

# De novo transcriptome reveals blood coagulation/antithrombin factors and infection mechanisms in *Angiostrongylus cantonensis* adult worms

## Research Article

\*These authors contributed equally to this study.

**Cite this article:** de Mattos Pereira L, de Jezuz MPG, Rangel AR, Baldasso BD, Zaluski AB, Graeff-Teixeira C, Morassutti AL (2021). *De novo* transcriptome reveals blood coagulation/antithrombin factors and infection mechanisms in *Angiostrongylus cantonensis* adult worms. *Parasitology* **148**, 857–870. <https://doi.org/10.1017/S0031182021000469>

Received: 12 December 2020

Revised: 17 February 2021

Accepted: 5 March 2021




First published online: 17 March 2021

### Key words:

*Angiostrongylus cantonensis*; *A. cantonensis* transcriptome; anticoagulant/antithrombotic pathways; cerebral angiostrongyliasis; eosinophilic meningitis; functional annotation; molecular diagnostics

### Author for correspondence:

Alessandra Loureiro Morassutti,  
E-mail: [almorassutti@gmail.com](mailto:almorassutti@gmail.com)

Leandro de Mattos Pereira<sup>1,2,\*</sup> , Milene Pereira Guimarães de Jezuz<sup>1,\*</sup>,  
Amaranta Ramos Rangel<sup>1</sup>, Bruna Dalcin Baldasso<sup>1</sup>, Amanda Bungi Zaluski<sup>3</sup>,  
Carlos Graeff-Teixeira<sup>1,4</sup>  and Alessandra Loureiro Morassutti<sup>5,6</sup> 

<sup>1</sup>Laboratório de Biologia Parasitária, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Escola de Ciências, Porto Alegre, RS, Brazil; <sup>2</sup>Databiomics, Parque Tecnológico Tecnovates, Lajeado, RS 95914-014, Brazil; <sup>3</sup>Laboratório de Biologia e Desenvolvimento do Sistema Nervoso, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Escola de Ciências, Porto Alegre, RS, Brazil; <sup>4</sup>Núcleo de Doenças Infecciosas, Centro de Ciências da Saúde, Universidade Federal do Espírito Santo, Vitória, ES, Brazil; <sup>5</sup>Escola de Medicina IMED, Passo Fundo, RS 99070-220, Brazil and <sup>6</sup>Instituto de Patologia de Passo Fundo, Passo Fundo, RS 99010-081, Brazil

### Abstract

*Angiostrongylus cantonensis* is the main aetiological agent of eosinophilic meningoencephalitis in humans. Several outbreaks have been documented around the world, cementing its status as an emerging global public health concern. As a result, new strategies for the diagnosis, prophylaxis and treatment of cerebral angiostrongyliasis are urgently needed. In this study, we report on the *de novo* assembly of the *A. cantonensis* transcriptome, its full functional annotation and a reconstruction of complete metabolic pathways. All results are available at AngiostrongylusDB (<http://angiostrongylus.lad.pucrs.br/admin/welcome>). The aim of this study was to identify the active genes and metabolic pathways involved in the mechanisms of infection and survival inside *Rattus norvegicus*. Among 389 metabolic mapped pathways, the blood coagulation/antithrombin pathways of heparan sulphate/heparin are highlighted. Moreover, we identified genes codified to GP63 (leishmanolysin), CALR (calreticulin), ACE (peptidyl-dipeptidase A), myoglobin and vWD (von Willebrand factor type D domain protein) involved in the infection invasion and survival of the parasite. The large dataset of functional annotations provided and the full-length transcripts identified in this research may facilitate future functional genomics studies and provides a basis for the development of new techniques for the diagnosis, prevention and treatment of cerebral angiostrongyliasis.

### Introduction

*Angiostrongylus cantonensis* is the main aetiological agent of eosinophilic meningoencephalitis in humans (Graeff-Teixeira *et al.*, 2009). Its life cycle includes mollusks as intermediate host and rodents as definitive hosts, with the adult worms living inside the pulmonary arteries (Alicata, 1965). Humans can become infected through the ingestion of undercooked mollusks, crustaceans or vegetables contaminated with third-stage larvae (L3). In human hosts, the third-stage larvae (L3) do not complete the cycle and are trapped in the central nervous system.

Cerebral angiostrongyliasis is an emerging global public health concern, with outbreaks documented across the world, including in Southeast Asia, Pacific Islands, parts of South and Central America and the Caribbean (Wang *et al.*, 2008; Martin-Alonso *et al.*, 2011; do Espírito-Santo *et al.*, 2013; Burns *et al.*, 2014). To help elucidate the mechanisms and effects of cerebral angiostrongyliasis in humans, both whole genome sequencing (Morassutti *et al.*, 2013; Yong *et al.*, 2015a) and a complete mitochondrial genome (Yong *et al.*, 2015b; Červená *et al.*, 2019) comparative transcriptomic analysis between the fifth-stage larvae and *A. cantonensis* adult lineages from China have been performed in previous studies (Yong *et al.*, 2015a).

In the current study, we used *de novo* transcriptome assembly, complete functional annotation and metabolic reconstruction from adult worms of *A. cantonensis* isolated from Porto Alegre (30°01'40" south; 51°13'43" west), in the southernmost Brazilian state, Rio Grande do Sul, with the aim of identifying the expressed and active genes in adult worms to improve our understanding of this disease and its main aetiological agent. Our findings represent a valuable source of novel molecular targets for the diagnosis and treatment of cerebral angiostrongyliasis. The annotation results are available in the database AngiostrongylusDB at: <http://angiostrongylus.lad.pucrs.br/admin/welcome>.

## Materials and methods

### Isolation of worms and RNA extraction

*Angiostrongylus cantonensis* female worms were recovered (42 days post-infection) from the pulmonary arteries and right-heart cavities of experimentally infected Wistar *Rattus norvegicus*, washed in saline solution, and stored at  $-80^{\circ}\text{C}$  in RNAlater (Qiagen, Inc., Valencia, CA) until use. Approximately 30 mg of worms were homogenized in  $600\ \mu\text{L}$  of lysis buffer RA1 using a T8 homogenizer (IKA WORKS, Inc., Wilmington, NC, USA). Total RNA was isolated using the NucleoSpin RNA II kit (Machery-Nagel, Inc., Bethlehem, PA, USA), according to the manufacturer's protocol.

### De novo assembly, prediction of coding regions and functional annotation

The *A. cantonensis* adult worm transcriptome was sequenced at Macrogen using Illumina HiSeq2500, obtaining  $2 \times 150$  paired end reads in the fastq format. The Illumina adapter sequences were removed from the reads, and low-quality reads towards the 3' and 5' ends of the reads were trimmed. Then, the reads were scanned with a 4 base-wide sliding window, and leading or trailing bases with an average Phred quality score lower than 20 were removed using Trimmomatic (Bolger *et al.*, 2014). Reads shorter than 50 bp were also discarded. Transcriptome *de novo* assembly was performed using Trinity 2.0.6 (Grabherr *et al.*, 2011) with default *k*-mer. Gene open reading frames (ORFs) or protein coding regions within the transcripts were predicted in TransDecoder (<https://github.com/TransDecoder/TransDecoder/wiki>) using the Ab-initio model, while homology searches were performed using the blastp results against all the proteins of the Swiss-Prot databases, PFAM functional domains (El-Gebali *et al.*, 2019) and THMM 2.0c (Transmembrane helix Markov model) software packages. Only predicted ORFs that were at least 100 amino acids long were retained. Subsequently, non-redundant (NR) sequences were obtained by comparing against all sequences in Cluster Database at High Identity with Tolerance (CD-HIT; <http://weizhongli-lab.org/cd-hit/>) using a cut-off of 100% identity (parameters: -c 1.0, -aL 1.0, -AL 1.0). The genes corresponding to these ORFs were recovered using the Perl script and functionally annotated using BLASTx (BLAST+ v2.2) (Camacho *et al.*, 2009) against NR (NCBI, non-redundant database) with an *E*-value cut-off of  $1.0 \times 10^{-6}$  and with InterProScan5 (Jones *et al.*, 2014). All annotations were performed and recovered using Blast2GO (Conesa *et al.*, 2005). Genes annotated with gene ontology (GO) term Oviposition were recovered using a complete list of GO annotation level terms assigned with Blast2GO with bash programming language scripts. The ORFs were annotated with BLASTp against TrEMBL/UniprotKB with an *E*-value cut-off of  $1.0 \times 10^{-6}$  and annotation recovery with Blast2GO.

### Transcriptome quality evaluation

The Trinity Assembly statistics, such as Nx statistics (e.g. the contig N50 value), total trinity genes, total of trinity transcripts and other transcripts, were obtained with the Perl script 'TrinityStats.pl' of the Trinity toolkit. The completeness and contiguity of the *A. cantonensis* transcriptome was assessed by comparing their assembly transcripts to benchmarking sets of universal single-copy (BUSCO) of the Eukaryota and Nematoda in BUSCO v3 (Simão *et al.*, 2015), based on the content of near-universal single-copy orthologues selected from the OrthoDB v9 database. The full-length transcript or near full-length transcript was identified using BLASTX against the predicted proteome of *Dictyocaulus viviparus* (GCA\_000816705.1),

*Ancylostoma ceylanicum* (GCA\_000688135.1), *Ancylostoma duodenale* (GCA\_000816745.1) and *Caenorhabditis elegans* (GCF\_000002985.6) obtained from WormBase (<https://parasite.wormbase.org/index.html>), with an *E*-value cut-off of  $1 \times 10^{-20}$  and parameters -max\_target\_seqs 1. The resulting BLASTX table (outfmt6) was processed with the Perl script 'analyze\_blastPlus\_topHit\_coverage.pl' from the Trinity toolkit. For each Blast hit in the target protein database, the best matching Trinity transcript was identified and selected. The percentage of Blast hit lengths covered by the Trinity transcript was identified. Global alignment between the encoding proteins was performed using Needle ([https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/)).

### Transcript abundance estimation

To estimate transcript abundance, sequenced pair-end reads were re-aligned to the assembled transcripts for quantification by using RSEM software (RNA-Seq by Expectation Maximization) using the Trinity script (align\_and\_estimate\_abundance.pl) in the Trinity toolkit. TPM values with cut-off  $\geq 100$  were considered highly expressed, while  $3 \leq \text{TPM} < 100$  were considered moderately expressed, according to previous studies (Shadeo *et al.*, 2007; Piras *et al.*, 2018; Duan *et al.*, 2019).

