RESEARCH

Open Access

CrossMark

CAP protein superfamily members in *Toxocara canis*

Andreas J. Stroehlein^{1*}, Neil D. Young¹, Ross S. Hall¹, Pasi K. Korhonen¹, Andreas Hofmann^{1,2}, Paul W. Sternberg³, Abdul Jabbar¹ and Robin B. Gasser^{1*}

Abstract

Background: Proteins of the cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 (CAP) superfamily are recognized or proposed to play roles in parasite development and reproduction, and in modulating host immune attack and infection processes. However, little is known about these proteins for most parasites.

Results: In the present study, we explored CAP proteins of *Toxocara canis*, a socioeconomically important zoonotic roundworm. To do this, we mined and curated transcriptomic and genomic data, predicted and curated full-length protein sequences (n = 28), conducted analyses of these data and studied the transcription of respective genes in different developmental stages of *T. canis*. In addition, based on information available for *Caenorhabditis elegans*, we inferred that selected genes (including *lon-1*, *vap-1*, *vap-2*, *scl-1*, *scl-8* and *scl-11* orthologs) of *T. canis* and their interaction partners likely play central roles in this parasite's development and/or reproduction via TGF-beta and/or insulin-like signaling pathways, or via host interactions.

Conclusion: In conclusion, this study could provide a foundation to guide future studies of CAP proteins of *T. canis* and related parasites, and might assist in finding new interventions against diseases caused by these parasites.

Keywords: Nematodes, Toxocara canis, CAP (SCP/TAPS) protein genes, Transcription profiles, Genetic interactions

Background

Toxocara canis (Werner, 1782) of canids is recognized as the principal causative agent of toxocariasis; this parasitic worm has a complex life cycle, which can also involve rodents and other animals as paratenic hosts [1]. In humans, particularly children, following the accidental ingestion of infective eggs of T. canis, infective larvae penetrate the intestinal wall and invade various tissues, leading to ocular larva migrans (OLM), visceral larva migrans (VLM), neurotoxocariasis (NT) and/or covert toxocariasis (CT) [2]. Some reports (e.g. [3, 4]) have also shown an association between T. canis infection or toxocariasis and allergic disorders in humans, such as urticaria, chronic pruritus and/or asthma. The T. canis-mouse and -dog models [5, 6] provide practical tools for studying T. canis biology, parasite-host interactions and human toxocariasis.

* Correspondence: astroehlein@unimelb.edu.au; robinbg@unimelb.edu.au ¹Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia

Full list of author information is available at the end of the article

Some excretory/secretory (ES) proteins from T. canis likely play key roles in these interactions and the infection process, supported, to some extent, by recent research showing abundant transcription of genes encoding peptidases, cell-adhesion molecules (including integrins and cadherins) and lectins (C-type) in parasitic stages of T. canis [7]. Moreover, the cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 (CAP) superfamily [8], defined based on sperm-coating protein (SCP)-like extracellular domains (InterProScan codes: IPR014044 and IPR001283), represents some ES molecules inferred from the genome of T. canis (cf. [7]). Although six CAP proteins encoded in the draft genome of T. canis had been described and were predicted to be ES molecules [7], preliminary curation work suggested the presence of a larger panel of related molecules in this worm (Stroehlein et al., unpublished). In spite of the major socioeconomic importance of toxocariasis globally and some promise for CAP proteins as drug or vaccine targets (e.g. [9-11]), detailed information on this group of proteins in ascaridoids has been lacking.



© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated. The completion of a draft genome and transcriptomes for *T. canis* [7] using an Illumina-based sequencing approach provides an opportunity to explore, in detail, these CAP proteins and other key protein groups of this nematode. Therefore, in the present study, using available data sets, we now define the complement of CAP proteins, establish their sequence features, explore the transcription profiles of their genes in selected life cycle stages and tissues, and predict their roles in *T. canis*. This work should guide structural and functional studies of CAP proteins of *T. canis*, and might assist in finding new interventions against toxocariasis.

Methods

For the present study, we utilized data from the published draft genome and developmental transcriptomes of T. canis from Denmark (NCBI BioProject accession no. PRJNA248777; WormBase [7, 12]); the draft genome of this parasite is 317 Mb and is predicted to have 18,596 protein-coding genes. Amino acid (aa) sequences (inferred from coding gene regions and de novo-assembled transcripts; cf. [7]) were subjected to protein domain/family/profile searches using InterProScan v.5.15.54 utilizing default settings [13]. Sequences with homology to one or more domains present in known CAP proteins were extracted, and de novo-assembled transcripts mapped to genomic scaffolds using BLAT [14]. These BLAT-transcript alignments (PSL files) as well as RNA-sequence reads that mapped to the genome (BAM files; cf. [7]) and original gene predictions (Maker2 GFF file; cf. [7]) were then loaded into the Integrative Genomics Viewer (IGV; [15]). All of this gene prediction information was then integrated to verify and define individual genes and corresponding transcripts. Final open reading frames (ORFs) were inferred using the program EMBOSS getorf v.6.4.0.0 [16].

