* 23: 205-211.
- Emms, D. M., and S. Kelly, 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* 20: 238.
- Fairfax, K. C., J. J. Vermeire, L. M. Harrison, R. D. Bungiro, W. Grant *et al.*, 2009 Characterisation of a fatty acid and retinol binding protein orthologue from the hookworm *Ancylostoma ceylanicum*. *International Journal for Parasitology* 39: 1561-1571.
- Ferguson, A. A., J. M. Inclan-Rico, D. Lu, S. D. Bobardt, L. Hung *et al.*, 2023 Hookworms dynamically respond to loss of Type 2 immune pressure. *PLoS Pathog* 19: e1011797.
- Ferguson, A. A., H. L. Rossi and D. R. Herbert, 2024 The secretome of adult murine hookworms is shaped by host expression of STAT6. *Parasite Immunol* 46: e13056.
- Finn, R. D., P. Coghill, R. Y. Eberhardt, S. R. Eddy, J. Mistry *et al.*, 2016 The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* 44: D279-285.
- Fujiwara, R. T., B. Zhan, S. Mendez, A. Loukas, L. L. Bueno *et al.*, 2007 Reduction of worm fecundity and canine host blood loss mediates protection against hookworm infection elicited by vaccination with recombinant *Ac-16*. *Clin Vaccine Immunol* 14: 281-287.
- Garside, P., and J. M. Behnke, 1989 *Ancylostoma ceylanicum* in the hamster: observations on the host-parasite relationship during primary infection. *Parasitology* 98 Pt 2: 283-289.
- Gems, D., 2000 Longevity and ageing in parasitic and free-living nematodes. *Biogerontology* 1: 289-307.
- Gibbs, G. M., K. Roelants and M. K. O'Bryan, 2008 The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins--roles in reproduction, cancer, and immune defense. *Endocrine Reviews* 29: 865-897.
- Gillis-Germitsch, N., T. Kockmann, L. M. Asmis, L. Tritten and M. Schnyder, 2021 The *Angiostrongylus vasorum* excretory/secretory and surface proteome contains putative modulators of the host coagulation. *Front Cell Infect Microbiol* 11: 753320.
- Gounaris, K., and M. E. Selkirk, 2005 Parasite nucleotide-metabolizing enzymes and host purinergic signalling. *Trends Parasitol* 21: 17-21.
- Grant, R. P., S. M. Buttery, G. C. Ekman, T. M. Roberts and M. Stewart, 2005 Structure of MFP2 and its function in enhancing MSP polymerization in *Ascaris* sperm amoeboid motility. *J Mol Biol* 347: 583-595.
- Gruning, B., R. Dale, A. Sjobin, B. A. Chapman, J. Rowe *et al.*, 2018 Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nature methods* 15: 475-476.
- Gu, Z., 2022 Complex heatmap visualization. *iMeta* 1: e43.
- Guigó, R., 2023 Genome annotation: From human genetics to biodiversity genomics. *Cell Genom* 3: 100375.
- Guiguet, A., G. Dubreuil, M. O. Harris, H. M. Appel, J. C. Schultz *et al.*, 2016 Shared weapons of blood- and plant-feeding insects: Surprising commonalities for manipulating hosts. *J Insect Physiol* 84: 4-21.
- Harcus, Y., G. Nicoll, J. Murray, K. Filbey, N. Gomez-Escobar *et al.*, 2009 C-type lectins from the nematode parasites *Heligmosomoides polygyrus* and *Nippostrongylus brasiliensis*. *Parasitol Int* 58: 461-470.
- Harnett, M. M., A. J. Melendez and W. Harnett, 2010 The therapeutic potential of the filarial nematode-derived immunomodulator, ES-62 in inflammatory disease. *Clin Exp Immunol* 159: 256-267.
- Hart, A. J., S. Ginzburg, M. S. Xu, C. R. Fisher, N. Rahmatpour *et al.*, 2020 EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol Ecol Resour* 20: 591-604.
- Hartmann, S., and R. Lucius, 2003 Modulation of host immune responses by nematode cystatins. *International Journal for Parasitology* 33: 1291-1302.

- Hatje, K., S. Muhlhausen, D. Simm and M. Kollmar, 2019 The protein-coding human genome: annotating high-hanging fruits. *Bioessays* 41: e1900066.
- Hawdon, J. M., B. F. Jones, D. R. Hoffman and P. J. Hotez, 1996 Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. *Journal of Biological Chemistry* 271: 6672-6678.
- He, L., A. M. I. Abuzeid, T. Zhuang, Q. Zhao, S. Zhu *et al.*, 2021 Expression and biological functions of *Ancylostoma ceylanicum* saposin-like protein. *Parasitol Res* 120: 3805-3813.
- Heger, P., M. Kroiher, N. Ndifon and E. Schierenberg, 2010 Conservation of MAP kinase activity and MSP genes in parthenogenetic nematodes. *BMC developmental biology* 10: 51.
- Hernandez-Plaza, A., D. Szklarczyk, J. Botas, C. P. Cantalapiedra, J. Giner-Lamia *et al.*, 2023 eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic Acids Res* 51: D389-D394.
- Hewitson, J. P., Y. Harcus, J. Murray, M. van Agtmaal, K. J. Filbey *et al.*, 2011 Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of venom allergen-like (VAL) proteins. *Journal of proteomics* 74: 1573-1594.
- Hewitson, J. P., A. C. Ivens, Y. Harcus, K. J. Filbey, H. J. McSorley *et al.*, 2013 Secretion of protective antigens by tissue-stage nematode larvae revealed by proteomic analysis and vaccination-induced sterile immunity. *PLoS Pathogens* 9: e1003492.
- Hoagland, K. E., and G. A. Schad, 1978 *Necator americanus* and *Ancylostoma duodenale*: life history parameters and epidemiological implications of two sympatric hookworms of humans. *Experimental Parasitology* 44: 36-49.