### Annotation and reconstruction of metabolic pathways

The annotation of signal transduction pathways from the transcriptome of *A. cantonensis* was performed through KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007), which provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database (Kanehisa *et al.*, 2016). To accomplish this, the proteomes in the representative set for eukaryotes were used as a reference. The set contained the following organisms: hsa: *Homo sapiens*, mmu: *Mus musculus*, rno: *Rattus norvegicus*, dre: *Danio rerio*, dme: *Drosophila melanogaster*, ath: *Arabidopsis thaliana*, sce: *Saccharomyces cerevisiae*, ago: *Ashbya gossypii*, cal: *Candida albicans*, spo: *Schizosaccharomyces pombe*, ecu: *Encephalitozoon cuniculi*, pfa: *Plasmodium falciparum* 3D7, cho: *Cryptosporidium hominis*, ehi: *Entamoeba histolytica*, eco: *Escherichia coli* K-12 MG1655, nme: *Neisseria meningitidis* MC58, hpy: *Helicobacter pylori* 26695, bsu: *Bacillus subtilis* subsp. subtilis 168, lla: *Lactococcus lactis* subsp. lactis Il1403, mge: *Mycoplasma genitalium* G37, mtu: *Mycobacterium tuberculosis* H37Rv, syn: *Synechocystis* sp. PCC 6803, aae: *Aquifex aeolicus*, mja: *Methanocaldococcus jannaschii*, ape: *Aeropyrum pernix* jointly with available proteomes of nematodes: cel: *Caenorhabditis elegans*, cbr: *Caenorhabditis briggsae*, nai: *Necator americanus*, bmy: *Brugia malayi*, loa: *Loa*, tsp: *Trichinella spiralis* and trematode: smm: *Schistosoma mansoni*. The results contained a list of KO assignments (KEGG Orthology) and functional annotations for the *A. cantonensis* transcripts that were processed with the Perl script. Complete and almost complete modules were assigned with Reconstruct Module of KEGG ([https://www.genome.jp/kegg/tool/map\\_module.html](https://www.genome.jp/kegg/tool/map_module.html)), while metabolic pathway figures were obtained with KEGG Mapper ([https://www.genome.jp/kegg/tool/map\\_pathway2.html](https://www.genome.jp/kegg/tool/map_pathway2.html)). The complete modules were tighter functional units for pathways, and protein complexes often corresponded to sub-pathways in the KEGG pathway maps. Each module was identified by M and was manually defined as a combination of K numbers (representing individual genes in each genome) revealing whether a module was complete, and a gene set was present. The list of *A. cantonensis* genes with KOs assigned (Supplementary Table S6) with KASS was used for module reconstruction.

### Database of *A. cantonensis*: AngiostrongylusDB

The annotation of the *A. cantonensis* transcriptome is available as refined data established in the relational structure of MySQL Database Manager System 14 (<https://dev.mysql.com/>), and can be accessed using the web interface AngiostrongylusDB, <http://angiostrongylus.lad.pucrs.br/admin/welcome>, with access to the control profiles for queries, curation of the data or inclusion of new annotations using the programming language Ruby on Rails 4.2.6. This allows for the easy maintenance and adaptation of research requirements in the database and provides a framework that facilitates navigation between the application's functionalities. The web interface AngiostrongylusDB provides the following annotation datasets: (i) functional annotation retrieved from BLASTx and BLASTp against *non-redundant* protein sequence database (NCBI); (ii) annotation of KASS; (iii) UniProt Knowledge base (*UniProtKB*) and (iv) search facilitator corresponding to filters for these annotation listings. These listings contain links to the refined details of each list item according to the sequence identifier and links for the editing or curation of these data, depending on the user's profile.

## Results

### De novo assembly of transcriptome and completeness assessment

The transcriptome of *A. cantonensis* (Supplementary Fig. 1) generated 58 418 364 paired reads (A2\_1 and A2\_2, respectively) with 93.58% of these reads having a Phred score >Q30 (per base sequence quality). *De novo* assembly generated 82 779 transcripts, with 53 467 genes with an N50 length of 1870, a median contig length of 535, an average contig length of 1034.77 and a total of assembled contigs of 85 657 588 bp (Supplementary Table S1). The completeness of Transcriptome Assessment by the Benchmarking Universal Single-Copy Orthologues (BUSCO v3) indicated that 87.1% ( $N = 852$ ) of the 978 core Metazoan genes were identified as complete, fragmented 1.3% ( $N = 14$ ), while 11.6% ( $N = 113$ ) of the genes were missing from the assembly (Supplementary Table S1). Regarding the eukaryotic core genes, 98.3% ( $N = 298$ ) of the genes were identified as complete, 1.0% ( $N = 3$ ) as fragmented and 0.7% ( $N = 2$ ) as missing. The quantity of the full-length or nearly full-length transcripts, determined by the number of transcripts that were reconstructed to full-length (100% alignment) or near full-length (>70% alignment) by alignment to proteome of nematodes, is shown in Table 1. We identified a total of 14 502 full-length or near full-length (100–70%) using as reference the proteomes of *A. ceylanicum*, *D. viviparus*, *A. duodenale* and *C. elegans* (Table 1 and Supplementary Table S1). Our transcriptome has 22 292 NR proteins codified, more than the reference genome of *A. cantonensis* deposited at the NCBI Genome database (GCA\_009735665.1, Xu *et al.*, 2019) what has 10 314 proteins and a genome size of 293 Mb.

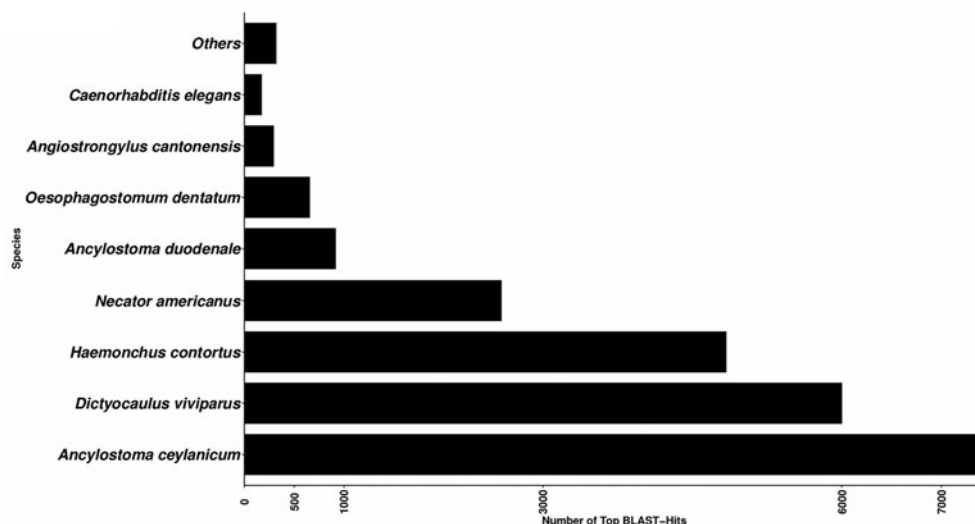
### Functional annotation of the transcriptome

From the 82 779 transcripts, 50 207 coding regions were predicted using TransDecoder with an *ab initio* model and homology for known sequences and functional domains. Subsequently, 22 292 NR-predicted transcripts were obtained with CD-HIT using 100% identity with all other default parameters unchanged. The distribution of the most significant hits (Top-Hits) in the BLASTx results against the NR database (NCBI) is shown in Fig. 1 and Supplementary Table S1. From the 22 292 proteins, nematode parasites accounted for 94.33% of all hits, most of them (7175 proteins, 32.10%) with high similarity to *A. ceylanicum*, 5596 (26.82%) to *D. viviparus*, 4706 (21.05%) to

**Table 1.** Full-length transcript reconstruction comparative analysis of the transcriptome of *Angiostrongylus cantonensis* adult worms

Pct_cov_Hit (%)	<i>Ancylostoma ceylanicum</i>			<i>Dictyocaulus viviparus</i>			<i>Ancylostoma duodenale</i>			<i>Caenorhabditis elegans</i>		
	PC	PCS	Pct_cov_Hit (%)	PC	PCS	Pct_cov_Hit (%)	PC	PCS	Pct_cov_Hit (%)	PC	PCS	Pct_cov_Hit (%)
100	1774	1774	100	2091	2091	100	1917	1917	100	787	787	100
90	5946	7720	90	5966	8057	90	6290	8207	90	7086	7873	90
80	2246	9966	80	2101	10 158	80	2515	10 722	80	1897	9770	80
70	1980	11 946	70	1599	11 757	70	1677	12 399	70	1463	11 233	70

The cumulative number of proteins of *A. ceylanicum*, *D. viviparus*, *A. duodenale* and *C. elegans* recovered in the assembly of the *A. cantonensis* transcriptome at 80–100% of coverage. Transcripts identified in *A. cantonensis* were annotated as full-length transcripts if they matched a protein in the reference proteome database with an *E*-value threshold of  $1 \times 10^{-20}$ . Pct. cov. Hit, percentage of coverage of top matching hits of reference proteome aligned across more than X% (80–100) with the *A. cantonensis* transcript. PC, protein counts of target reference proteome aligned by at least one transcript of *A. cantonensis*; PCS, protein counts sum of reference proteins from reference proteome aligned at X% (80–100) coverage by at least one *A. cantonensis* transcript. Assembly ID of *A. ceylanicum*: GCA\_000688135.1, *D. viviparus*: GCA\_000816705.1, *A. duodenale*: GCA\_000816745.1, *C. elegans*: GCF\_000002985.6.



**Fig. 1.** Species distribution of predicted homologues in the transcriptome of *Angiostrongylus cantonensis* adult worms. Homologues were predicted using a BLASTX search against the Nr NCBI database at an  $E$ -value cut-off of  $1.0 \times 10^{-6}$ . The top eight species with the most homologues are shown.

*Haemonchus contortus*, 1292 (5.78%) to *N. americanus*, 891 (3.98%) with *A. duodenale*, 652 (2.91%) to *Oesophagostomum dentatum* and 288 (1.28%) to *A. cantonensis*, 84 (0.37%) to *C. elegans*, 397 (1.77%) to other organisms, and the remaining 870 (3.89%) did not show similarity to any other organisms in the NR database. Taken together, all 22 292 sequences were annotated using BLASTx, InterProscan and GO (Supplementary Table S2). The datasets of the annotation of *A. cantonensis* and the fasta files were stored in a database created in MySQL and can be accessed at <https://angiostrongylus.lad.pucrs.br/admin>.