Quality-filtered, decontaminated paired-end RNAsequence reads were mapped to all 19,004 transcript sequences of *T. canis* [7]. For all life cycle stages and tissues investigated herein, genewise transcripts per million (TPM) values were calculated using the program RSEM v.1.2.11 [17]. These TPM values were then log-transformed [log_e(TPM + 1)] to display transcription profiles for CAP protein-encoding genes in a heat map utilizing the gplots package (heatmap.2 function) in R v.3.2.4 [18].

Differential transcription analysis was conducted employing the program edgeR v.3.12.1 [19], using expected count values from RSEM as input. Genes were included in the analysis if they had more than one read count per million (CPM) in at least three samples (= biological replicates; cf. [7]). Differential transcription of genes was computed using a general dispersion of 0.43, considering a false discovery rate (FDR) value of \leq 0.01 to be statistically significant. Log₂-fold change (FC) values were employed to infer an upregulation (log₂-fold change: > 2) of transcription of genes based on pairwise comparisons of stages/sexes/tissues.

Genetic interactions were inferred based on network analysis described previously [20, 21]. Identities between aa sequences encoded in *T. canis* and those from other free-living and parasitic worms represented in ParaSite within WormBase (WBPS5) were established using BLASTp v.2.2.28+ [22]. Pairwise global aa sequence similarities were calculated using the program EMBOSS Needle v.6.4.0.0 [16].

Results and discussion

CAP proteins encoded in *T. canis* and their features

Homology-based profile/domain searches using InterProScan and manual sequence curation predicted 28 CAP protein sequences (cf. [8]), which mapped to 23 genomic scaffolds (Additional file 1: Table S1). Subsequent mapping of de novo-assembled transcripts to these genomic scaffolds revealed 23 full-length (with start and stop codons at the beginning and end of the coding sequence, respectively) and five partial transcripts (Additional file 1: Table S1). Of the 28 protein sequences encoded by these transcripts, 25 were single- and three were double-domain CAP proteins, respectively. The features of these 28 predicted proteins (including length, InterProScan domains and signal peptides) are summarized in Additional file 1: Table S1. Specifically, the single-domain proteins (72 to 546 aa in total length; average length: 225 aa) were usually shorter than the double-domain proteins (416 to 490 aa in total length; average length: 449 aa) and exhibited 10-87 % sequence similarity among each other, whereas double-domain sequences had 21-44 % sequence similarity among each other upon overall pairwise aa sequence comparison.

Based on InterProScan analysis, 22 of the 28 predicted proteins belonged to the "cysteine-rich secretory protein, allergen V5/Tpx-1-related subfamily" (PTHR10334), of which 16 and 5 had the "allergen V5/Tpx-1 family" (PR00837) and "venom-allergen 5" (PR00838) signatures, respectively (Additional file 1: Table S1). All 28 proteins had the InterProScan signature IPR014044 (CAP) and contained a PR-1-like (SSF55797) domain. Additionally, 20 sequences had an SCP (SM00198), 27 a pathogenesis-related (G3DSA:3.40.33.10) and 25 a CAP (PF00188) domain (Additional file 1: Table S1). In addition, seven singledomain and two double-domain proteins belonged to "allergen V5/Tpx-1 related, conserved site CRISP family 1" (PS01009). In contrast, only two single-domain proteins and none of the double-domain proteins were assigned to "allergen V5/Tpx-1 related, conserved site CRISP family 2" (PS01010). Moreover, 11 single-domain proteins had a signal peptide that was absent from the other 14 singledomain proteins and all three double-domain proteins (Additional file 1: Table S1).

Amino acid sequence comparisons of *T. canis* CAP proteins with those of free-living and parasitic worms showed that most (n = 21) *T. canis* CAP proteins shared the highest sequence identity (36-89 %) to those of *Ascaris suum* (*Asu*) or *A. lumbricoides* (*Al*), ascaridoids of pigs and humans, respectively. Close homologs of one to two *T. canis* CAP proteins (39-83 % aa identity) were identified in the equine ascaridoid *Parascaris equorum* (*Pe*), the free-living nematode *Panagrellus redivivus* (*Pr*) and the entomopathogenic nematodes *Steinernema monticolum* (*Smo*) and *Steinernema scapterisci* (*Ss*) (Additional file 1: Table S2). Next, we explored transcription profiles for all 28 individual genes in different life stages/sexes/tissues.

Transcription profiles

Subsets of the 28 genes encoding CAP proteins were differentially transcribed between/among different developmental stages, sexes and/or tissues (Fig. 1; Additional file 1: Table S3). Seven genes (*Tc-cap-4, Tc-cap-7, Tc-cap-16, Tc-cap-19, Tc-cap-20, Tc-cap-21* and *Tc-cap-22*) were consistently and highly upregulated in the third larval (L3) stage, with respect to all other developmental stages and tissues (FC-range: 2.3 to 15.8; Fig. 1; Additional file 1: Table S3).