- Hoekendorf, A., and M. Leippe, 2012 SPP-3, a saposin-like protein of *Caenorhabditis elegans*, displays antimicrobial and pore-forming activity and is located in the intestine and in one head neuron. *Dev Comp Immunol* 38: 181-186.
- Hoekendorf, A., M. Stanisak and M. Leippe, 2012 The saposin-like protein SPP-12 is an antimicrobial polypeptide in the pharyngeal neurons of *Caenorhabditis elegans* and participates in defence against a natural bacterial pathogen. *Biochem J* 445: 205-212.
- Hong, M., X. Zhou, C. Zeng, D. Xu, T. Xu *et al.*, 2024 Nucleolar stress induces nucleolar stress body formation via the NOSR-1/NUMR-1 axis in *Caenorhabditis elegans*. *Nat Commun* 15: 7256.
- Hotez, P. J., S. Narasimhan, J. Haggerty, L. Milstone, V. Bhopale *et al.*, 1992 Hyaluronidase from infective *Ancylostoma* hookworm larvae and its possible function as a virulence factor in tissue invasion and in cutaneous larva migrans. *Infect Immun* 60: 1018-1023.
- Howe, K. L., B. J. Bolt, M. Shafie, P. Kersey and M. Berriman, 2017 WormBase ParaSite - a comprehensive resource for helminth genomics. *Molecular and biochemical parasitology* 215: 2-10.
- Hu, Y., B. Zhan, B. Keegan, Y. Y. Yiu, M. M. Miller *et al.*, 2012 Mechanistic and single-dose *in vivo* therapeutic studies of Cry5B anthelmintic action against hookworms. *PLoS Negl Trop Dis* 6: e1900.
- Huang, Y., A. M. I. Abuzeid, Y. Liu, L. He, Q. Zhao *et al.*, 2020 Identification and localization of hookworm platelet inhibitor in *Ancylostoma ceylanicum*. *Infect Genet Evol* 77: 104102.
- Hunt, V. L., I. J. Tsai, M. E. Selkirk and M. Viney, 2017 The genome of *Strongyloides* spp. gives insights into protein families with a putative role in nematode parasitism. *Parasitology* 144: 343-358.
- Irvine, A., S. A. Huws, L. E. Atkinson and A. Mousley, 2023 Exploring the antimicrobial peptidome of nematodes through phylum-spanning *in silico* analyses highlights novel opportunities for pathogen control. *PLoS Negl Trop Dis* 17: e0011618.
- Jenkins, R. E., M. J. Taylor, N. Gilvary and A. E. Bianco, 1996 Characterization of a secreted antigen of *Onchocerca volvulus* with host-protective potential. *Parasite Immunol* 18: 29-42.
- Jex, A. R., P. Nejsum, E. M. Schwarz, L. Hu, N. D. Young *et al.*, 2014 Genome and transcriptome of the porcine whipworm *Trichuris suis*. *Nature genetics* 46: 701-706.
- Jia, T. W., S. Melville, J. Utzinger, C. H. King and X. N. Zhou, 2012 Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. *PLoS neglected tropical diseases* 6: e1621.
- Jung, S., F. D. Sonnichsen, C. W. Hung, A. Tholey, C. Boidin-Wichlacz *et al.*, 2012 Macin family of antimicrobial proteins combines antimicrobial and nerve repair activities. *J Biol Chem* 287: 14246-14258.
- Käll, L., A. Krogh and E. L. Sonnhammer, 2004 A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* 338: 1027-1036.

- Kanehisa, M., M. Furumichi, Y. Sato, M. Kawashima and M. Ishiguro-Watanabe, 2023 KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 51: D587-D592.
- Katz, K., O. Shutov, R. Lapoint, M. Kimelman, J. R. Brister *et al.*, 2022 The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res* 50: D387-D390.
- Keiser, J., and J. Utzinger, 2008 Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. *JAMA : the journal of the American Medical Association* 299: 1937-1948.
- Klotz, C., T. Ziegler, A. S. Figueiredo, S. Rausch, M. R. Hepworth *et al.*, 2011 A helminth immunomodulator exploits host signaling events to regulate cytokine production in macrophages. *PLoS pathogens* 7: e1001248.
- Knox, D., 2012 Proteases as vaccines against gastrointestinal nematode parasites of sheep and cattle, pp. 399-420 in *Parasitic Helminths: Targets, Screens, Drugs, and Vaccines*, edited by C. R. Caffrey. Wiley-VCH Verlag & Co. KGaA., Weinheim, Germany.
- Knox, D. P., 2007 Proteinase inhibitors and helminth parasite infection. *Parasite Immunol* 29: 57-71.
- Knox, D. P., and D. G. Jones, 1992 A comparison of superoxide dismutase (SOD, EC:1.15.1.1) distribution in gastro-intestinal nematodes. *Int J Parasitol* 22: 209-214.
- Kokot, T., and M. Kohn, 2022 Emerging insights into serine/threonine-specific phosphoprotein phosphatase function and selectivity. *J Cell Sci* 135.
- Kvist, S., A. Manzano-Marin, D. de Carle, P. Trontelj and M. E. Siddall, 2020 Draft genome of the European medicinal leech *Hirudo medicinalis* (Annelida, Clitellata, Hirudiniformes) with emphasis on anticoagulants. *Sci Rep* 10: 9885.
- Laing, R., T. Kikuchi, A. Martinelli, I. J. Tsai, R. N. Beech *et al.*, 2013 The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome biology* 14: R88.
- Lee, R. Y. N., K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain *et al.*, 2017 WormBase 2017: molting into a new stage. *Nucleic acids research* 46: D869-D874.
- Li, W., and A. Godzik, 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658-1659.
- Lightowlers, M. W., and M. D. Rickard, 1988 Excretory-secretory products of helminth parasites: effects on host immune responses. *Parasitology* 96 Suppl: S123-166.
- Lo, S. K., A. Rahman, N. Xu, M. Y. Zhou, P. Nagpala *et al.*, 1999 Neutrophil inhibitory factor abrogates neutrophil adhesion by blockade of CD11a and CD11b beta(2) integrins. *Mol Pharmacol* 56: 926-932.