GO annotation [biological processes, molecular functions (MF) and cellular components] at level 2 were predicted for the 22 292 total mRNA sequences of adults of *A. cantonensis* (Fig. 2 and Supplementary Table S2). The top GO term assignments for all mapped genes in the biological processes (BP) category included: cellular process ( $N = 10\,281$ ), metabolic process ( $N = 9545$  proteins), developmental process ( $N = 6824$ ), multicellular organism process ( $N = 6799$ ), biological regulation ( $N = 6106$ ), regulation of biological process ( $N = 5324$ ), localization ( $N = 4933$ ), reproduction ( $N = 4303$ ), cellular component organization or biogenesis ( $N = 3422$ ), response to stimulus ( $N = 3278$ ), locomotion ( $N = 2831$ ), reproductive process ( $N = 2639$ ), signalling ( $N = 2234$ ), positive regulation of biological process ( $N = 1617$ ), negative regulation of biological process ( $N = 1405$ ), multi-organism process (1190), growth ( $N = 948$ ), behaviour ( $N = 829$ ), cell proliferation ( $N = 292$ ) and biological adhesion ( $N = 253$ ) (Fig. 2A). In the MF category, the GO terms assignment include: binding ( $N = 9064$ ), catalytic activity ( $N = 7567$ ), transport activity ( $N = 1260$ ), regulator ( $N = 544$ ), molecular transducer activity ( $N = 499$ ), transcription regulator activity ( $N = 440$ ), structural molecule activity ( $N = 421$ ), translation regulator activity ( $N = 230$ ), antioxidant activity ( $N = 82$ ), molecular carrier activity ( $N = 30$ ), cargo receptor activity ( $N = 18$ ), small molecule sensor activity ( $N = 4$ ) and protein tag ( $N = 2$ ) (Fig. 2B). For the cellular component category, the GO terms assigned include: cell ( $N = 8613$ ), cell part ( $N = 8542$ ), organelle ( $N = 6009$ ), membrane ( $N = 4394$ ), protein containing complex ( $N = 3429$ ), membrane part ( $N = 3313$ ), organelle part ( $N = 3184$ ), membrane enclosed lumen ( $N = 922$ ), supramolecular complex ( $N = 651$ ), extracellular region ( $N = 466$ ), extracellular region part (372), cell junction ( $N = 347$ ), synapse ( $N = 283$ ), synapse part ( $N = 173$ ), virion part ( $N = 53$ ), virion ( $N = 53$ ), nucleoid ( $N = 13$ ), other organism ( $N = 4$ ) and other organism part ( $N = 4$ ) (Fig. 2C). The GO annotation (biological processes, MF and

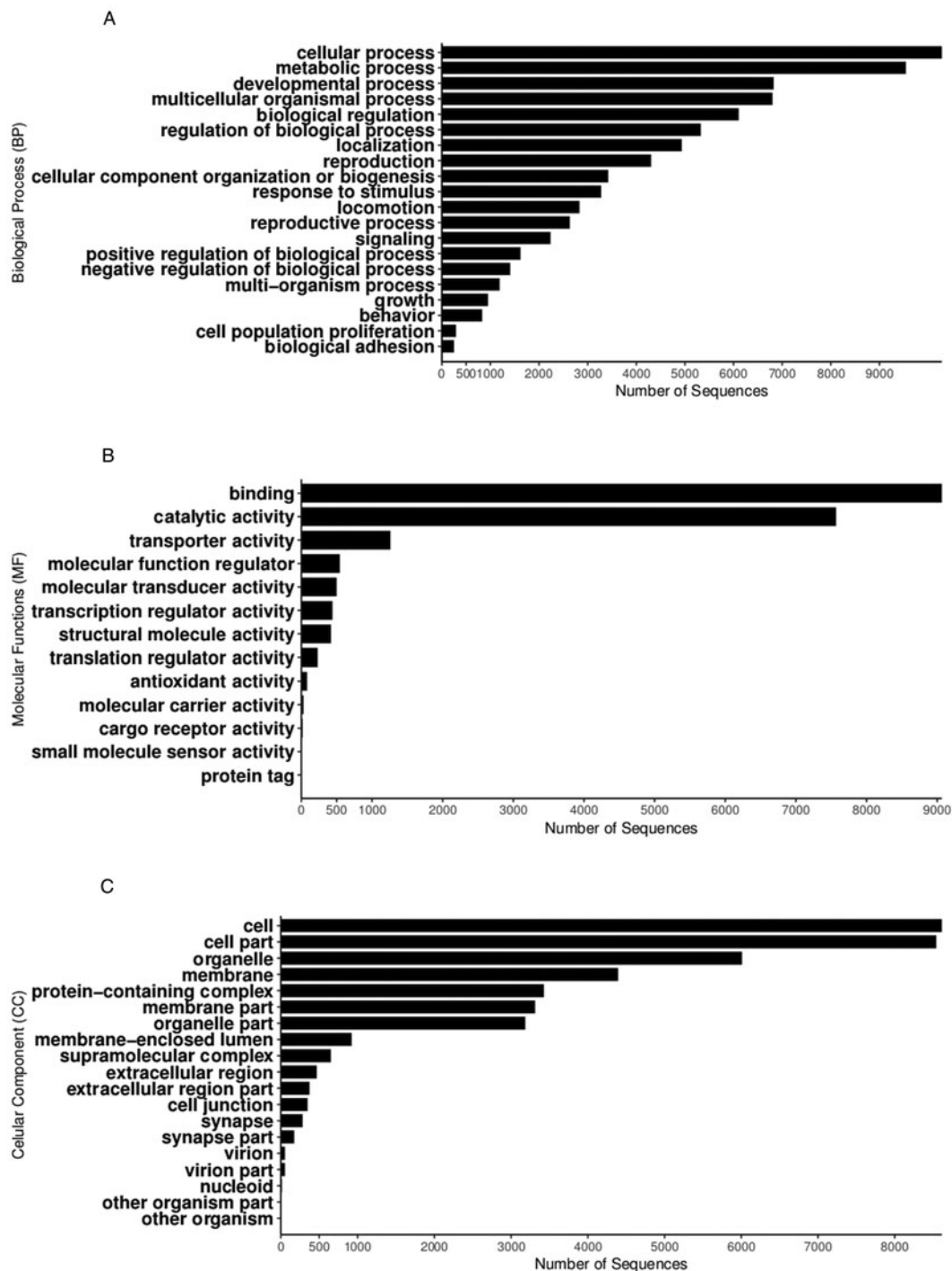
cellular components) at level 9 was also predicted for all sequences (Supplementary Table S2). We found 1053 GO terms, where the top five most abundant GO terms were: alpha-amino acid metabolic process (GO:1901605) with 1218 transcript sequences; nucleic acid-templated transcription (GO:0097659) with 1173 transcript sequences; hermaphrodite genitalia development (GO:0040035) with 1173 transcript sequences; regulation of RNA biosynthetic process (GO:2001141) with 984 transcript sequences and regulation of vulval development (GO:0040028) with 447 transcript sequences. Within the predicted proteins from ORFs, we identified 3012 signatures present in the PFAM, 578 in SMART and 618 in the SUPERFAMILY database. The top 20 most abundant signatures identified by these software are shown in Fig. 3A–3C, respectively.

#### Estimation of transcript abundance in adult *A. cantonensis* worms

The FPKM and TPM (>400) values for the 20 top highly expressed genes are given in Table 2 and Supplementary Table S3. Among the highly expressed genes were those coding for: hypothetical protein DICVIV\_07050, an unnamed protein product, rRNA promoter binding protein, von Willebrand factor type D domain protein, nematode fatty acid retinoid binding protein, ribosomal protein L40, collagen alpha-1(I) chain-like, core histone H2A/H2B/H3/H4, major sperm protein (MSP), Hsp20/alpha crystallin family, putative cuticle protein, intracellular globin, 14-3-3 protein, aspartyl protease inhibitor, Ubiquitin, glyceraldehyde-3-phosphate dehydrogenase, type I, parasitic stage specific protein 1, chondroitin proteoglycan 3, putative DNA-binding response regulator CreB, putative DNA-binding response regulator CreB and Transthyretin-like family protein (Table 2). Among the top 20 transcripts we found signal peptides for parasitic stage specific protein 1, von Willebrand factor type D domain protein and nematode fatty acid retinoid binding protein (Supplementary Table S4). Altogether 1694 predicted proteins were identified with signal peptide (Supplementary Table S4).

#### Annotation, reconstruction of metabolic pathways and functional modules

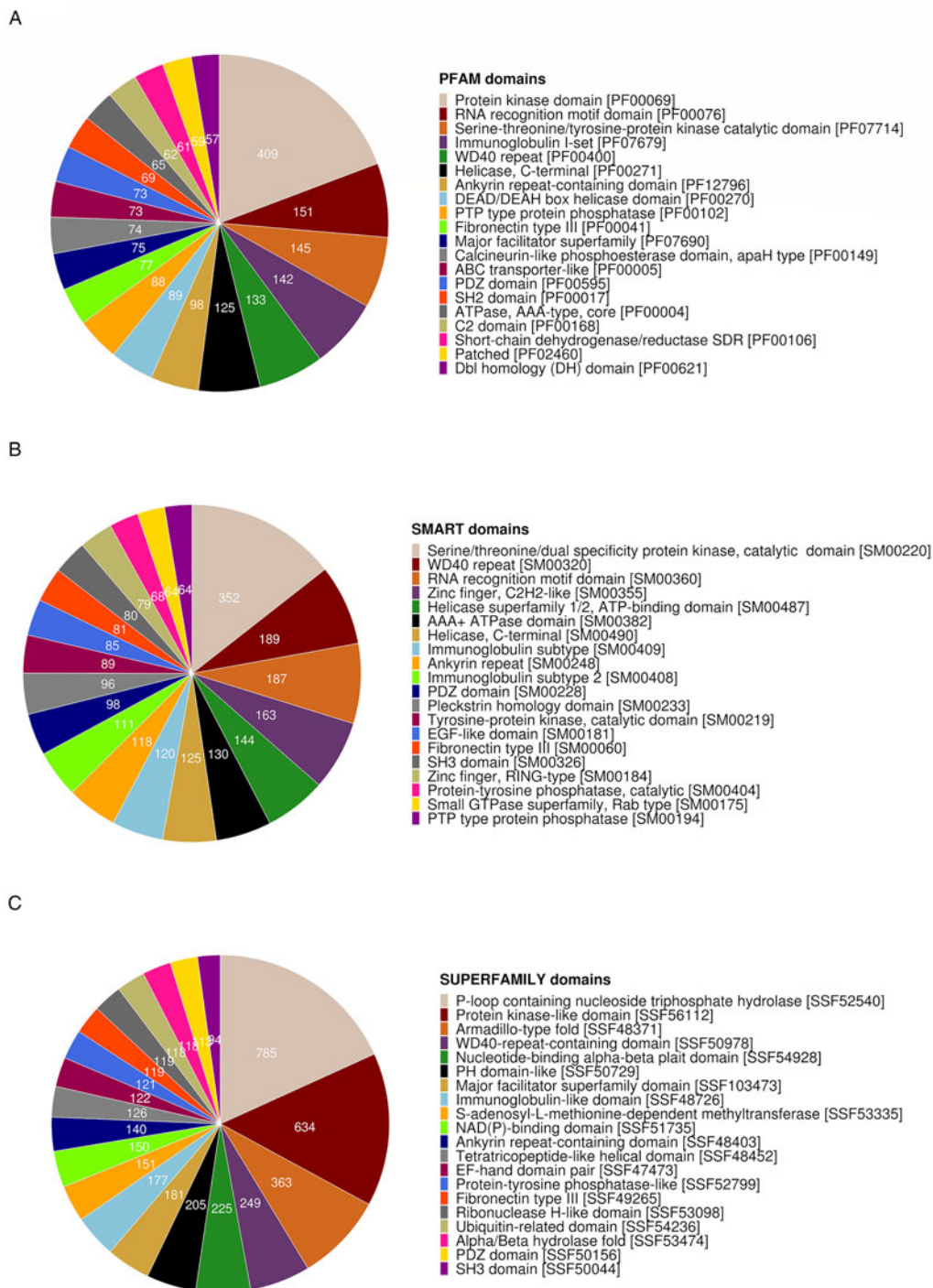
In total, 4203 coding proteins were assigned a KO (orthologous group) with KASS, annotating 389 metabolic pathways containing at least one gene, and 92 complete modules, such as glycolysis module



**Fig. 2.** Functional categorization of transcripts of *A. cantonensis* based on GO annotations terms at level 2. Functional annotation was performed using Blast2GO. The transcriptome was functionally mapped to GO terms and annotated using an *E*-value-hit-filter of  $1.0 \times 10^6$  with all other parameters as default.