Two additional genes were significantly upregulated in L3 (FC-range: 3.5 to 6.2); one (Tc-cap-18) with respect to all adult male tissues and another (Tc-cap-28) in relation to all adult female tissues. Conversely, a significant upregulation (FC-range: 3.8 to 16.6) was seen in the adult male stage (whole-worm) of T. canis for Tc-cap-5, Tc-cap-6, Tc-cap-13, Tc-cap-14, Tc-cap-25 and Tc-cap-26 compared with all other stages and tissues. Thereof, four genes (Tc-cap-13, Tc-cap-14, Tc-cap-25 and Tc-cap-26) were also highly transcribed in the gut and significantly upregulated compared with all other stages/tissues, except male whole-worm (Fig. 1), suggesting a role for these genes in digestion, but also in other processes, given the major upregulation in male whole-worm compared with male gut. Additionally, although not significantly upregulated with respect to the other two adult male tissues, Tc-cap-11 was upregulated in the adult male whole-worm (FC-range: 4.8 to 8.0) compared with the female and L3 stages. Taken together, these findings suggest that these seven molecules play key roles in male reproduction, digestion and/or parasite-host interactions.

Although no significant female-enriched or -specific genes were inferred, *Tc-cap-2* was upregulated in both male and female whole-worms with respect to adult

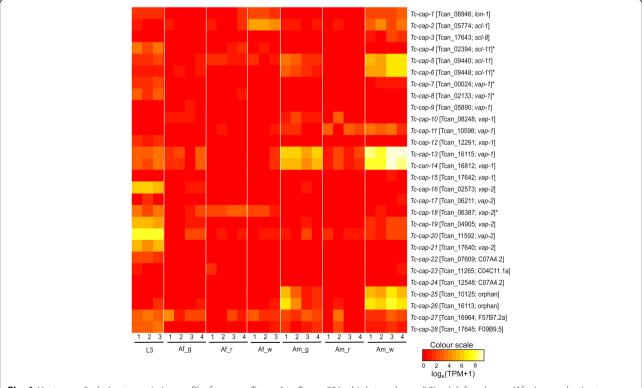


Fig. 1 Heat map displaying transcription profiles for genes *Tc-cap-1* to *Tc-cap-28* in third-stage larvae (L3); adult female: gut (Af_g), reproductive tract (Af_r) and whole-worm (Af_w); adult male: gut (Am_g), reproductive tract (Am_r) and whole-worm (Am_w) of *Toxocara canis*; biological replicates are indicated (Lanes 1 to 3 or 1 to 4). The identifiers and closest *Caenorhabditis elegans* homologs of individual *Tc-cap* genes are in square brackets and partial sequences are indicated with an asterisk. The heat map was drawn based on log_e-transformed "transcripts per million" (TPM) values (cf. Additional file 1: Table S4); a colour-scale indicates the level of transcription: low (*red*), medium (*orange*), high (*yellow*) and very high (*white*)

reproductive and gut tissues and the L3 stage, suggesting an exclusive role of this gene in both sexes of adult worms in tissues other than those studied, possibly by regulating longevity and/or stress resistance in mature worms, similar to the role that its ortholog (*scl-1*) plays in *Caenorhabditis elegans* [23].

The transcription profiles for *cap* genes of *T. canis* appear to be quite different from those of blood-feeding (haematophagous) nematodes, including the hookworms Ancylostoma caninum, A. ceylanicum and Necator americanus, in which cap genes are upregulated predominantly in parasitic larval stages [24-29], as well as the barber's pole worm, Haemonchus contortus, where a substantial proportion of cap genes are highly transcribed in haematophagous stages [20]. There are also major differences in the numbers of CAP protein-encoding genes in A. ceylani*cum* (n = 432; [29]), *N. americanus* (n = 137; [30]) and *H. contortus* (*n* = 45; [20]) compared with *T. canis.* In addition, the relatively high proportion (93 %; n = 26; Fig. 1) of CAP proteins of T. canis with homologs in C. elegans contrasts the situation for A. ceylanicum and N. americanus [29, 30], for which only 14 of 432 (3 %) and six of 137 (4 %) CAP protein homologs, respectively, were identified in the freeliving nematode. Taken together, these qualitative and quantitative differences indicate considerable variation among these nematodes in their biology and immune recognition, modulation or evasion mechanisms.

Functional inferences

Of all 28 predicted CAP proteins of T. canis, 26 had C. elegans homologs (Additional file 1: Table S2). Specifically, all three double-domain proteins of T. canis had homologs in C. elegans, including Ce-VAP-1 (two homologs: Tc-CAP-7 and Tc-CAP-15), Ce-VAP-2 (one homolog: Tc-CAP-16) (Additional file 1: Table S2). Additionally, seven single-domain Ce-VAP-1 homologs and five single-domain Ce-VAP-2 homologs were detected (Additional file 1: Table S2). Of the remaining 13 single-domain proteins of T. canis, 11 had homologs in C. elegans, namely Ce-LON-1 (Tc-CAP-1), Ce-SCL-1 (Tc-CAP-2), Ce-SCL-8 (Tc-CAP-3), Ce-SCL-11 (Tc-CAP-4 to Tc-CAP-6), C07A4.2 (Tc-CAP-22 and Tc-CAP-24), C04C11.1a (Tc-CAP-23), F57B7.2 (Tc-CAP-27) and F09B9.5 (Tc-CAP-28) (Additional file 1: Table S2). For the two shortest full-length T. canis CAP protein sequences (Tc-CAP-25 and Tc-CAP-26; 90-95 aa), we did not detect any C. elegans homologs and thus labelled them "orphan" molecules (Fig. 1). However, both sequences did have homologs in a range of parasitic nematodes (see Additional file 1: Table S2), and Tc-CAP-26 also shared moderate aa sequence identity (46 %) with a sequence of the free-living nematode Panagrellus redivivus.