- Loeb, L., and A. J. Smith, 1904 The presence of a substance inhibiting the coagulation of the blood in *Anchylostoma*. *Proceedings of the Pathological Society of Philadelphia* 7: 173-178.
- Logan, J., M. S. Pearson, S. S. Manda, Y. J. Choi, M. Field *et al.*, 2020 Comprehensive analysis of the secreted proteome of adult *Necator americanus* hookworms. *PLoS neglected tropical diseases* 14: e0008237.
- Lomsadze, A., P. D. Burns and M. Borodovsky, 2014 Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* 42: e119.
- Loukas, A., A. Doedens, M. Hintz and R. M. Maizels, 2000 Identification of a new C-type lectin, TES-70, secreted by infective larvae of *Toxocara canis*, which binds to host ligands. *Parasitology* 121 Pt 5: 545-554.
- Loukas, A., P. J. Hotez, D. Diemert, M. Yazdanbakhsh, J. S. McCarthy *et al.*, 2016 Hookworm infection. *Nature reviews. Disease primers* 2: 16088.
- Lupas, A., 1996 Prediction and analysis of coiled-coil structures. *Methods in Enzymology* 266: 513-525.
- Maizels, R. M., 2020 Regulation of immunity and allergy by helminth parasites. *Allergy* 75: 524-534.
- Maizels, R. M., J. P. Hewitson and W. C. Gause, 2011 *Heligmosomoides polygyrus*: one species still. *Trends in parasitology* 27: 100-101.
- Maizels, R. M., H. H. Smits and H. J. McSorley, 2018 Modulation of host immunity by helminths: the expanding repertoire of parasite effector molecules. *Immunity* 49: 801-818.
- Manoury, B., W. F. Gregory, R. M. Maizels and C. Watts, 2001 *Bm*-CPI-2, a cystatin homolog secreted by the filarial parasite *Brugia malayi*, inhibits class II MHC-restricted antigen processing. *Current biology : CB* 11: 447-451.
- Maruszewska-Cheruiyot, M., L. Szewczak, K. Krawczak-Wojcik, M. Kierasinska, M. Stear *et al.*, 2023 The impact of intestinal inflammation on nematode's excretory-secretory proteome. *Int J Mol Sci* 24.

- Matouskova, P., I. Vokral, J. Lamka and L. Skalova, 2016 The role of xenobiotic-metabolizing enzymes in anthelmintic deactivation and resistance in helminths. *Trends Parasitol* 32: 481-491.
- McCarthy, J. S., M. Wieseman, J. Tropea, D. Kaslow, D. Abraham *et al.*, 2002 *Onchocerca volvulus* glycolytic enzyme fructose-1,6-bisphosphate aldolase as a target for a protective immune response in humans. *Infect Immun* 70: 851-858.
- McNeilly, T. N., D. Frew, S. T. G. Burgess, H. Wright, D. J. Bartley *et al.*, 2017 Niche-specific gene expression in a parasitic nematode; increased expression of immunomodulators in *Teladorsagia circumcincta* larvae derived from host mucosa. *Sci Rep* 7: 7214.
- Mistry, J., S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar *et al.*, 2021 Pfam: The protein families database in 2021. *Nucleic Acids Res* 49: D412-D419.
- Morante, T., C. Shepherd, C. Constantinoiu, A. Loukas and J. Sotillo, 2017 Revisiting the *Ancylostoma caninum* secretome provides new information on hookworm-host interactions. *Proteomics* 17.
- Moreno, Y., P. P. Gros, M. Tam, M. Segura, R. Valanparambil *et al.*, 2011 Proteomic analysis of excretory-secretory products of *Heligmosomoides polygyrus* assessed with next-generation sequencing transcriptomic information. *PLoS Negl Trop Dis* 5: e1370.
- Moyle, M., D. L. Foster, D. E. McGrath, S. M. Brown, Y. Laroche *et al.*, 1994 A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *The Journal of biological chemistry* 269: 10008-10015.
- Neph, S., M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman *et al.*, 2012 BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28: 1919-1920.
- Nisbet, A. J., T. N. McNeilly, D. R. G. Price, E. M. Oliver, Y. Bartley *et al.*, 2019 The rational simplification of a recombinant cocktail vaccine to control the parasitic nematode *Teladorsagia circumcincta*. *Int J Parasitol* 49: 257-265.
- Noble, W. S., 2009 How does multiple testing correction work? *Nature Biotechnology* 27: 1135-1137.
- Noon, J. B., E. M. Schwarz, G. R. Ostroff and R. V. Aroian, 2019 A highly expressed intestinal cysteine protease of *Ancylostoma ceylanicum* protects vaccinated hamsters from hookworm infection. *PLoS neglected tropical diseases* 13: e0007345.
- O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufu, D. Haddad *et al.*, 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44: D733-745.
- Oliveira, P. L., and F. A. Genta, 2021 Blood digestion in triatomine insects in *Triatominae - The Biology of Chagas Disease Vectors*, edited by A. Guarneri and M. Lorenzo. Springer, Cham.
- Orr, A. R., J. E. Quagrain, P. Suwondo, S. George, L. M. Harrison *et al.*, 2019 Genetic markers of benzimidazole resistance among human hookworms (*Necator americanus*) in Kintampo North Municipality, Ghana. *The American journal of tropical medicine and hygiene* 100: 351-356.
- Ow, M. C., M. A. Nishiguchi, A. R. Dar, R. A. Butcher and S. E. Hall, 2024 RNAi-dependent expression of sperm genes in ADL chemosensory neurons is required for olfactory responses in *Caenorhabditis elegans*. *Front Mol Biosci* 11: 1396587.
- Pala, Z. R., E. S. T. L. Alves, M. Minai, B. Crews, E. Patino-Martinez *et al.*, 2023 *Anopheles* salivary apyrase regulates blood meal hemostasis and drives malaria parasite transmission. *bioRxiv*.