(M00001), citrate cycle, first carbon oxidation (M00010), pyruvate oxidation (M00307), pentose phosphate pathway (M00006), galactose degradation (M00632), glycogen biosynthesis (M00854) and degradation (M00855). The genes necessary for the module of complete fatty acid biosynthesis initiation (M00082), fatty acid biosynthesis, elongation (M00083), fatty acid biosynthesis, elongation, mitochondria (M00085), elongation (M00415), sphingosine degradation (M00100), fatty acid biosynthesis, elongation, endoplasmic reticulum (M00415) and triacylglycerol biosynthesis (M00089). *De novo* lipid synthesis was found, as well as beta-oxidation for synthesis of palmitoyl-coA (M00086 and M00087). We identified the pathways of phosphatidylcholine (M00090) and phosphatidylethanolamine (M00092), both the

largest components of phospholipid of biological membranes. We also identified the module for *N*-glycan precursor biosynthesis (M00055), *N*-glycosylation by oligosaccharyltransferase (M00072), *N*-glycan precursor trimming (M00073), GPI-anchor biosynthesis, core oligosaccharide (M00065), *N*-glycan metabolism such as glycosaminoglycan biosynthesis and linkage tetra saccharide (M00057). The module M00065 (GPI-anchor biosynthesis, core oligosaccharide), which includes the enzyme PIGA (phosphatidylinositol glycan, class A), catalyses the first step of GPI-anchor synthesis. The transcriptome annotation analysis indicates that *A. cantonensis* can synthesize *de novo* purine adenine (M00049) and guanine (M00050), pyrimidines such uridine (M00051) and other ribonucleotides (for more details see module



**Fig. 3.** Top 20 most abundant signatures identified in PFAM (A), SMART (B) and SUPERFAMILY databases (C). The most abundant functional domains identified in the coding predicted proteins of *A. cantonensis* are shown.

M00052). The modules of metabolic pathways of serine biosynthesis from glycerate-3P (M00020), methionine salvage pathway (M00034) and proline biosynthesis from glutamate are present (M00015). For example, the complete module of glycosaminoglycan biosynthesis was comprised of a heparan sulphate (HS) backbone (M00059) containing all enzymes involved in the synthesis of heparan sulphate or heparin (HP) (Fig. 4). The final product of HS biosynthesis was the form of HS proteoglycan.

Among the incomplete metabolic pathways, genes homologous to *gp63* (EC:3.4.24.36, KO: K01404, comp16391\_c0, TPM: 4.59), which encodes leishmanolysin of *Leishmania* sp. promastigotes (Fig. 5) and TcCRT (calreticulin, KO: K08057, comp6270\_c0, TPM: 182.93) of *Trypanosoma cruzi*, were annotated in

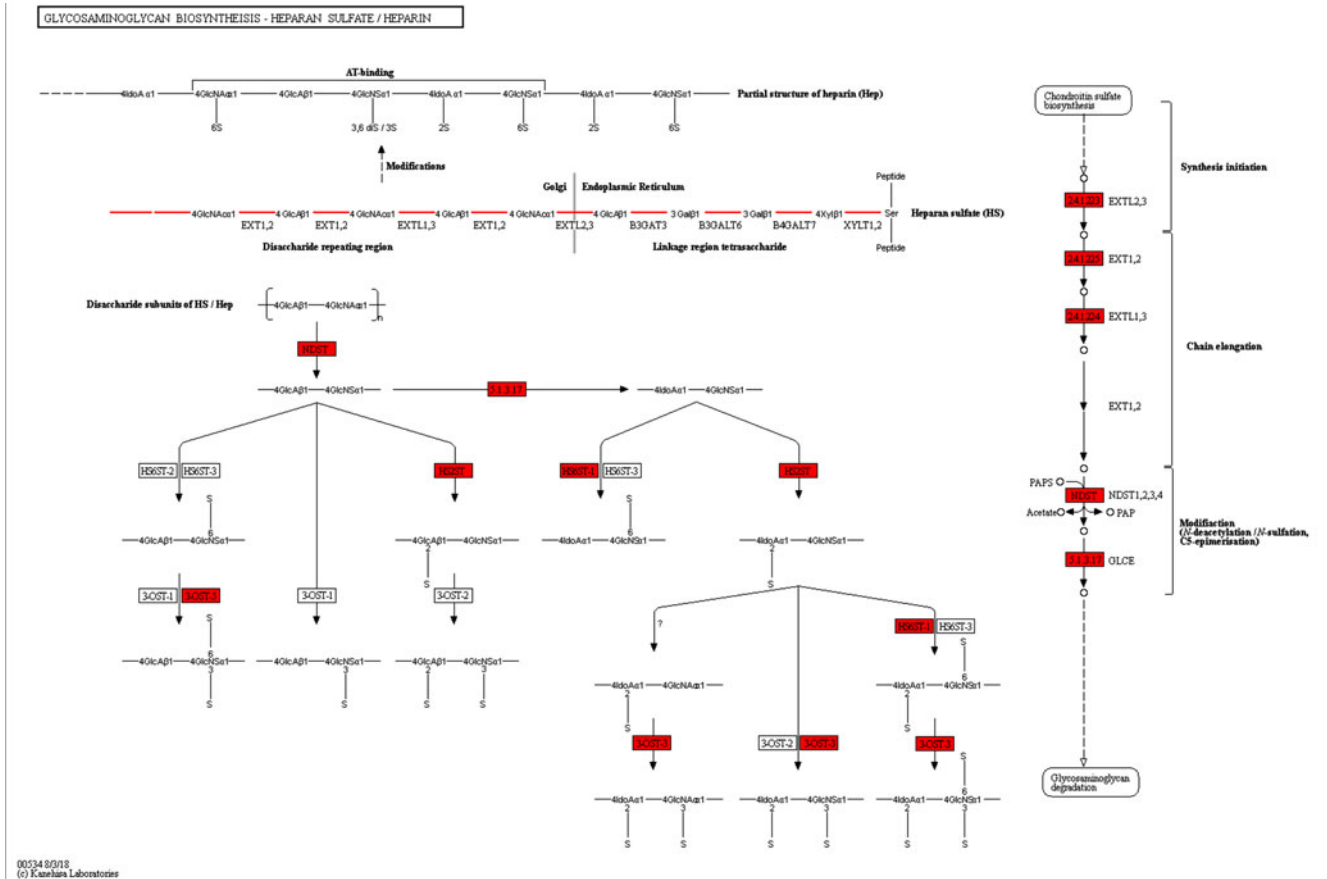
*A. cantonensis* (Fig. 6). In this last pathway, ACE (peptidyl-dipeptidase A, EC: 3.4.15.1, KO: K01283, comp30541\_c0, TPM: 1.39) was also identified. In addition, the coding sequences for *laminin gamma-1* (KO: K05635, LAMC1, ID: comp27259\_c0\_seq1, TPM: 1.94) and heat shock 70 kDa protein (HSP70) 1/2/6/8 (HSPA1s, KO: K03283, comp19023\_c0\_seq1, TPM: 362.36), homologous proteins components of the mechanism of infection of *Toxoplasma*, were also annotated.

The hemoglobin Hb [hemoglobin subunit alpha (K13822, comp7751\_c0\_seq1, TPM: 3.70), and its beta subunit (K13823, comp7748\_c0\_seq1\_ORF, TPM: 2.19)] was also annotated (Supplementary Table S6). These sequences were found to have the best hit values (BLASTx, 100% of identity) against the sequences

**Table 2.** Highly expressed genes in female adult worms from *A. cantonensis*, southern Brazil isolate from Porto Alegre, after transcriptome analysis

Gene ID	Transcript ID(s)	Length	TPM	FPKM	Hit description NR	Hit ACC	E-value	Similarity (%)
comp17882_c1	comp17882_c1_seq1	289	184 071.04	243 954.67	Hypothetical protein DICVIV_07050	KJH46859.1	$3.00 \times 10^{-38}$	78
comp18968_c0	comp18968_c0_seq1	206	32 726.77	43 373.73	Unnamed protein product, partial	VDL69858	$3.00 \times 10^{-18}$	82
comp6464_c0	comp6464_c0_seq1	1005	16 185.45	21 451.05	rRNA promoter binding protein	XP_001891902	$4.00 \times 10^{-37}$	71
comp6474_c0	comp6474_c0_seq1	4851	6860.29	9092.14	von Willebrand factor type D domain protein	RCN42163	0	80
comp10575_c1	comp10575_c1_seq1	599	2721.98	3607.51	Nematode fatty acid retinoid binding protein	KJH44317.1	$9.00 \times 10^{-84}$	91
comp9400_c0	comp9400_c0_seq1	228	2297.31	3044.69	Ribosomal protein L40	AEL28882.1	$2.00 \times 10^{-47}$	100
comp18989_c0	comp18989_c0_seq1	391	2050.42	2717.48	Collagen alpha-1(I) chain-like	XP_027819650.1	$1.00 \times 10^{-41}$	99
comp17947_c0	comp17947_c0_seq1, comp17947_c0_seq2	429.75	1470.13	1948.4	Core histone H2A/H2B/H3/H4	EPB74136.1	$4.00 \times 10^{-51}$	97
comp15792_c0	comp15792_c0_seq1, comp15792_c0_seq2	780.57	1470.1	1948.37	Major sperm protein	AAB27962.2	$2.00 \times 10^{-88}$	100
comp18995_c0	comp18995_c0_seq1	620	1203.55	1595.1	Hsp20/alpha crystallin family protein	KJH45534.1	$2.00 \times 10^{-81}$	78
comp18254_c0	comp18254_c0_seq1, comp18254_c0_seq2, comp18254_c0_seq3, comp18254_c0_seq4, comp18254_c0_seq5, comp18254_c0_seq6	509.75	1007.47	1335.23	Putative cuticle protein	CAR63653.1	$6.00 \times 10^{-21}$	87
comp18985_c0	comp18985_c0_seq1	1238	963.23	1276.59	Intracellular globin	AAL56427.1	$7.00 \times 10^{-67}$	64
comp18994_c0	comp18994_c0_seq1	1125	871.26	1154.7	14-3-3 protein	AEK98129.1	$1.00 \times 10^{-179}$	99
comp19002_c0	comp19002_c0_seq1	858	797.43	1056.85	Aspartyl protease inhibitor	APS24030.1	$2.00 \times 10^{-91}$	100
comp9142_c0	comp9142_c0_seq1	497	730.41	968.03	Ubiquitin	EFN65063.1	$2.00 \times 10^{-68}$	100
comp18240_c0	comp18240_c0_seq1	1223	707.97	938.3	Unnamed protein product	VDM54411.1	$1.00 \times 10^{-49}$	53
comp19004_c0	comp19004_c0_seq1	1822	649.86	861.28	Glyceraldehyde-3-phosphate dehydrogenase, type I	XP_013303290.1	0	92
comp19005_c0	comp19005_c0_seq1	1625	608.46	806.4	Parasitic stage specific protein 1	ADN00782.1	$3.00 \times 10^{-65}$	71
comp19010_c0	comp19010_c0_seq1	736	570.69	756.36	Chondroitin proteoglycan 3	CDJ80611.1	$2.00 \times 10^{-36}$	58
comp7732_c0	comp7732_c0_seq1	711	477.06	632.26	Putative DNA-binding response regulator CreB	CAR63606.1	$4.00 \times 10^{-21}$	100
comp19044_c0	comp19044_c0_seq1	633	404.55	536.16	Transthyretin-like family protein	PIO69026.1	$2.00 \times 10^{-67}$	92

TPM, transcripts per kilobase million; FPKM, fragments per kilobase million.