As most (26 of 28) *Tc*-CAP proteins predicted have homologs in *C. elegans* (Additional file 1: Table S2), we elected to infer functions from information available for the free-living worm in WormBase. We propose that the *C. elegans* LON-1 homolog, *Tc*-CAP-1, regulates body length and polyploidisation in hypodermal cells in adult and larval *T. canis*. In *C. elegans*, the protein encoded by *lon-1* is a target of TGF-beta signaling and is expressed in hypodermal tissues; the *lon-1* gene is negatively regulated by Sma/Mab pathway signaling and epistatic to *dbl-1* [31–33]. Knockdown of *lon-1* function in *C. elegans* by double-stranded RNA interference (dsRNAi) results in long worms, indicating that it is a negative regulator of body length [32]. *Ce-lon-1* encodes a protein (312 aa) with a sequence motif (GHYVQVVW) that is conserved in *T. canis Tc-cap-1* (284 aa) and across many other metazoan organisms (e.g. [32, 34–38]).

The *lon-1* gene interacts with *sma-2*, -3 and -4 (complex), -6, -9, -10, -12, -13, -14, -16, -17, -18 and -19 encoding various Smad proteins (transcription factors), and dbl-1 (see above), daf-4, kin-29 (encoding a kinase involved in regulating the expression of chemosensory receptors and entry into the dauer pathway), rnt-1, crm-1, lgg-1 and che-2 (encoding a protein with G-protein beta-like WD-40 repeats that affects chemotaxis, longevity and dauer formation; expressed in male tail rays, and some head and sensory neurons [39–44]), most of which are involved (up- or downstream) in TGF-beta signaling. Given the relative conservation of C. elegans and T. canis LON-1 (overall aa similarity: 70 %), this protein might indirectly regulate body size, morphogenesis, germline quality and/or reproductive aging. In C. elegans, lon-1 is expressed in the head region, intestine and hypodermis (cf. [12]; WBGene00003055); in T. canis, Tc-lon-1 is transcribed at moderate levels in adult worms of both sexes, at low levels in the L3 stage and at very low levels in all other stages (Additional file 1: Table S4).

Both Tc-CAP-27 and Tc-CAP-28 are homologs of the C. elegans glioma pathogenesis-related protein 2 (GLIPR2) orthologs F57B7.2 and F09B9.5 (50-54 % aa identity), but these proteins share limited sequence similarity to human GLIPR2 (accession number: Q5VZR0; 16-30 % aa identity) and are slightly more similar to the human Golgi-associated plant pathogenesis-related protein 1 (GAPR1; PDB accession number: 4AIW; 18-34 % aa identity) of the same pathogenesis-related 1 (PR-1) family. Some of the genes encoding these proteins are well characterized in mammals [8, 38, 45]. In contrast to the moderate to low levels of transcription of F57B7.2 in the larval and young adult stages of C. elegans (cf. FPKM expression data from selected modENCODE libraries [46, 47]), Tc-cap-27 and Tc-cap-28 have variable transcription within and among the stages and tissues of T. canis studied here (Fig. 1), although Tc-cap-27 is mostly transcribed at higher levels than Tc-cap-28. For Tc-cap-27, no significant upregulation was detected for any of the stages with respect to other stages/tissues, but Tc-cap-28 was significantly upregulated in the L3

stage with respect to all tissues of female *T. canis* (FC-range: 5.2 to 6.2).

Originally, the GLIPR2 gene was identified and localized to human chromosome 9p12-p13, with orthologous genes localized in syntenic regions in other species (cf. [8]). The GLIPR2 subfamily possesses several unique features that suggest that this subfamily might contain the most "primitive" mammalian CAP proteins. GLIPR2 proteins do not contain a predicted signal sequence, which is consistent with an intracellular localization to the Golgi membrane [48]. Interestingly, the conserved PY dipeptide sequence present in most other CAP superfamily members [8] is absent, as is the Hinge-like sequence present in other mammalian CAP proteins. In this latter respect, GLIPR2 is more like a non-mammalian CAP. This interpretation is supported by phylogenetic analysis [8], showing the GLIPR2 clade to be the "earliest diverged" CAP subfamily in mammals, using the protein PRY1 from the yeast Saccharomyces cerevisiae as the outgroup (cf. [49]). Analyses have shown GLIPR2 orthologs to be present in a broad range of species including S. cerevisiae, Drosophila melanogaster, C. elegans, other nematodes, insects, ascidians, bony fish, amphibians, birds and mammals (e.g. [8, 20]). In mammals, experimental studies have shown that GLIPR2 is mainly expressed in/localized to tissues with immunological functions, including developing and circulating leukocytes within the spleen. High levels of GLIPR2 have also been localized in the embryo, kidney, lung, pancreas, placenta and uterus, and relatively low levels in the brain, skeletal muscles and testes [48, 50]. Taken together, these findings suggest that Tc-cap-27 and Tc-cap-28 might play an immunological role in the nematode and/or an immune modulatory role in the host animal.