- Parks, S. C., S. Nguyen, S. Nasrolahi, C. Bhat, D. Juncaj *et al.*, 2021 Parasitic nematode fatty acid- and retinol-binding proteins compromise host immunity by interfering with host lipid signaling pathways. *PLoS Pathog* 17: e1010027.
- Parks, S. C., O. K. Okakpu, P. Azizpor, S. Nguyen, S. Martinez-Beltran *et al.*, 2023 Parasitic nematode secreted phospholipase A(2) suppresses cellular and humoral immunity by targeting hemocytes in *Drosophila melanogaster*. *Front Immunol* 14: 1122451.
- Patro, R., G. Duggal, M. I. Love, R. A. Irizarry and C. Kingsford, 2017 Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* 14: 417-419.
- Paysan-Lafosse, T., M. Blum, S. Chuguransky, T. Grego, B. L. Pinto *et al.*, 2023 InterPro in 2022. *Nucleic Acids Res* 51: D418-D427.
- Pearson, M. S., L. Tribolet, C. Cantacessi, M. V. Periago, M. A. Valero *et al.*, 2012 Molecular mechanisms of hookworm disease: stealth, virulence, and vaccines. *Journal of Allergy and Clinical Immunology* 130: 13-21.

- Perez-Riverol, Y., J. Bai, C. Bandla, D. Garcia-Seisdedos, S. Hewapathirana *et al.*, 2022 The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50: D543-D552.
- Pertea, G., X. Huang, F. Liang, V. Antonescu, R. Sultana *et al.*, 2003 TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19: 651-652.
- Pertea, G., and M. Pertea, 2020 GFF utilities: GffRead and GffCompare. *F1000Res* 9.
- Pollo, S. M. J., H. Liu, A. L. Coria, N. Rosin, E. Labit *et al.*, 2024 A single-cell transcriptomics atlas for the parasitic nematode *Heligmosomoides bakeri*: Extrapolating model organism information to non-model systems. *bioRxiv*.
- Price, D. R. G., A. J. Nisbet, D. Frew, Y. Bartley, E. M. Oliver *et al.*, 2019 Characterisation of a niche-specific excretory-secretory peroxiredoxin from the parasitic nematode *Teladorsagia circumcincta*. *Parasit Vectors* 12: 339.
- Price, D. R. G., P. Steele, D. Frew, K. McLean, D. Androsuk *et al.*, 2024 Characterisation of protective vaccine antigens from the thiol-containing components of excretory/secretory material of *Ostertagia ostertagi*. *Vet Parasitol* 328: 110154.
- Pritchard, D. I., 1986 Antigens of gastrointestinal nematodes. *Trans R Soc Trop Med Hyg* 80: 728-734.
- Pullan, R. L., J. L. Smith, R. Jasrasaria and S. J. Brooker, 2014 Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites & vectors* 7: 37.
- Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842.
- Ranjit, N., B. Zhan, B. Hamilton, D. Stenzel, J. Lowther *et al.*, 2009 Proteolytic degradation of hemoglobin in the intestine of the human hookworm *Necator americanus*. *Journal of Infectious Diseases* 199: 904-912.
- Ray, D. K., K. K. Bhopale and V. B. Shrivastava, 1972 Migration and growth of *Ancylostoma ceylanicum* in golden hamsters *Mesocricetus auratus*. *Journal of Helminthology* 46: 357-362.
- Raza, A., A. R. Williams and M. M. Abeer, 2023 Importance of ABC transporters in the survival of parasitic nematodes and the prospect for the development of novel control strategies. *Pathogens* 12.
- Rezansoff, A. M., R. Laing, A. Martinelli, S. Stasiuk, E. Redman *et al.*, 2019 The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode *Haemonchus contortus*. *International journal for parasitology* 49: 847-858.
- 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.
- Rödelsperger, C., A. Ebbing, D. R. Sharma, M. Okumura, R. J. Sommer *et al.*, 2021 Spatial transcriptomics of nematodes identifies sperm cells as a source of genomic novelty and rapid evolution. *Mol Biol Evol* 38: 229-243.
- Roeder, T., M. Stanisak, C. Gelhaus, I. Bruchhaus, J. Grotzinger *et al.*, 2010 Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. *Dev Comp Immunol* 34: 203-209.
- Romeo, Y., and B. Lemaitre, 2008 *Drosophila* immunity: methods for monitoring the activity of Toll and Imd signaling pathways. *Methods in Molecular Biology* 415: 379-394.
- Ryan, S. M., R. Ruscher, W. A. Johnston, D. A. Pickering, M. W. Kennedy *et al.*, 2022 Novel antiinflammatory biologics shaped by parasite-host coevolution. *Proc Natl Acad Sci U S A* 119: e2202795119.
- Scanlan, J. L., P. Battlay and C. Robin, 2022 Ecdysteroid kinase-like (EcKL) paralogs confer developmental tolerance to caffeine in *Drosophila melanogaster*. *Curr Res Insect Sci* 2: 100030.
- Schmitz, C., P. Kinge and H. Hutter, 2007 Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive *Caenorhabditis elegans* strain *nre-1(hd20) lin-15b(hd126)*. *Proceedings of the National Academy of Sciences of the United States of America* 104: 834-839.
- Schurch, N. J., P. Schofield, M. Gierlinski, C. Cole, A. Sherstnev *et al.*, 2016 How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22: 839-851.
- Schwarz, E. M., Y. Hu, I. Antoshechkin, M. M. Miller, P. W. Sternberg *et al.*, 2015 The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. *Nature genetics* 47: 416-422.

- Schwarz, E. M., M. Kato and P. W. Sternberg, 2012 Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America 109: 16246-16251.
- Shinoda, K., A. Choe, K. Hirahara, M. Kiuchi, K. Kokubo *et al.*, 2022 Nematode ascarosides attenuate mammalian type 2 inflammatory responses. Proc Natl Acad Sci U S A 119.
- Shumate, A., and S. L. Salzberg, 2020 Liftoff: accurate mapping of gene annotations. Bioinformatics 37: 1639-1643.