**Fig. 4.** Pathways of glycosaminoglycan biosynthesis. Biosynthesis of heparan sulphate and heparin backbone. Red rectangles: enzymes annotated in *A. cantonensis* with KEGG Automatic Annotation Server (KAAS) composing the complete functional module of heparan sulphate and heparin backbone. Other rectangles (white): unidentified enzymes.

of *R. norvegicus* (Supplementary Table S6). However, additional sequences of intracellular myoglobin (comp18983\_c0\_seq1, TPM: 1658.86), homologous for nematode sequences of *Syngamus trachea*, were also annotated (Supplementary Table S6).

**Discussion**

In the current study, all NR ORFs in adult *A. cantonensis* worms from a Brazilian isolate were identified (available at <http://angiostrongylus.lad.pucrs.br/admin/welcome>). Our assembly had an N50 of 1870, totalling 85 657 Mb and 22 294 NR ORF sequences. We identified and annotated a higher number of predicted proteins than the previously deposited reference genomes of *A. cantonensis* (Xu *et al.*, 2019), which could make our current transcriptome analysis a complement for the functional knowledge of the coding proteins present in *A. cantonensis*. In addition, our reference database makes all ORF sequences annotated available for functional genomics studies. In the reference genome of *A. cantonensis* (Xu *et al.*, 2019), the fasta files with annotation are not available ([https://parasite.wormbase.org/Angiostrongylus\\_cantonensis\\_prjna350391/](https://parasite.wormbase.org/Angiostrongylus_cantonensis_prjna350391/)). Among the sequences with the best BLAST hits within nematodes, only a small number were from *A. cantonensis*. This is because the *A. cantonensis* proteome is absent from the NR NCBI database. The species list with the best BLAST hit matched the phylogenetic clade V, in which *A. cantonensis* is grouped (Kiontke *et al.*, 2007; van den Elsen *et al.*, 2009).

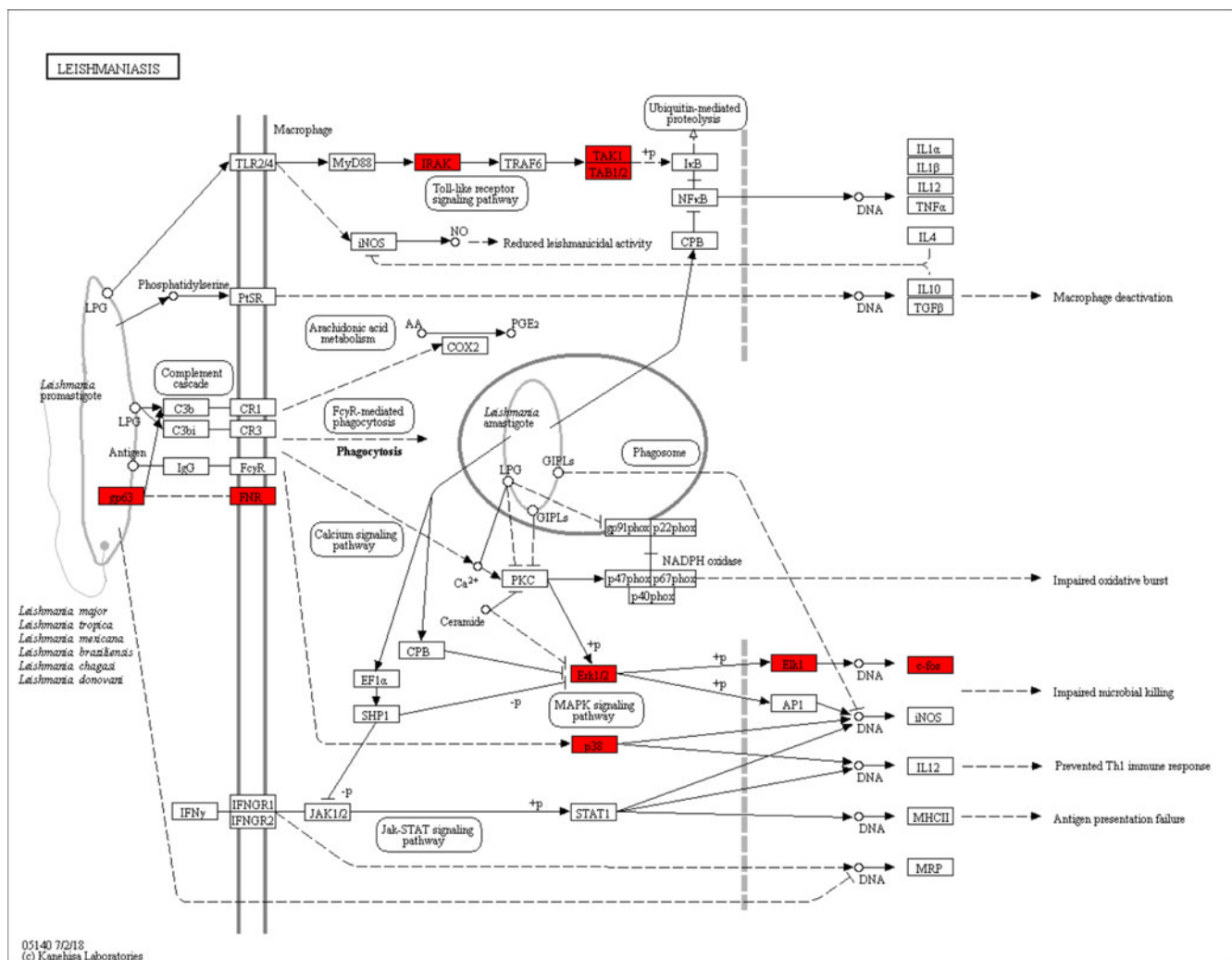
The GO terms are organized hierarchically from the more general level terms to the more specific parent terms (Alexa *et al.*, 2006). In our analysis, we found a similar pattern of GO terms

as in a previous transcriptome study (Yu *et al.*, 2017), but we have additionally identified the GO terms behaviour, cell population proliferation and biological adherence (Fig. 2). Although the same study by Yu and colleagues aimed to compare L5 and adult worms' transcriptome in *A. cantonensis*, the current study focused on the complete functional annotation, ORFs' identification and reconstruction of metabolic pathways important for the survival of the adult worm within the host.

The homologue for glyceraldehyde-3-phosphate dehydrogenase (Table 2) was annotated as one of the most expressed mRNAs in *A. cantonensis*. This enzyme was shown to act as a complement-binding protein, inhibiting the complement-mediated lysis of sensitized erythrocytes (Sahoo *et al.*, 2013), and was present in the parasite secretion product of *H. contortus* (Vedamurthy *et al.*, 2015), suggesting that it could be a common strategy used by parasites to suppress the host's immune response.

The MSP genes (comp19031\_c0 and comp19047\_c0; Supplementary Table S3) were found to be highly expressed in adult females of *A. cantonensis*. Since these genes have not been previously reported in *Angiostrongylus* species, we investigated the MSP transcripts further and assembled the raw reads of *A. cantonensis* adult females from another transcriptome study, deposited in the SRA database (SRR3199277) (Yu *et al.*, 2017). As a result, the same ORFs encoding the MSP proteins were found (Supplementary Table S5). MSP genes have been described to encode sperm-specific cytoskeletal proteins, where their expression in nematodes is restricted to sperm (Cottee *et al.*, 2004; Strube *et al.*, 2009). The MSP protein of *Ascaris* males is an essential protein for the development and locomotion of amoeboid sperm (King *et al.*, 1994; Kuwabara, 2003). In *C. elegans*,





**Fig. 5.** Metabolic pathways of *Leishmania* spp. Red rectangles: enzymes annotated in *A. cantonensis* with KEGG Automatic Annotation Server (KAAS) with reciprocal best-hit homologous in *Leishmania* spp.

hermaphrodite MSP is secreted to stimulate oocyte maturation and ovulation (Miller, *et al.*, 2001; Han *et al.*, 2010). This study is the first to report on MSP transcripts present in female nematodes. This may be due to MSP mRNA molecules present in recent sperm fertilized females just before collection, since in *C. elegans* 10% of oocyte mRNAs are of paternal origin from mRNA transferred from sperm to oocytes (Stoeckius *et al.*, 2014).

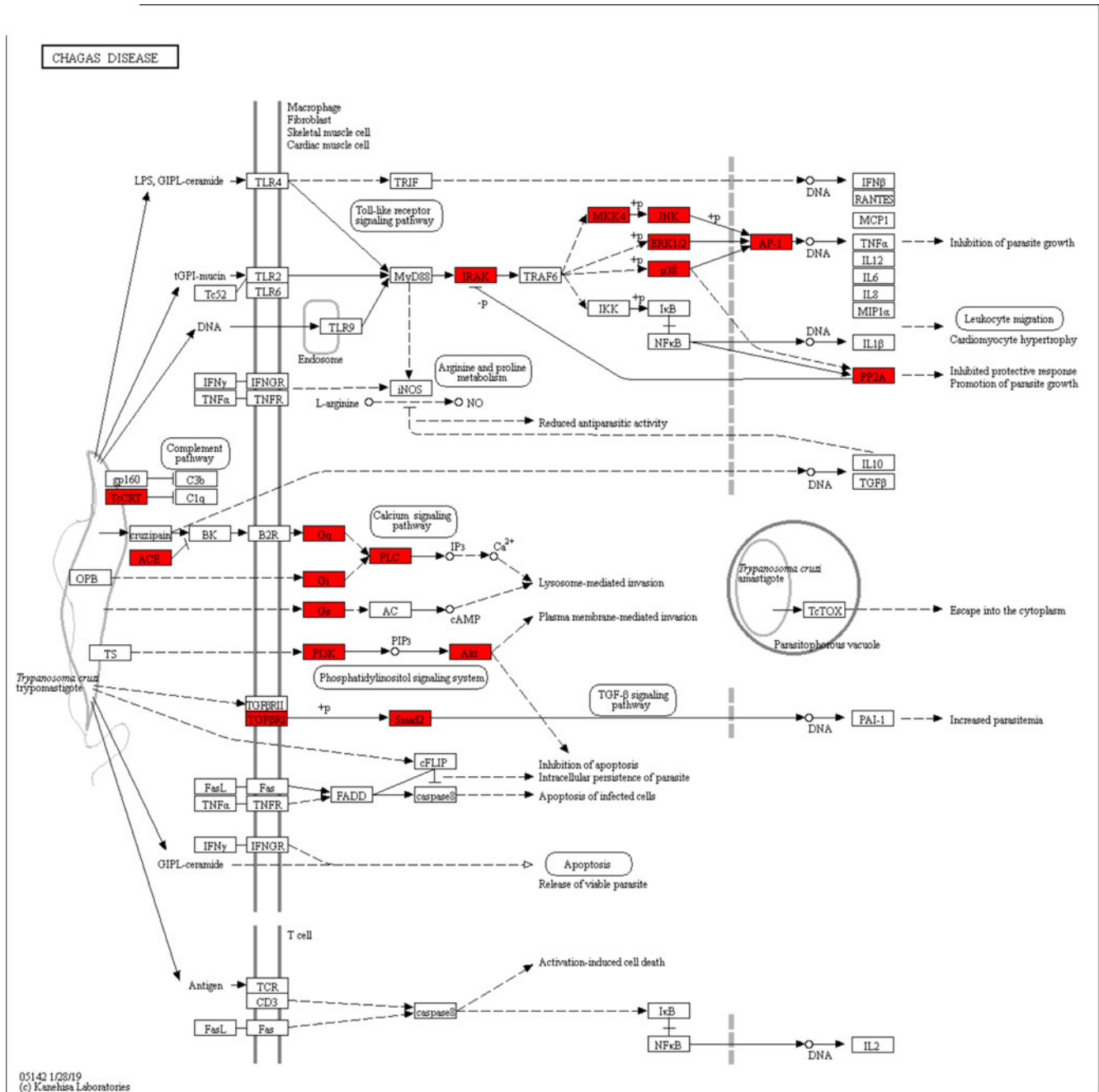
An intracellular globin (TPM: 1276.59) highly expressed in *A. cantonensis* was also annotated. As adult worms inhabit arteries, we hypothesized that the biological role of this protein is associated with the acquisition of blood oxygen and haem group oxygen from the host. The myoglobin-predicted protein identified here presents with 60.1% of identity with the myoglobin sequence (GLB2) of *Nippostrongylus brasiliensis*. This GLB2 was found to be expressed when the nematode enters the host and on the L4 lung stage as an anaerobic pre-adaptation to the intestine (Bouchery *et al.*, 2018). This suggests the metabolic dependence of *A. cantonensis* on its host, since parasite nematodes lack a complete and functional haem biosynthetic pathway (Rao *et al.*, 2005), and must acquire haem from exogenous sources (Luck *et al.*, 2016).