The nine CAP proteins encoded by Tc-cap-7 to Tccap-15 are homologs of Ce-VAP-1 (Additional file 1: Table S2), which is a secreted protein similar to some venom allergen-like (VAL) proteins found in a number of invertebrates, including parasitic nematodes, and a human homolog of cysteine-rich secretory protein 3 isoform 2 precursor [51]. Transcription of the genes Tc*cap-13* and *Tc-cap-14* was significantly upregulated in the adult male whole-worm (FC-range: 3.8 to 16.6) and male gut (FC-range: 3.7 to 12.7) with respect to all other stages/tissues of *T. canis* studied (Additional file 1: Table S3). Although the transcription levels of Tc-cap-10 and -11 were very low in the L3 and female stages, transcription in male stages was slightly higher than in females, whereas Tc-cap-7, -8 and -12 were transcribed at moderate levels in L3s, and Tc-cap-9 and Tc-cap-15 consistently at very low levels across all stages and tissues investigated (Additional file 1: Table S4). The set of nine VAP-1 homologs in T. canis suggests diversified functional roles of these molecules in this parasite, which appear to be reflected in considerable differences in transcription profiles. Based on functional data for *C. elegans*, we propose that at least some of the *Tc-cap* genes encoding VAP-1 homologs are involved in motility and expressed in chemosensory organs. In *C. elegans*, a *Ce-vap-1* reporter fusion is expressed specifically within sheath cells of such organs (i.e. amphids), and knockdown of *Ce-vap-1* results in an uncoordinated (Unc) phenotype (WBPhenotype:0000643; [52]). In addition, we propose that male-enriched expression of *Tc-cap-13* and *Tc-cap-14* in whole adult worms and gut tissues might relate to developmental and/or digestive processes, possibly acting in pathways associated with *Tc-cap-25* and *Tc-cap-26* (encoding "orphan" CAP proteins), which showed very similar transcription profiles.

Based on genetic network analysis, Ce-vap-1 is inferred to interact with 17 Ce-scl genes (Ce-scl-1, -2, -3, -5, -6, -7, -8, -9, -10, -11, -12, -13, -14, -15, -17, -18 and -20) in C. elegans [20, 53], and Ce-scl-1 interacts with age-1, daf-2 and daf-16 [21, 23, 54-58]. Although Ce-vap-1 and Ce-vap-2 do not interact directly with each other, all 17 Ce-scl genes are predicted to interact with Ce-vap-2 (cf. [20]) that interacts with egl-9 [59]. Independent interactions of Ce-vap-1 and Ce-vap-2 with Ce-scl-1, and their association with parts of the insulin-like signaling pathway via Ce-age-1, Ce-daf-2 and Ce-daf-16 [20, 60] suggest an integrated but relatively complex role of these molecules/pathways in regulating nematode development and growth in T. canis and/or roles in the parasite-host relationship, which is supported by substantial differences in the number of *T. canis vap-1* (n = 9) and *scl* (n = 5) genes, compared with *C. elegans* (n = 1 and 26, respectively; [61]).

The Ce-vap-2 homologs of T. canis, including Tc-cap-16 to Tc-cap-21, represent homologs of human cysteinerich secretory proteins involved in receptor-mediated endocytosis [62]. These receptors specifically recognize and bind extracellular macromolecules (ligands); the area of the plasma membrane with the receptor-ligand complex then undergoes endocytosis and forms a transport vesicle containing the receptor-ligand complex. Although involved in the specific uptake of particular substances (e.g. iron or low density lipoproteins) required by cells, endocytosis has also been linked to the transduction of signals from the periphery of cells to their nuclei [63]. The upregulation of Tc-cap-16, -19, -20 and -21 in the L3 stage might indicate a substantial requirement for nutrient uptake as the L3 stage enters the host animal and undergoes migration. Interestingly, Tccap-18 is upregulated both in this stage and in the female adult reproductive tract (compared with all male tissues; FC-range: 3.5 to 6.2), which further supports the notion of a role for vap-2 orthologs in nutrient uptake by both L3s and by progeny within gravid females. This proposal is relatively consistent with the finding that

Ce-vap-2 genetically interacts with Ce-egl-9 (cf. [20]), a gene involved in yolk uptake by oocytes in C. elegans [64]. Functional genomic studies have shown that Cevap-2 suppresses egl-9, leading to an egg-laying defect (Egl phenotype; WormBase entry WBRNAi00071630; [62]). The protein Ce-EGL-9 is known to function in a conserved hypoxia-sensing pathway to negatively regulate Ce-HIF-1 (hypoxia-inducible factor) by hydroxylating prolyl Ce-HIF-1 residues [65]; EGL-9 activity is negatively regulated by its physical association with Ce-CYSL-1, a protein with similarity to cysteine synthases that transduces signals linked to environmental O₂ levels via hydrogen sulfide (H₂S) signaling [66]. Ce-EGL-9 belongs to a protein superfamily (representing leprecan and AlkB) implicated in the hydroxylation of proteins and oxidative detoxification of alkylated bases [67]. This protein is expressed in hypodermis, musculature and neurons, and needed for muscle function for egg laying [68]. Based on this information, we suggest that *Tc-vap-2* is involved in aspects of reproduction, including egg-yolk uptake and egg laying, in intimate association with a Ce-egl-9 ortholog and/or other complementary genes.