- Smallwood, T. B., S. Navarro, B. Cristofori-Armstrong, T. S. Watkins, K. Tungatt *et al.*, 2021 Synthetic hookworm-derived peptides are potent modulators of primary human immune cell function that protect against experimental colitis *in vivo*. J Biol Chem 297: 100834.
- Smith, H. E., 2014 Nematode sperm motility. WormBook: 1-15.
- Sotillo, J., A. Sanchez-Flores, C. Cantacessi, Y. Harcus, D. Pickering *et al.*, 2014 Secreted proteomes of different developmental stages of the gastrointestinal nematode *Nippostrongylus brasiliensis*. Mol Cell Proteomics 13: 2736-2751.
- Sparkes, A., P. De Baetselier, K. Roelants, C. De Trez, S. Magez *et al.*, 2017 The non-mammalian MIF superfamily. Immunobiology 222: 473-482.
- 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.
- Srivastava, A., L. Malik, H. Sarkar, M. Zakeri, F. Almodaresi *et al.*, 2020 Alignment and mapping methodology influence transcript abundance estimation. Genome Biol 21: 239.
- Srivastava, A., M. Zakeri, H. Sarkar, C. Sonesson, C. Kingsford *et al.*, 2021 Accounting for fragments of unexpected origin improves transcript quantification in RNA-seq simulations focused on increased realism. bioRxiv.
- 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.
- Stassens, P., P. W. Bergum, Y. Gansemans, L. Jespers, Y. Laroche *et al.*, 1996 Anticoagulant repertoire of the hookworm *Ancylostoma caninum*. Proceedings of the National Academy of Sciences of the United States of America 93: 2149-2154.
- Stevens, L., I. Martinez-Ugalde, E. King, M. Wagah, D. Absolon *et al.*, 2023 Ancient diversity in host-parasite interaction genes in a model parasitic nematode. Nat Commun 14: 7776.
- Stirewalt, M. A., 1963 Chemical biology of secretions of larval helminths. Ann N Y Acad Sci 113: 36-53.
- Stracke, K., A. R. Jex and R. J. Traub, 2020 Zoonotic ancylostomiasis: an update of a continually neglected zoonosis. The American journal of tropical medicine and hygiene 103: 64-68.
- Tang, Y. T., X. Gao, B. A. Rosa, S. Abubucker, K. Hallsworth-Pepin *et al.*, 2014 Genome of the human hookworm *Necator americanus*. Nature Genetics 46: 261-269.
- Taylor, S. S., and A. P. Kornev, 2011 Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem Sci 36: 65-77.
- Tellier, A., S. Moreno-Gamez and W. Stephan, 2014 Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. Evolution 68: 2211-2224.
- Thomas, J. H., and H. M. Robertson, 2008 The *Caenorhabditis* chemoreceptor gene families. BMC biology 6: 42.
- Tian, X., M. Lu, C. Jia, Y. Bu, K. Aimulajiang *et al.*, 2020 *Haemonchus contortus* transthyretin domain - containing protein (*HcTTR*): A promising vaccine candidate against *Haemonchus contortus* infection. Vet Parasitol 279: 109045.
- Tian, X., M. Lu, W. Wang, C. Jia, E. Muhammad *et al.*, 2019 *HcTTR*: a novel antagonist against goat interleukin 4 derived from the excretory and secretory products of *Haemonchus contortus*. Vet Res 50: 42.
- Traub, R. J., 2013 *Ancylostoma ceylanicum*, a re-emerging but neglected parasitic zoonosis. International Journal for Parasitology 43: 1009-1015.
- Tribolet, L., C. Cantacessi, D. A. Pickering, S. Navarro, D. L. Doolan *et al.*, 2015 Probing of a human proteome microarray with a recombinant pathogen protein reveals a novel mechanism by which hookworms suppress B-cell receptor signaling. The Journal of infectious diseases 211: 416-425.
- Tritten, L., C. Ballesteros, R. Beech, T. G. Geary and Y. Moreno, 2021 Mining nematode protein secretomes to explain lifestyle and host specificity. PLoS Negl Trop Dis 15: e0009828.

- Tritten, L., U. Nwosu, M. Vargas and J. Keiser, 2012 In vitro and in vivo efficacy of tribendimidine and its metabolites alone and in combination against the hookworms *Heligmosomoides bakeri* and *Ancylostoma ceylanicum*. *Acta Tropica* 122: 101-107.
- Tvermoes, B. E., W. A. Boyd and J. H. Freedman, 2010 Molecular characterization of *numr-1* and *numr-2*: genes that increase both resistance to metal-induced stress and lifespan in *Caenorhabditis elegans*. *J Cell Sci* 123: 2124-2134.
- UniProt, C., 2023 UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 51: D523-D531.
- Uzoechi, S. C., B. A. Rosa, K. S. Singh, Y. J. Choi, B. K. Bracken *et al.*, 2023 Excretory/secretory proteome of females and males of the hookworm *Ancylostoma ceylanicum*. *Pathogens* 12.
- van Megen, H., S. van den Elsen, M. Holterman, G. Karssen, P. Mooyman *et al.*, 2009 A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 11: 927-950.
- Venkatesan, A., P. D. Jimenez Castro, A. Morosetti, H. Horvath, R. Chen *et al.*, 2023 Molecular evidence of widespread benzimidazole drug resistance in *Ancylostoma caninum* from domestic dogs throughout the USA and discovery of a novel beta-tubulin benzimidazole resistance mutation. *PLoS Pathog* 19: e1011146.
- Vlaminck, J., P. Cools, M. Albonico, S. Ame, M. Ayana *et al.*, 2019 Therapeutic efficacy of albendazole against soil-transmitted helminthiasis in children measured by five diagnostic methods. *PLoS Negl Trop Dis* 13: e0007471.
- Walker, M., P. Cools, M. Albonico, S. M. Ame, M. Ayana *et al.*, 2021 Individual responses to a single oral dose of albendazole indicate reduced efficacy against soil-transmitted helminths in an area with high drug pressure. *PLoS Negl Trop Dis* 15: e0009888.