Another transcript that was identified was FAR-1, which encodes an unusual secreted  $\alpha$ -helix rich lipid-binding protein found exclusively in nematodes (Rey-Burusco *et al.*, 2015). In agreement with our findings, the transcripts encoding for FARs are expressed in adult females of *Aphelenchoides besseyi*

(Cheng *et al.*, 2013), *A. ceylanicum* (Fairfax *et al.*, 2009), *Heterodera avenae* and *Heterodera filipjevi* (Qiao *et al.*, 2016). FAR-1 protein antibodies have been used for the immunization of animals infected with *A. ceylanicum* (Fairfax *et al.*, 2009) and *B. malayi* (Zhan *et al.*, 2018).

A putative DNA-binding response regulator CreB (TPM: 632.26, 100% of identity with CAR63606.1) was among the top 20 genes that were highly expressed (Table 2) and is specific to *A. cantonensis*. There were no BLASTx/BLASTp hits found against any other organisms when these sequences were compared against the NR NCBI database and Worm Parasite database. Considering its uniqueness, it has potential for use as a marker in the diagnosis of angiostrongyliasis. Currently, the function of the putative DNA-binding response regulator CreB in *A. cantonensis* remains unknown. However, corroborated with our findings, the cDNA of this transcript was previously identified in the fourth-stage larvae of *A. cantonensis* in a previous study (He *et al.*, 2009).

The annotation results of the metabolic pathways indicated that adult *A. cantonensis* can synthesize and metabolize major organic macromolecules (Supplementary Table S6). Unlike cestodes and schistosomes, which have lost their capacity for the *de novo* synthesis of lipids and have become entirely dependent on hosts (Barrett, 1983; Tyagi *et al.*, 2015b), *A. cantonensis* has complete modules for the biosynthesis, initiation, elongation of fatty acids, as well as the beta-oxidation of lipids. As a result,



**Fig. 6.** Metabolic pathways of *Trypanosoma cruzi*. Red rectangles: enzymes annotated in *A. cantonensis* with KEGG Automatic Annotation Server (KAAS) with reciprocal best-hit homologous in *T. cruzi*.

*A. cantonensis* is also able to synthesize several amino acids, namely serine, cysteine, methionine and proline, given that the complete functional modules for these pathways are present (Supplementary Table S6). However, the pathways for the biosynthesis and degradation of histidine, lysine, phenylalanine, threonine, leucine, isoleucine and valine were found to be incomplete, suggesting its metabolic dependence on the host for these amino acids. The parasites *Leishmania*, *Plasmodium* and *Cryptosporidium* lost their ability to synthesize the nine amino acids (Phe, Trp, Ile, Leu, Val, Lys, His, Thr and Met) in nine independent pathways as soon as they evolved the ability to feed on other organisms (Payne and Loomis, 2006). In a future study, we will use a comparative genomic approach to compare the metabolic dependence of *A. cantonensis* and *Angiostrongylus costaricensis* on their hosts.

The complete modules for the biosynthesis and trimming of the *N*-glycan precursor (M00055) were found to be present in

*A. cantonensis*. The *N*- and *O*-glycans are major constituents of glycoproteins and outer layers covered by a carbohydrate-rich glycocalyx or cuticle surface coat. They have attracted the attention of researchers in parasitic nematodes due to their immunogenic and immunomodulatory characteristics (Verissimo et al., 2019). Other worms, such as *Schistosoma* spp., *T. spiralis*, *B. malayi*, *Meloidogyne incognita*, *Bursaphelenchus xylophilus*, *Pristionchus pacificus* and *Ascaris sum*, lack the ability to synthesize the *N*-glycan precursor (M00055) (Tyagi et al., 2015a), and may acquire macromolecules from the hosts. For the biosynthesis of GPI-anchor, core oligosaccharide (M00065), which is indispensable for the germline development of the nematode *C. elegans* (Murata et al., 2012), was also found in *A. cantonensis*. In addition, two PIGA genes involved in the first step of GPI-anchor synthesis (K03857, phosphatidylinositol *N*-acetylglucosaminyltransferase subunit A, EC 2.4.1.198, ID: comp18316\_c0\_seq13 and comp18514\_c0\_seq25, Supplementary Table S2) were also annotated. These

PIGA genes are indispensable for the maintenance of the mitotic germline cell number, germline formation and the normal development of oocytes and eggs, with knockout worms displaying 100% lethality (Murata *et al.*, 2012). Therefore, this opens a new line of investigation for the treatment of angiostrongyliasis.

Several other candidate targets for the treatment of angiostrongyliasis were found. Chondroitin proteoglycan 3 (CS3, comp19010\_c0\_seq1: 756.36) was found among the top 20 most highly expressed genes with the highest expression value. Chondroitin proteoglycan 3 is a heterogeneous group of heavily glycosylated proteins with covalently attached glycosaminoglycan (GAG) chains of chondroitin sulphate (CS) (Noborn *et al.*, 2018). It was found to have greater expression in the parasitic female stage of *Strongyloides ratti* compared with other lifecycle stages (Spinner *et al.*, 2012). Furthermore, it was also reported as essential in the oogenesis or early embryogenesis of *C. elegans*, given that treated worms were not viable, showed poor gonad formation, and laid fewer eggs (Galvin *et al.*, 2003).

In addition to chondroitin proteoglycan 3, other proteoglycans (HSPGs) may be useful as treatment targets. The HSPGs that can attach to GAGs are chondroitin, dermatan sulphate and heparan sulphate (HS) (Fig. 4A and Supplementary Table S6). At the first time, the complete module for the HS biosynthetic pathways was identified in the transcriptome of *A. cantonensis*, with the genes showing moderate levels of expression ( $3 \leq \text{TPM} < 100$ ) (Supplementary Table S6). The proteoglycans of heparan sulphate exhibit anticoagulant effects (Teien, *et al.*, 1976) and play important biological roles in development, cell adhesion, cell migrations, cytokine liberation of secretion, growth factors and morphogens (Esko and Selleck, 2002; Kreuger and Kjellén, 2012). Among the HS pathway genes, hst-2 (heparan sulphate 2-O-sulphotransferase) an essential enzyme for the HS sulphation of proteoglycans and tyrosine-sulphated protein. The RNAi-mediated inhibition of hst-2 in *C. elegans* was found to result in abnormal oocyte formation and inhibited egg production (Akiyoshi *et al.*, 2015).

In humans, the von Willebrand factor (VWF) has two essential functions in haemostasis: (i) mediating the adhesion of platelets to subendothelial connective tissue and (ii) binding to blood clotting factor VIII. Patients who lack VWF have shown defects both in blood clotting and in the formation of platelet plugs at vascular injury sites (Sadler, 2008). The protein domains of von Willebrand/integrin A have been proposed to play a biological role in nematodes as a cell adhesion factor (Whittaker and Hynes, 2002). The functional Von Willebrand factor type D domain (vWF-D) is present in different forms of vitellogenin (Vg), the precursor of egg-yolk proteins in both vertebrates and invertebrates (Akasaka *et al.*, 2013; Sun *et al.*, 2013). The fact that vWF-D transcript isoforms (comp6474\_c0\_seq1, comp6476\_c0\_seq1 and comp6473\_c1\_seq1) were found to be highly expressed in the adult worms of *Angiostrongylus* gave rise the hypothesis of a possible survival strategy, since worms live inside arteries and are in close contact with the vascular endothelium. The predicted proteins of these transcript isoforms contain signal peptides, lipoprotein N-terminal, and vWF-D domains, as predicted in SMART (Supplementary Table S4). We hypothesized that HS and vWF-D modulate coagulation and prevent thrombosis in the intravascular habitat of *A. cantonensis*.

Furthermore, a leishmanolysin homologue (TPM: 4.59) with moderate expression was also annotated in *A. cantonensis*. Leishmanolysin (gp63) was first identified in *Leishmania promastigotes*, where it facilitates migration through the host extracellular layer, as well as it affects the AK, MAP and IRAK-1 kinase signalling pathways (McGwire *et al.*, 2002). This protein plays a crucial role in migration through the extracellular matrix and basement

membrane proteins. This may facilitate the access of parasites to the blood or lymph circulation, enabling dissemination to distant sites (Ghosh *et al.*, 1999). The gp63 homologue in schistosomes has been expanded to at least 12 putative family members compared to a single orthologue in humans, fruit fly and *C. elegans* (Brindley *et al.*, 2009; Zhou *et al.*, 2009). Functional studies on schistosome cercaria have found that this protein may facilitate tissue invasion (Curwen *et al.*, 2006). The homologue to the metalloprotease leishmanolysin in *Entamoeba histolytica*, MSP-1 (*EhMSP-1*), was shown to be a surface metalloprotease involved in the regulation of amoebic adherence and cell motility (Teixeira *et al.*, 2012).

The coding sequences for peptidyl-dipeptidase A (angiotensin-converting enzyme, peptidyl-dipeptidase A, EC 3.4.15.1) were identified in *A. cantonensis* with low levels of expression. *Caenorhabditis elegans* peptidyl-dipeptidase A lacks crucial active site residues and cannot function as a typical ACE. However, its coding gene is essential for larval development and adult morphogenesis, and is expressed in hypodermal cells, the developing vulva, and ray papillae of the male tail (Brooks *et al.*, 2003).