Conclusions

Surprisingly, the CAP protein sequences of T. canis predicted in the recent T. canis genome project [7] that have been curated and described in detail in the present work were not described in previous molecular studies (reviewed by [69–71]). The number of transcripts encoding various CAP proteins in T. canis compares with those inferred from transcriptomic and genomic sequence data sets for A. suum and Trichuris suis [72, 73], but is substantially less than those for A. caninum, A. ceylanicum, N. americanus and H. contortus [20, 24, 29, 30, 74-76]. The reasons for this apparent difference are unclear, but might relate to variation in developmental and reproductive biology as well as varying modes of host invasion and immune modulation or evasion among nematode species (cf. [9, 10]). Moreover, we found no evidence of neutrophil inhibitory factor (NIF) homologs amongst the predicted CAP proteins of T. canis. NIFs can be relatively abundant in ES products from some parasitic worms, such as hookworms [77]. For instance, SCP-1, a NIF homolog of A. caninum, has been reported to bind host integrin CR3 (CD11b/ CD18), leading to the inhibition of neutrophil function [78, 79]. The absence of a NIF homolog in T. canis suggests that other molecules might assume NIF-like roles in this parasite.

The generation of improved genomic and transcriptomic assemblies for parasitic helminths (cf. [80, 81]) and the availability of expanded datasets for *T. canis* [7], as well as the characterization of the gene silencing machinery in *T. canis* and *A. suum* [7, 72] and advances in functional genomics of these ascaridoids [82, 83] will

likely create opportunities for investigations of CAP protein-encoding genes and their products in different developmental stages of these nematodes. The development of a medium- to high-throughput gene-silencing screening system for *T. canis*, based on initial evidence of effective knockdown of some genes [82], could accelerate our understanding of essential orphan CAP proteins and their genes (e.g. *Tc-cap-25* and *Tc-cap-26*), and should elucidate their involvement in biological and developmental pathways in this important parasite, and in parasite-host interactions. In conclusion, this study provides a basis to guide future studies of CAP proteins of *T. canis* and related ascaridoids, and might assist in finding new interventions against diseases caused by these nematodes.

Additional file

Additional file 1: Table S1. CAP protein-encoding genes of Toxocara canis. For each predicted protein, the genomic scaffolds to which they map, including the length of the inferred amino acid sequences and InterProScan matches (including domains, motifs and signal peptides), are shown. Table S2. Pairwise comparisons of 28 predicted Toxocara canis CAP protein sequences with homologs in Caenorhabditis elegans and parasitic nematodes including Ascaris lumbricoides (AI), Anisakis simplex (As), Ascaris suum (Asu), Brugia malayi (Bm), Caenorhabditis japonica (Cj), Caenorhabditis sinica (Cs), Dirofilaria immitis (Di), Dracunculus medinensis (Dm), Elaeophora elaphi (Ee), Enterobius vermicularis (Ev), Loa loa (LI), Litomosoides sigmodontis (Ls), Oesophagostomum dentatum (Od), Onchocerca ochenai (Oo) Onchocerca volvulus (Ov) Parascaris eauorum (Pe), Panagrellus redivivus (Pr), Steinernema carpocapsae (Sc), Steinernema feltiae (Sf), Steinernema glaseri (Sg), Syphacia muris (Sm), Steinernema monticolum (Smo), Steinernema scapterisci (Ss) and Thelazia callipaeda (Tc), with gene identification (ID) and identity (%) indicated in brackets. Table S3. Toxocara canis cap genes (Tc-cap) differentially transcribed (upregulated; log₂-fold change: > 2) in individual life stages/ sexes/tissues (pairwise comparisons). Table S4. Transcription of Toxocara canis cap genes (Tc-cap) in different life stages/sexes/tissues. Values are transcripts per million (TPM). (XLSX 44 kb)

Abbreviations

BLAST, basic local alignment search tool; CAP, cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1; CRISP, cysteine-rich secretory protein; dsRNAi, double-stranded RNA interference; ES, excretory/secretory; FC, log₂-fold change; FPKM, fragments per kilobase million; GAPR1, Golgi-associated plant pathogenesis-related protein 1; GLIPR2, glioma pathogenesis-related protein 2; SCP, sperm-coating protein; TPM, transcripts per million; VAL/VAP, venom allergen-like (protein)

Acknowledgements

Support from the Victorian Life Sciences Computation Initiative, Australia (VLSCI; grant no. VR0007) on its Peak Computing Facility at the University of Melbourne, Australia, an initiative of the Victorian Government, Australia, is gratefully acknowledged (RBG), as is other support from the Medicines for Malaria Venture (MMV), YourGene Bioscience (Taiwan) and Melbourne Water (RBG). We acknowledge the contributions of all of the staff members at WormBase (www.wormbase.org).

Funding

The present study was funded by the Australian Research Council (ARC), the National Health and Medical Research Council of Australia (NHMRC) and the Wellcome Trust (Pathfinder) (RBG). NDY is a NHMRC Career Development Fellow (CDF). AJS is a recipient of a Melbourne International Research Scholarship (MIRS) and a Melbourne International Fee Remission Scholarship (MIFRS) from The University of Melbourne. Funding bodies had no role in the design of the study or collection, analysis or interpretation of data, or in writing the manuscript.

Availability of data and material

The datasets supporting the conclusions of this article are included within the article and the additional file.