- Wang, H., G. C. Paesen, P. A. Nuttall and A. G. Barbour, 1998 Male ticks help their mates to feed. *Nature* 391: 753-754.
- Wang, J., M. M. Barr and A. M. Wehman, 2024 Extracellular vesicles. *Genetics* 227.
- Wang, T., G. Ma, C. S. Ang, P. K. Korhonen, A. V. Koehler *et al.*, 2019 High throughput LC-MS/MS-based proteomic analysis of excretory-secretory products from short-term in vitro culture of *Haemonchus contortus*. *J Proteomics* 204: 103375.
- Wang, Z., Q. Q. Lu, M. M. Weng, Y. L. Li, L. L. Han *et al.*, 2023 Binding of *Trichinella spiralis* C-type lectin with syndecan-1 on intestinal epithelial cells mediates larval invasion of intestinal epithelium. *Vet Res* 54: 86.
- Waterhouse, R. M., M. Seppey, F. A. Simao, M. Manni, P. Ioannidis *et al.*, 2018 BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution* 35: 543-548.
- Wei, J., A. Damania, X. Gao, Z. Liu, R. Mejia *et al.*, 2016 The hookworm *Ancylostoma ceylanicum* intestinal transcriptome provides a platform for selecting drug and vaccine candidates. *Parasites & vectors* 9: 518.
- Wilbers, R. H. P., R. Schneiter, M. H. M. Holterman, C. Drurey, G. Smant *et al.*, 2018 Secreted venom allergen-like proteins of helminths: Conserved modulators of host responses in animals and plants. *PLoS Pathog* 14: e1007300.
- Williamson, A. L., S. Lustigman, Y. Oksov, V. Deumic, J. Plieskatt *et al.*, 2006 *Ancylostoma caninum* MTP-1, an astacin-like metalloprotease secreted by infective hookworm larvae, is involved in tissue migration. *Infect Immun* 74: 961-967.
- Wisniewski, M., S. Jaros, P. Baska, M. Cappello, E. Dlugosz *et al.*, 2016 Hamsters vaccinated with *Ace-mep-7* DNA vaccine produced protective immunity against *Ancylostoma ceylanicum* infection. *Exp Parasitol* 163: 1-7.
- Wisniewski, M., S. Jaros, P. Baska, M. Cappello and H. Wedrychowicz, 2013 *Ancylostoma ceylanicum* metalloprotease 6 DNA vaccination induces partial protection against hookworm challenge infection. *Acta Parasitol* 58: 376-383.
- Wootton, J. C., 1994 Non-globular domains in protein sequences: automated segmentation using complexity measures. *Computers and Chemistry* 18: 269-285.
- Wu, C. W., K. Wimberly, A. Pietras, W. Dodd, M. B. Atlas *et al.*, 2019 RNA processing errors triggered by cadmium and integrator complex disruption are signals for environmental stress. *BMC Biol* 17: 56.
- Xiao, S., B. Zhan, J. Xue, G. N. Goud, A. Loukas *et al.*, 2008 The evaluation of recombinant hookworm antigens as vaccines in hamsters (*Mesocricetus auratus*) challenged with human hookworm, *Necator americanus*. *Experimental Parasitology* 118: 32-40.

- Xie, Y., E. P. Hoberg, Z. Yang, J. F. Urban, Jr. and G. Yang, 2017 *Ancylostoma ailuopodae* n. sp. (Nematoda: Ancylostomatidae), a new hookworm parasite isolated from wild giant pandas in Southwest China. *Parasit Vectors* 10: 277.
- Yadav, S., P. Sharma, A. Sharma, L. Ganga, J. K. Saxena *et al.*, 2021 Immunization with *Brugia malayi* calreticulin protein generates robust antiparasitic immunity and offers protection during experimental lymphatic filariasis. *ACS Infect Dis* 7: 790-799.
- Yang, X., S. Khan, X. Zhao, J. Zhang, A. Nisar *et al.*, 2020 Suppression of hyaluronidase reduces invasion and establishment of *Haemonchus contortus* larvae in sheep. *Vet Res* 51: 106.
- Yang, Y., Y. Wen, Y. N. Cai, I. Vallee, P. Boireau *et al.*, 2015 Serine proteases of parasitic helminths. *Korean J Parasitol* 53: 1-11.
- Yatsuda, A. P., M. Eysker, M. C. Vieira-Bressan and E. De Vries, 2002 A family of activation associated secreted protein (ASP) homologues of *Cooperia punctata*. *Res Vet Sci* 73: 297-306.
- Yoshida, A., E. Nagayasu, Y. Horii and H. Maruyama, 2012 A novel C-type lectin identified by EST analysis in tissue migratory larvae of *Ascaris suum*. *Parasitology Research* 110: 1583-1586.
- Zawawi, A., and K. J. Else, 2020 Soil-transmitted helminth vaccines: are we getting closer? *Frontiers in immunology* 11: 576748.
- Zhan, B., P. J. Hotez, Y. Wang and J. M. Hawdon, 2002 A developmentally regulated metalloprotease secreted by host-stimulated *Ancylostoma caninum* third-stage infective larvae is a member of the astacin family of proteases. *Mol Biochem Parasitol* 120: 291-296.
- Zhan, B., Y. Liu, M. Badamchian, A. Williamson, J. Feng *et al.*, 2003 Molecular characterisation of the *Ancylostoma*-secreted protein family from the adult stage of *Ancylostoma caninum*. *International Journal for Parasitology* 33: 897-907.
- Zhan, B., S. Perally, P. M. Brophy, J. Xue, G. Goud *et al.*, 2010 Molecular cloning, biochemical characterization, and partial protective immunity of the heme-binding glutathione S-transferases from the human hookworm *Necator americanus*. *Infection and Immunity* 78: 1552-1563.