Finally, we annotated the coding sequence transcripts for HSP70 in *A. cantonensis*, the homologue for HSP70 of *Toxoplasma gondii* (TgHSp70), which is also known as TgHSp70 cyclophilin 18 (TgCyp18). TgHSp70 is highly antigenic, and mice immunized with rTgHSp70 or rTgHSp70 were found to exhibit a significantly reduced number of cysts in the brains and reduced tissue damage (Czarnewski *et al.*, 2017). The secreted *T. gondii* TgHSp70 is involved in the triggering of cysteine–cysteine chemokine receptor 5 (CCR5) in dendritic cells and macrophages, and induces interleukin-12 (Ibrahim *et al.*, 2009). In *H. contortus*, HSP70 is expressed in all larval stages and adult forms, with the highest expression found in eggs (Zhang *et al.*, 2013). It is one of the most abundant proteins in *S. mansoni* egg secretions (Cass *et al.*, 2007), and was found in the somatic extracts of the larval and adult stages of *T. spiralis*. Mice vaccinated with *T. spiralis* HSP70 exhibit strong humoral immune responses and significant reduction in the number of parasites (Wang *et al.*, 2009). Moreover, HSP70 is a potent immunogen in several parasitic infections, such as *B. malayi* (Selkirk *et al.*, 1989), *Onchocerca volvulus* (Rothstein *et al.*, 1989), *Wuchereria bancrofti* (Ravi *et al.*, 2004), *Echinostoma caproni* (Higón *et al.*, 2008) and *Litomosoides sigmodontis* (Hartmann *et al.*, 2014). However, further investigation is needed in order to evaluate *A. cantonensis* Hsp70 as a potential vaccine candidate.

In the current study, several highly expressed genes involved in essential pathways for *A. cantonensis* survival and immune system modulation inside of blood vessels of *R. norvegicus*, some of these mechanisms appear conserved during evolution, because they have already been described in other parasites. Here, the full-length transcripts were also identified, which may contribute to the development of vaccine candidates and aid functional genomics studies in elucidating their biological roles. Overall, this study provides a comprehensive source of open data for use in future research on *Angiostrongylus* nematodes.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0031182021000469>

**Data.** The raw FASTQ files analysed during the current study are publicly available and have been deposited in BioProject under accession number PRJNA549533. The FASTA files of all ORFs and functional annotations generated during this study are publicly available at <http://angiostrongylus.lad.pucrs.br/admin/welcome>.

**Acknowledgements.** The authors would like to thank the Laboratório de Alto Desempenho (LAD) of the PUCRS Polytechnic School and the High-Performance Computing Technician Rafael Lorenzo Bellé for all

technical support and Dr Malcolm Jones for the fruitful discussions about *A. cantonensis* biology. The graphical abstract was created with [BioRender.com](https://BioRender.com).

**Author contributions.** L. M. P. was responsible for conceptualization and design/analysis of bioinformatics, interpretation of data, wrote and revision the article. M. P. G. J. carried out web database construction, took part in writing of the article and revision. B. D. B. participated in part of bioinformatic analysis and revision of the article. A. R. R. and A. B. Z. participated in part of bioinformatic analysis. C. C. T. was responsible for the conceptualization and design of the project, search for financial support, participated in the writing and revision of the manuscript. A. L. M. was responsible for the conceptualization and design of the project, responsible for all experimental laboratory analyses, search for financial support, participated in the writing and revision of the article.

**Financial support.** This study was supported by the Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico do Brasil [Grant numbers CNPq 401904/2013-0 (2013) and CNPq PQ1D 307005/2014-3 (2014)] and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Brasil (Edital 32, 2010).

**Conflict of interest.** The authors declare there are no conflicts of interest.

## References

- Akasaka M, Kato KH, Kitajima K and Sawada H (2013) Identification of novel isoforms of vitellogenin expressed in ascidian eggs. *Journal of Experimental Zoology* **320**, 118–128.
- Akiyoshi S, Nomura KH, Dejima K, Murata D, Matsuda A, Kanaki N, Takaki T, Mihara H, Nagaishi T, Furukawa S, Ando KG, Yoshina S, Mitani S, Togayachi A, Suzuki Y, Shikanai T, Narimatsu H and Nomura K (2015) RNAi screening of human glycogene orthologs in the nematode *Caenorhabditis elegans* and the construction of the *C. elegans* glycogene database. *Glycobiology* **25**, 8–20.
- Alexa A, Rahnenführer J and Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics (Oxford, England)* **22**, 1600–1607.
- Alicata JE (1965) Biology and distribution of the rat lungworm, *Angiostrongylus cantonensis*, and its relationship to eosinophilic meningoencephalitis and other neurological disorders of man and animals. *Advances in Parasitology* **3**, 223–248.
- Barrett J (1983) Lipid metabolism. In: Arme C and Pappas PW (eds), *Biology of the Eucestoda*. London: Academic Press, vol. 2, pp. 391–419.
- Bolger AM, Lohse M and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114–2120.
- Bouchery T, Filbey K, Shepherd A, Chandler J, Patel D, Schmidt A, Camberis M, Peignier A, Smith AA, Johnston K, Painter G, Pearson M, Giacomini P, Loukas A, Bottazzi ME, Hotez P and LeGros G (2018) A novel blood-feeding detoxification pathway in *Nippostrongylus brasiliensis* L3 reveals a potential checkpoint for arresting hookworm development. *PLoS Pathogens* **14**, e1006931.
- Brindley PJ, Mitreva M, Ghedin E and Lustigman S (2009) Helminth genomics: the implications for human health. *PLoS Neglected Tropical Diseases* **3**, e538. <https://doi.org/10.1371/journal.pntd.0000538>
- Brooks DR, Appleford PJ, Murray L and Isaac RE (2003) An essential role in molting and morphogenesis of *Caenorhabditis elegans* for ACN-1, a novel member of the angiotensin-converting enzyme family that lacks a metallo-peptidase active site. *Journal of Biological Chemistry* **278**, 52340–52346.
- Burns RE, Bicknese EJ, Qvarnstrom Y, DeLeon-Carnes M, Drew CP, Gardiner CH and Rideout BA (2014) Cerebral *Angiostrongylus cantonensis* infection in a captive African pygmy falcon (*Polihierax semitorquatus*) in southern California. *Journal of Veterinary Diagnostic Investigation* **26**, 695–698.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.
- Cass CL, Johnson JR, Califf LL, Xu T, Hernandez HJ, Stadecker MJ, Yates JR and Williams DL (2007) Proteomic analysis of *Schistosoma mansoni* egg secretions. *Molecular and Biochemical Parasitology* **155**, 84–93.
- Červená B, Modrý D, Fecková B, Hrazdilová K, Foronda P, Alonso AM, Lee R, Walker J, Niebuhr CN, Malik R and Šlapeta J (2019) Low diversity of *Angiostrongylus cantonensis* complete mitochondrial DNA sequences from Australia, Hawaii, French Polynesia and the Canary Islands revealed using whole genome next-generation sequencing. *Parasites and Vectors* **12**, 1–13.
- Cheng X, Xiang Y, Xie H, Xu CL, Xie TF, Zhang C and Li Y (2013) Molecular characterization and functions of fatty acid and retinoid binding protein gene (Ab-far-1) in *Aphelenchoides besseyi*. *PLoS One* **8**, 1–9.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)* **21**, 3674–3676.
- Cottee PA, Nisbet AJ, Boag PR, Larsen M and Gasser RB (2004) Characterization of major sperm protein genes and their expression in *Oesophagostomum dentatum* (Nematoda: Strongylida). *Parasitology* **129**, 479–490.
- Curwen RS, Ashton PD, Sundaralingam S and Wilson RA (2006) Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Molecular & Cellular Proteomics* **5**, 835–844.
- Czarnewski P, Araújo ECB, Oliveira MC, Mineo TWP and Silva NM (2017) Recombinant TgHSP70 immunization protects against *Toxoplasma gondii* brain cyst formation by enhancing inducible nitric oxide expression. *Frontiers in Cellular and Infection Microbiology* **7**, 1–13.
- do Espírito-Santo MCC, Pinto PLS, da Mota DJG and Gryscek RCB (2013) The first case of *Angiostrongylus cantonensis* eosinophilic meningitis diagnosed in the city of São Paulo, Brazil. *Revista do Instituto de Medicina Tropical de São Paulo* **55**, 129–132.
- Duan J, Flock K, Jue N, Zhang M, Jones A, Al Seesi S, Mandoiu I, Pillai S, Hoffman M, O'Neill R, Zinn S, Govoni K, Reed S, Jiang H, Jiang Z and Tian X (2019) Dosage compensation and gene expression of the X chromosome in sheep. *G3 (Bethesda)* **9**, 305–314.
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladín L, Piovesan D, Tosatto SCE and Finn RD (2019) The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427–D432.
- Esko JD and Selleck SB (2002) Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annual Review of Biochemistry* **71**, 435–471.
- Fairfax KC, Vermeire JJ, Harrison LM, Bungiro RD, Grant W, Husain SZ and Cappello M (2009) Characterisation of a fatty acid and retinoid binding protein orthologue from the hookworm *Ancylostoma ceylanicum*. *International Journal for Parasitology* **39**, 1561–1571.
- Galvin B, Mizuguchi S, Uyama T, Kitagawa H, Nomura KKH, Dejima K, Gengyo-Ando K, Mitani S and Sugahara K (2003) Chondroitin proteoglycans are involved in cell division of *Caenorhabditis elegans*. *Nature* **423**, 443–448.
- Ghosh A, Bandyopadhyay K, Kole L and Das PK (1999) *Leishmania donovani* that may mediate cell adhesion. *Biochemical Journal* **337**, 551–558.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli E, Hachohen N, Gnirke A, Rhind N, di Palma, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652.
- Graeff-Teixeira C, Da Silva ACA and Yoshimura K (2009) Update on eosinophilic meningoencephalitis and its clinical relevance. *Clinical Microbiology Reviews* **22**, 322–348.
- Han SM, Cottee PA and Miller MA (2010) Sperm and oocyte communication mechanisms controlling *C. elegans* fertility. *Developmental Dynamics* **239**, 1265–1281.
- Hartmann W, Singh N, Rathaur S, Brenz Y, Liebau E, Fleischer B and Breloer M (2014) Immunization with *Brugia malayi* Hsp70 protects mice against *Litomosoides sigmodontis* challenge infection. *Parasite Immunology* **36**, 141–149.
- He H, Cheng M, Yang X, Meng J, He A, Zheng X, Li Z, Guo P, Pan Z and Zhan X (2009) Preliminary molecular characterization of the human pathogen *Angiostrongylus cantonensis*. *BMC Molecular Biology* **10**, 97.
- Higón M, Monteagudo C, Fried B, Esteban JG, Toledo R and Marcilla A (2008) Molecular cloning and characterization of *Echinostoma caproni* heat shock protein-70 and differential expression in the parasite derived from low- and high-compatible hosts. *Parasitology* **135**, 1469–1477.
- Ibrahim HM, Bannai H, Xuan X and Nishikawa Y (2009) *Toxoplasma gondii* cyclophilin 18-mediated production of nitric oxide induces bradyzoite