Authors' contributions

Conceived and designed the study and supervised the project: AJS and RBG. Undertook the study and data analysis: AJS and RSH. Contributed to analysis using various tools: NDY and PKK. Wrote the paper: AJS and RBG.

Contributed to the interpretation of findings, drafting of the manuscript and funding: NDY, AJ, AH, PWS and RBG. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia. ²Structural Chemistry Program, Eskitis Institute, Griffith University, Brisbane, Australia. ³HHMI and Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA.

Received: 2 May 2016 Accepted: 13 June 2016 Published online: 24 June 2016

References

- Macpherson CN. The epidemiology and public health importance of toxocariasis: a zoonosis of global importance. Int J Parasitol. 2013;43:999–1008.
- Rubinsky-Elefant G, Hirata CE, Yamamoto JH, Ferreira MU. Human toxocariasis: diagnosis, worldwide seroprevalences and clinical expression of the systemic and ocular forms. Ann Trop Med Parasitol. 2010;104:3–23.
- 3. Overgaauw PA, van Knapen F. Veterinary and public health aspects of *Toxocara* spp. Vet Parasitol. 2013;193:398–403.
- Pinelli E, Brandes S, Dormans J, Gremmer E, van Loveren H. Infection with the roundworm *Toxocara canis* leads to exacerbation of experimental allergic airway inflammation. Clin Exp Allergy. 2008;38:649–58.
- Akao N. Critical assessment of existing and novel model systems of toxocariasis. In: Holland CV, Smith HV, editors. *Toxocara*: the enigmatic parasite. Wallingford: CABI Publishing; 2006. p. 74–85.
- Schnieder T, Laabs EM, Welz C. Larval development of *Toxocara canis* in dogs. Vet Parasitol. 2011;175:193–206.
- Zhu XQ, Korhonen PK, Cai H, Young ND, Nejsum P, von Samson-Himmelstjerna G, et al. Genetic blueprint of the zoonotic pathogen *Toxocara canis*. Nat Commun. 2015;6:6145.
- Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins - roles in reproduction, cancer, and immune defense. Endocr Rev. 2008;29:865–97.
- Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, Nisbet AJ, et al. A portrait of the "SCP/TAPS" proteins of eukaryotes - developing a framework for fundamental research and biotechnological outcomes. Biotechnol Adv. 2009;27:376–88.
- 10. Cantacessi C, Gasser RB. SCP/TAPS proteins in helminths where to from now? Mol Cell Probes. 2012;26:54–9.
- Hewitson JP, Filbey KJ, Esser-von Bieren J, Camberis M, Schwartz C, Murray J, et al. Concerted activity of IgG1 antibodies and IL-4/IL-25-dependent effector cells trap helminth larvae in the tissues following vaccination with defined secreted antigens, providing sterile immunity to challenge infection. PLoS Pathog. 2015;11:e1004676.
- WormBase: Nematode Information Resource. http://www.wormbase.org/ (2016). Accessed 2 May 2016.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–40.

- 14. Kent WJ. BLAT the BLAST-like alignment tool. Genome Res. 2002;12:656-64.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.
- 16. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7.
- 17. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
- R: The R Project for Statistical Computing. http://www.R-project.org/ (2016). Accessed 2 May 2016.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
- Mohandas N, Young ND, Jabbar A, Korhonen PK, Koehler AV, Amani P, et al. The barber's pole worm CAP protein superfamily - a basis for fundamental discovery and biotechnology advances. Biotechnol Adv. 2015;33:1744–54.
- 21. Zhong W, Sternberg PW. Genome-wide prediction of *C. elegans* genetic interactions. Science. 2006;311:1481–4.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
- Ookuma S, Fukuda M, Nishida E. Identification of a DAF-16 transcriptional target gene, scl-1, that regulates longevity and stress resistance in Caenorhabditis elegans. Curr Biol. 2003;13:427–31.
- Datu BJ, Gasser RB, Nagaraj SH, Ong EK, O'Donoghue P, McInnes R, et al. Transcriptional changes in the hookworm, *Ancylostoma caninum*, during the transition from a free-living to a parasitic larva. PLoS Negl Trop Dis. 2008;2:e130.
- Osman A, Wang CK, Winter A, Loukas A, Tribolet L, Gasser RB, et al. Hookworm SCP/TAPS protein structure - A key to understanding host-parasite interactions and developing new interventions. Biotechnol Adv. 2012;30:652–7.
- Cantacessi C, Mitreva M, Jex AR, Young ND, Campbell BE, Hall RS, et al. Massively parallel sequencing and analysis of the *Necator americanus* transcriptome. PLoS Negl Trop Dis. 2010;4:e684.
- Cantacessi C, Loukas A, Campbell BE, Mulvenna J, Ong EK, Zhong W, et al. Exploring transcriptional conservation between *Ancylostoma caninum* and *Haemonchus contortus* by oligonucleotide microarray and bioinformatic analyses. Mol Cell Probes. 2009;23:1–9.
- Zhan B, Liu Y, Badamchian M, Williamson A, Feng J, Loukas A, et al. Molecular characterisation of the *Ancylostoma*-secreted protein family from the adult stage of *Ancylostoma caninum*. Int J Parasitol. 2003;33:897–907.
- Schwarz EM, Hu Y, Antoshechkin I, Miller MM, Sternberg PW, Aroian RV. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. Nat Genet. 2015;47:416–22.
- Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, et al. Genome of the human hookworm *Necator americanus*. Nat Genet. 2014; 46:261–9.
- Maduzia LL, Gumienny TL, Zimmerman CM, Wang H, Shetgiri P, Krishna S, et al. *Ion-1* regulates *Caenorhabditis elegans* body size downstream of the *dbl-1* TGF beta signaling pathway. Dev Biol. 2002;246:418–28.
- Morita K, Flemming AJ, Sugihara Y, Mochii M, Suzuki Y, Yoshida S, et al. A Caenorhabditis elegans TGF-beta, DBL-1, controls the expression of LON-1, a PR-related protein, that regulates polyploidization and body length. EMBO J. 2002;21:1063–73.
- 33. Tuck S. The control of cell growth and body size in *Caenorhabditis elegans*. Exp Cell Res. 2014;321:71–6.
- 34. Delannoy-Normand A, Cortet J, Cabaret J, Neveu C. A suite of genes expressed during transition to parasitic lifestyle in the trichostrongylid nematode *Haemonchus contortus* encode potentially secreted proteins conserved in *Teladorsagia circumcincta*. Vet Parasitol. 2010;174:106–14.
- Kasahara M, Gutknecht J, Brew K, Spurr N, Goodfellow PN. Cloning and mapping of a testis-specific gene with sequence similarity to a sperm-coating glycoprotein gene. Genomics. 1989;5:527–34.
- Kjeldsen L, Cowland JB, Johnsen AH, Borregaard N. SGP28, a novel matrix glycoprotein in specific granules of human neutrophils with similarity to a human testis-specific gene product and a rodent sperm-coating glycoprotein. FEBS Lett. 1996;380:246–50.
- Lu G, Villalba M, Coscia MR, Hoffman DR, King TP. Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. J Immunol. 1993;150:2823–30.
- Murphy EV, Zhang Y, Zhu W, Biggs J. The human glioma pathogenesisrelated protein is structurally related to plant pathogenesis-related proteins and its gene is expressed specifically in brain tumors. Gene. 1995;159:131–5.