- Zhang, Q., J. Pell, R. Canino-Koning, A. C. Howe and C. T. Brown, 2014 These are not the k-mers you are looking for: efficient online k-mer counting using a probabilistic data structure. *PLoS One* 9: e101271.

Figure 1

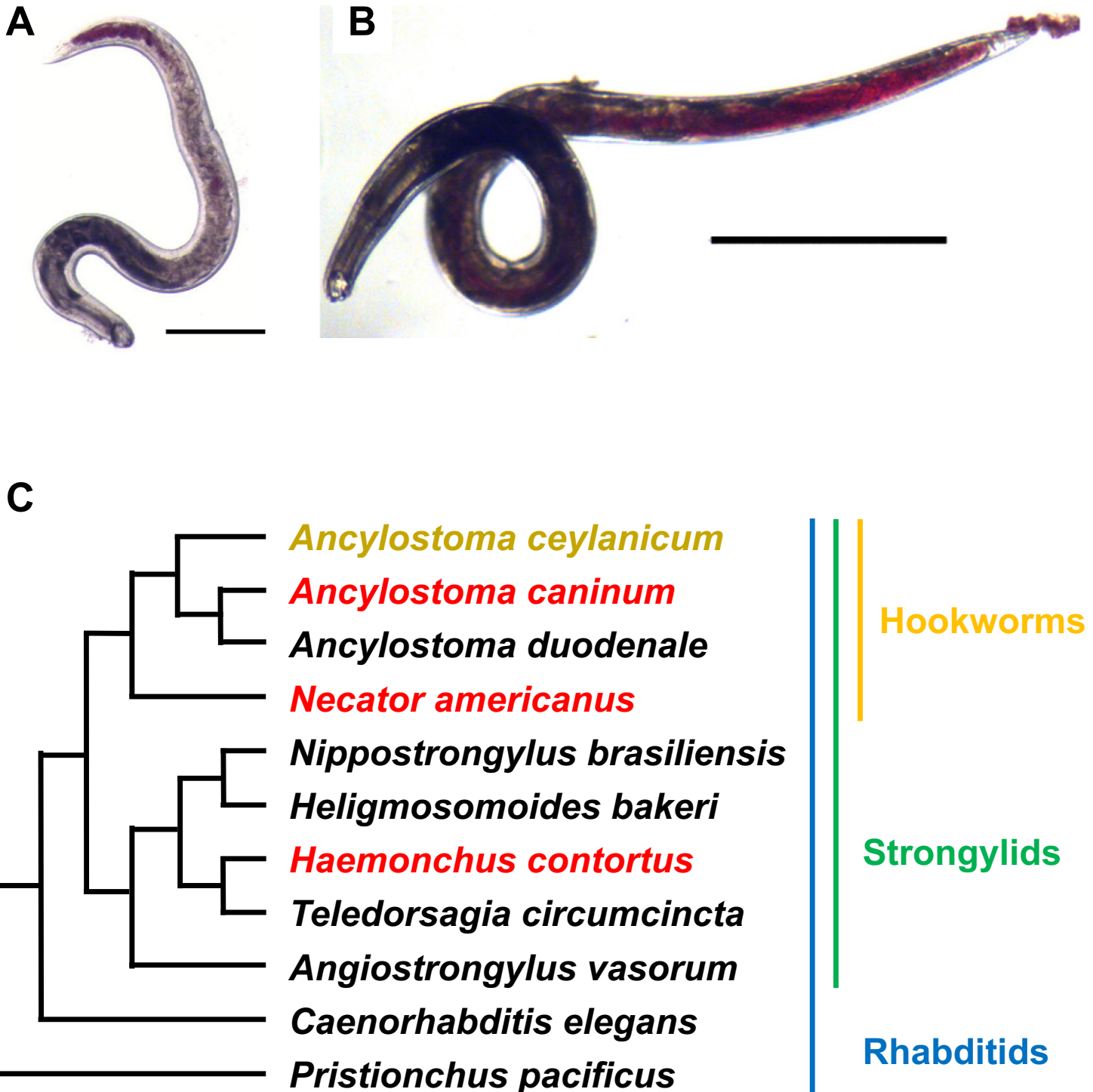
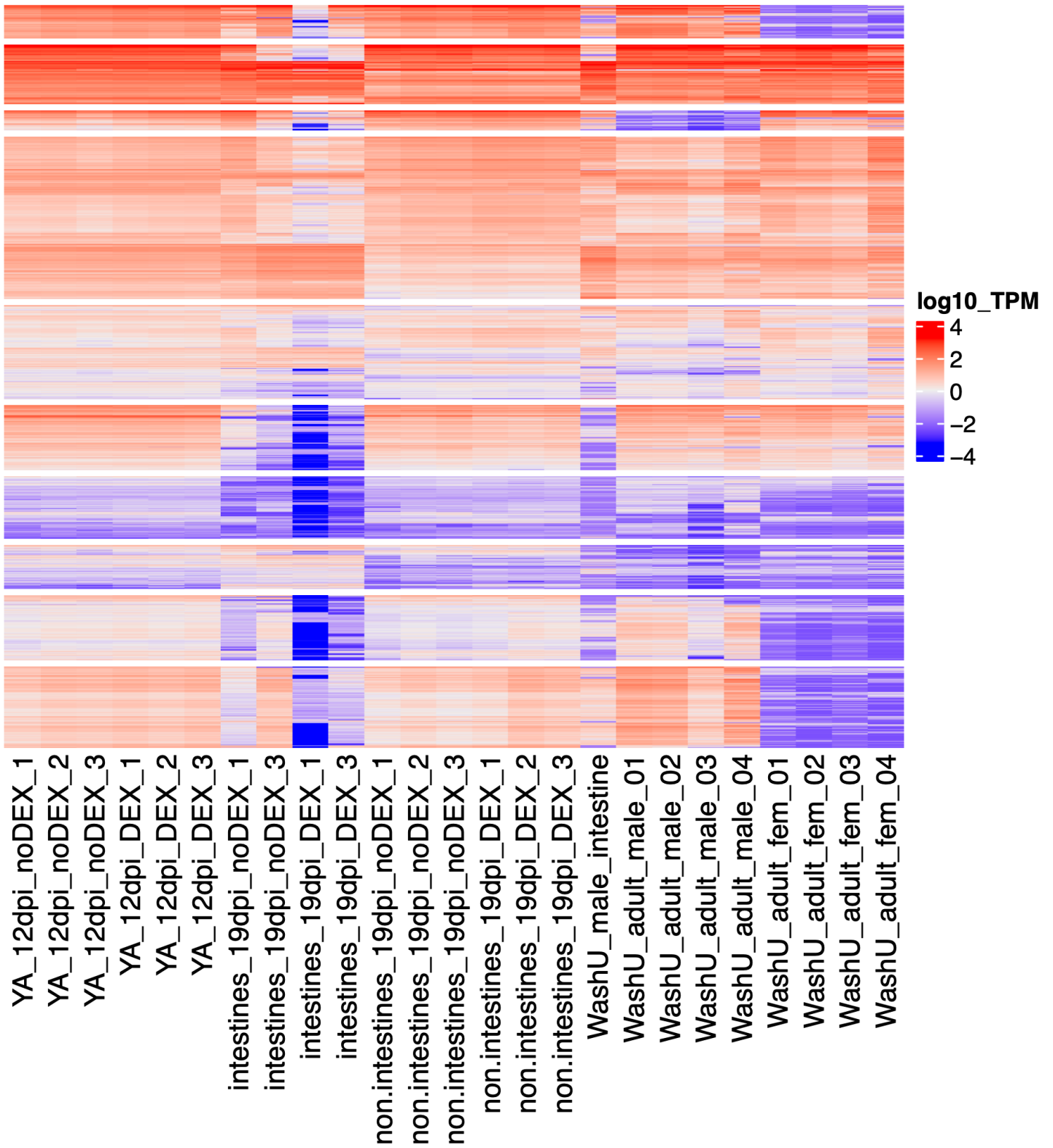
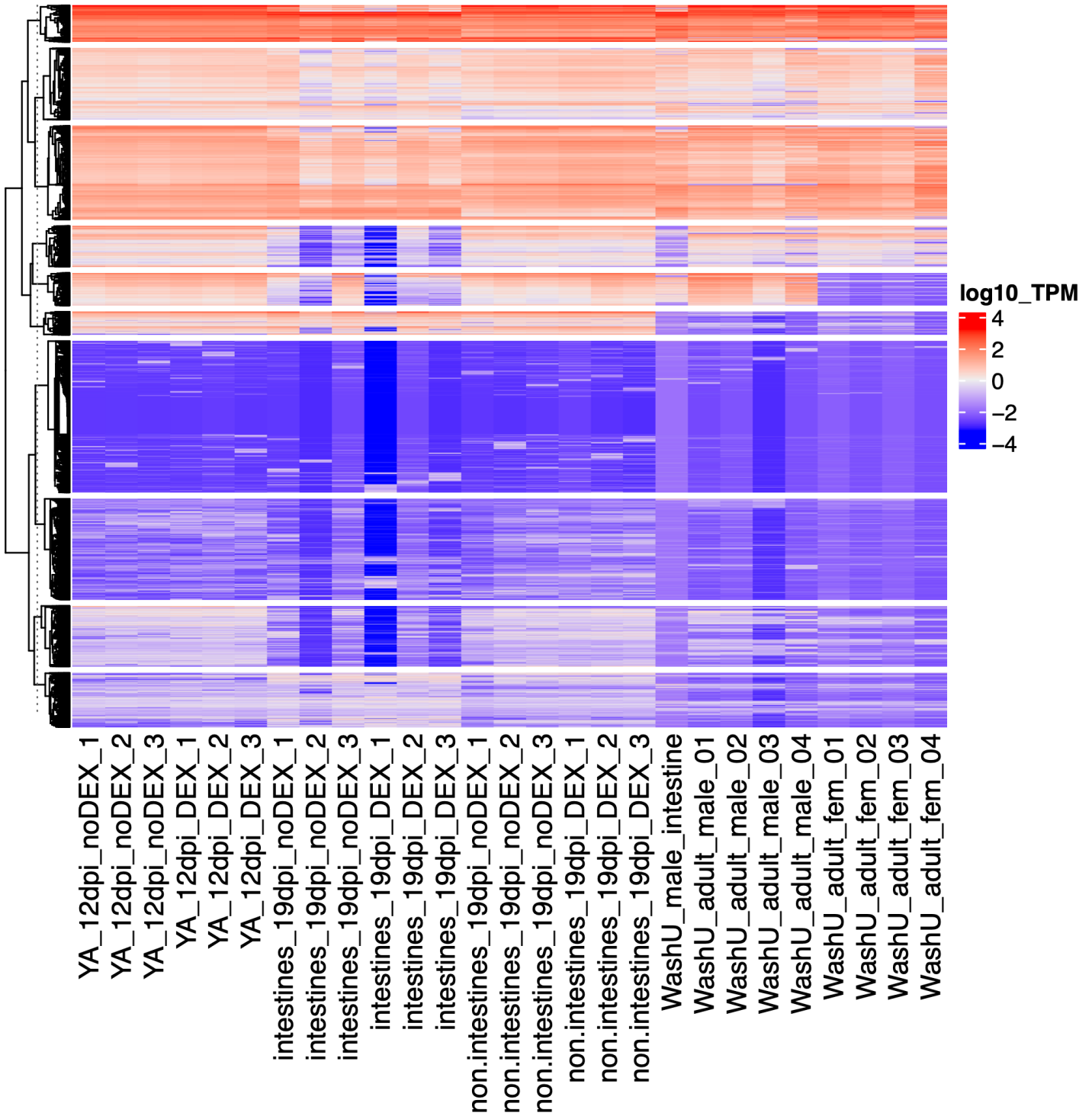


Figure 2



Supp. Figure 1

A



Supp. Figure 1

B