- conversion in a CCR5-dependent manner. *Infection and Immunity* 77, 3686–3695.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R and Hunter S (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* 30, 1236–1240.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M and Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44, D457–D462.
- King KL, Essig J, Roberts TM and Moerland TS (1994) Regulation of the *Ascaris major* sperm protein (MSP) cytoskeleton by intracellular pH. *Cell Motility Cytoskeleton* 27, 193–205.
- Kiontke K, Barrière A, Kolotuev I, Podbilewicz B, Sommer R, Fitch DHA and Félix MA (2007) Trends, stasis, and drift in the evolution of nematode vulva development. *Current Biology* 17, 1925–1937.
- Kreuger J and Kjellén L (2012) Heparan sulfate biosynthesis: regulation and variability. *Journal of Histochemistry & Cytochemistry* 60, 898–907.
- Kuwabara PE (2003) The multifaceted *C. elegans* major sperm protein: an ephrin signaling antagonist in oocyte maturation. *Genes & Development* 17, 155–161.
- Luck AN, Yuan X, Voronin D, Slatko BE, Hamza I and Foster JM (2016) Heme acquisition in the parasitic filarial nematode *Brugia malayi*. *The FASEB Journal* 30, 3501–3514.
- Martin-Alonso A, Foronda P, Quispe-Ricalde MA, Feliu C and Valladares B (2011) Seroprevalence of *Angiostrongylus cantonensis* in wild rodents from the Canary Islands. *PLoS One* 6, 2–6.
- McGwire BS, O'Connell WA, Chang KP and Engman DM (2002) Extracellular release of the glycosylphosphatidylinositol (GPI)-linked *Leishmania* surface metalloprotease, gp63, is independent of GPI phospholipolysis. Implications for parasite virulence. *Journal of Biological Chemistry* 277, 8802–8809.
- Miller MA, Nguyen VQ, Lee MH, Kosinski M, Schedl T, Caprioli RM and Greenstein D (2001) A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science (New York, N.Y.)* 291, 2144–2147.
- Morassutti AL, Perelygin A, De Carvalho MO, Lemos LN, Pinto PM, Frace M, Wilkins PP, Graeff-Teixeira C and Da Silva AJ (2013) High throughput sequencing of the *Angiostrongylus cantonensis* genome: a parasite spreading worldwide. *Parasitology* 140, 1304–1309.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC and Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35, W182–W185.
- Murata D, Nomura KH, Dejima K, Mizuguchi S, Kawasaki N, Matsuihi-Nakajima Y, Ito S, Gengyo-Ando K, Kage-Nakadai E, Mitani S and Nomura K (2012) GPI-anchor synthesis is indispensable for the germline development of the nematode *Caenorhabditis elegans*. *Molecular Biology of the Cell* 23, 982–995.
- Noborn F, Gomez Toledo A, Nasir W, Nilsson J, Dierker T, Kjellén L and Larson G (2018) Expanding the chondroitin glycoproteome of *Caenorhabditis elegans*. *Journal of Biological Chemistry* 293, 379–389.
- Payne SH and Loomis WF (2006) Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryotic Cell* 5, 272–276.
- Piras V, Chiow A and Selvarajoo K (2018) Long range order and short range disorder in *Saccharomyces cerevisiae* biofilm. *Engineering Biology* 3, 3, 2019.
- Qiao F, Luo L, Peng H, Luo S, Huang W, Cui J, Li X, Kong L, Jiang D, Chitwood DJ and Peng D (2016) Characterization of three novel fatty acid- and retinoid-binding protein genes (Ha-far-1, Ha-far-2 and Hf-far-1) from the cereal cyst nematodes *Heterodera avenae* and *H. filipjevi*. *PLoS One* 11, 1–17.
- Rao AU, Carta LK, Lesuisse E and Hamza I (2005) Lack of heme synthesis in a free-/living eukaryote. *Proceedings of the National Academy of Sciences of the United States of America* 102, 4270–4275.
- Rey-Burusco MF, Ibanez-Shimabukuro M, Gabrielsen M, Franchini GR, Roe AJ, Griffiths K, Zhan B, Cooper A, Kennedy MW, Corsico B and Smith BO (2015) Diversity in the structures and ligand-binding sites of nematode fatty acid and retinol-binding proteins revealed by Na-FAR-1 from *Necator americanus*. *Biochemical Journal* 471, 403–414.
- Rothstein NM, Higashi G, Yates J and Rajan TV (1989) *Onchocerca volvulus* heat shock protein 70 is a major immunogen in amicrofilaric individuals from a filariasis-endemic area. *Molecular and Biochemical Parasitology* 33, 229–235.
- Sadler JE (2008) Biochemistry and genetics of von Willebrand factor. *Annual Review of Biochemistry* 67, 395–424.
- Sahoo S, Murugavel S, Devi IK, Vedomurthy GV, Gupta SC, Singh BP and Joshi P (2013) Glyceraldehyde-3-phospho dehydrogenase of the parasitic nematode *Haemonchus contortus* binds to complement C3 and inhibits its activity. *Parasite Immunology* 35, 457–467.
- Selkirk ME, Denham DA, Partono F and Maizels RM (1989) Heat shock cognate 70 is a prominent immunogen in *Brugian filariasis*. *Journal of Immunology* 143, 299–308.
- Shadeo A, Chari R, Vatcher G, Campbell J, Lonergan KM, Maticic J, van Niekerk D, Ehlen T, Miller D, Follen M, Lam WL and MacAulay C (2007) Comprehensive serial analysis of gene expression of the cervical transcriptome. *BMC Genomics* 8, 1–11.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)* 31, 3210–3212.
- Spinner WG, Thompson FJ, Emery DC and Viney ME (2012) Characterization of genes with a putative key role in the parasitic lifestyle of the nematode *Strongyloides ratti*. *Parasitology* 139, 1317–1328.
- Stoeckius M, Grün D and Rajewsky N (2014) Paternal RNA contributions in the *Caenorhabditis elegans* zygote. *The EMBO Journal* 33, 1740–1750.
- Strube C, Buschbaum S and Schnieder T (2009) Molecular characterization and real-time PCR transcriptional analysis of *Dictyocaulus viviparus* major sperm proteins. *Parasitology Research* 104, 543–551.
- Sun C, Hu L, Liu S, Gao Z and Zhang S (2013) Functional analysis of domain of unknown function (DUF) 1943, DUF1944 and von Willebrand factor type D domain (VWD) in vitellogenin2 in zebrafish. *Developmental and Comparative Immunology* 41, 469–476.
- Teien AN, Abildgaard U and Höök M (1976) The anticoagulant effect of heparan sulfate and dermatan sulfate. *Thrombosis Research* 8, 859–867.
- Teixeira JE, Sateriale A, Bessoff KE and Huston CD (2012) Control of *Entamoeba histolytica* adherence involves metalloprotease 1, an M8 family surface metalloprotease with homology to leishmanolysin. *Infection and Immunity* 80, 2165–2176.
- Tyagi R, Joachim A, Rutkowski B, Rosa BA, Martin JC, Hallsworth-Pepin K, Zhang X, Ozersky P, Wilson RK, Ranganathan S, Sternberg PW, Gasser RB and Mitreva M (2015a) Cracking the nodule worm code advances knowledge of parasite biology and biotechnology to tackle major diseases of livestock. *Biotechnology Advances* 33, 980–991.
- Tyagi R, Rosa BA, Lewis WG and Mitreva M (2015b) Pan-phylum comparison of nematode metabolic potential. *PLoS Neglected Tropical Diseases* 9, 1–32.
- van den Elsen S, Holovachov O, Karssen G, van Megen H, Helder J, Bongers T, Bakker J, Holterman M and Mooyman P (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 11, 927–950.
- Vedomurthy GV, Sahoo S, Devi IK, Murugavel S and Joshi P (2015) The N-terminal segment of glyceraldehyde-3-phosphate dehydrogenase of *Haemonchus contortus* interacts with complements C1q and C3. *Parasite Immunology* 37, 568–578.
- Verissimo CM, Graeff-Teixeira C, Jones MK and Morassutti AL (2019) Glycans in the roles of parasitological diagnosis and host-parasite interplay. *Parasitology* 146, 1217–1232.
- Wang QP, Lai DH, Zhu XQ, Chen XG and Lun ZR (2008) Human angiostrongyliasis. *Lancet Infectious Diseases* 8, 621–630.
- Wang S, Zhu X, Yang Y, Yang J, Gu Y, Wei J, Hao R, Boireau P and Cui S (2009) Molecular cloning and characterization of heat shock protein 70 from *Trichinella spiralis*. *Acta Tropica* 110, 46–51.
- Whittaker CA and Hynes RO (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Molecular Biology of the Cell* 13, 3369–3387.
- Xu L, Xu M, Sun X, Xu J, Zeng X, Shan D, Yuan D, He P, He W, Yang Y, Luo S, Wei J, Wu X, Liu Z, Xu X, Dong Z, Song L, Zhang B, Yu Z, Wang L, Zhang C, Fang X, Gao Q, Lv Z and Wu Z (2019) The genetic basis of adaptive evolution in parasitic environment from the *Angiostrongylus cantonensis* genome. *PLoS Neglected Tropical Diseases* 13, e0007846.
- Yong HS, Eamsobhana P, Lim PE, Razali R, Aziz FA, Rosli NSM, Poole-Johnson J and Anwar A (2015a) Draft genome of neurotropic nematode parasite *Angiostrongylus cantonensis*, causative agent of human eosinophilic meningitis. *Acta Tropica* 148, 51–57.
- Yong HS, Song SL, Eamsobhana P, Goh SY and Lim PE (2015b) Complete mitochondrial genome reveals genetic diversity of *Angiostrongylus cantonensis* (Nematoda: Angiostrongylidae). *Acta Tropica* 152, 157–164.
- Yu L, Cao B, Long Y, Tukayo M, Feng C, Fang W and Luo D (2017) Comparative transcriptomic analysis of two important life stages of

- Angiostrongylus cantonensis*: fifth-stage larvae and female adults. *Genetics and Molecular Biology* **40**, 540–549.
- Zhan B, Arumugam S, Kennedy MW, Tricoche N, Lian LY, Asojo OA, Bennuru S, Bottazzi ME, Hotez PJ, Lustigman S and Klei TR (2018) Ligand binding properties of two *Brugia malayi* fatty acid and retinol (FAR) binding proteins and their vaccine efficacies against challenge infection in gerbils. *PLoS Neglected Tropical Diseases* **12**, e0006772.
- Zhang H, Zhou Q, Yang Y, Chen X, Yan B and Du A (2013) Characterization of heat shock protein 70 gene from *Haemonchus contortus* and its expression and promoter analysis in *Caenorhabditis elegans*. *Parasitology* **140**, 683–694.
- Zhou Y, Zheng H, Chen Y, Zhang L, Wang K, Guo J, Huang Z, Zhang B, Huang W, Jin K, Dou T, Hasegawa M, Wang L, Zhang Y, Zhou J, Tao L, Cao Z, Li Y, Vinar T, Brejova B, Brown D, Li M, Miller DJ, Blair D, Zhong Y, Chen Z, Liu F, Hu W, Wang ZQ, Zhang QH, Song HD, Chen S, Xu X, Xu B, Ju C, Huang Y, Brindley PJ, McManus DP, Feng Z, Han ZG, Lu G, Ren S, Wang Y, Gu W, Kang H, Chen J, Chen X, Chen S, Wang, L, Yan J, Wang, B, Lv X, Jin L, Wang, B, Pu S, Zhang X, Zhang W, Hu Q, Zhu G, Wang, J, Yu J, Wang, J, Yang H, Ning Z, Beriman M, Wei CL, Ruan Y, Zhao G and Wang S (2009) The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* **460**, 345–351.