- 39. Bargmann Cl, Hartwieg E, Horvitz HR. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. Cell. 1993;74:515–27.
- Fujiwara M, Ishihara T, Katsura I. A novel WD40 protein, CHE-2, acts cell-autonomously in the formation of *C. elegans* sensory cilia. Development. 1999;126:4839–48.
- Pujol N, Link EM, Liu LX, Kurz CL, Alloing G, Tan MW, et al. A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. Curr Biol. 2001;11:809–21.
- Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. Mol Cell. 2000;5: 411–21.
- Tonkin LA, Saccomanno L, Morse DP, Brodigan T, Krause M, Bass BL. RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. EMBO J. 2002;21:6025–35.
- 44. Vowels JJ, Thomas JH. Genetic analysis of chemosensory control of dauer formation in *Caenorhabditis elegans*. Genetics. 1992;130:105–23.
- Serrano RL, Kuhn A, Hendricks A, Helms JB, Sinning I, Groves MR. Structural analysis of the human Golgi-associated plant pathogenesis related protein GAPR-1 implicates dimerization as a regulatory mechanism. J Mol Biol. 2004; 339:173–83.
- F57B7.2 (gene) WormBase: Nematode Information Resource. http://www. wormbase.org/species/c_elegans/gene/WBGene00010192#01-9g-3 (2016). Accessed 2 May 2016.
- Reinke V, Gil IS, Ward S, Kazmer K. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. Development. 2004; 131:311–23.
- Eberle HB, Serrano RL, Fullekrug J, Schlosser A, Lehmann WD, Lottspeich F, et al. Identification and characterization of a novel human plant pathogenesis-related protein that localizes to lipid-enriched microdomains in the Golgi complex. J Cell Sci. 2002;115:827–38.
- 49. Miosga T, Schaaff-Gerstenschlager I, Chalwatzis N, Baur A, Boles E, Fournier C, et al. Sequence analysis of a 33.1 kb fragment from the left arm of *Saccharomyces cerevisiae* chromosome X, including putative proteins with leucine zippers, a fungal Zn(II)2-Cys6 binuclear cluster domain and a putative alpha 2-SCB-alpha 2 binding site. Yeast. 1995;11:681–9.
- Eisenberg I, Barash M, Kahan T, Mitrani-Rosenbaum S. Cloning and characterization of a human novel gene C9orf19 encoding a conserved putative protein with an SCP-like extracellular protein domain. Gene. 2002;293:141–8.
- 51. Shaye DD, Greenwald I. OrthoList: a compendium of *C. elegans* genes with human orthologs. PLoS One. 2011;6:e20085.
- Cronin CJ, Mendel JE, Mukhtar S, Kim YM, Stirbl RC, Bruck J, et al. An automated system for measuring parameters of nematode sinusoidal movement. BMC Genet. 2005;6:5.
- Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM. A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. Nat Genet. 2008;40:181–8.
- Harada H, Kurauchi M, Hayashi R, Eki T. Shortened lifespan of nematode Caenorhabditis elegans after prolonged exposure to heavy metals and detergents. Ecotoxicol Environ Saf. 2007;66:378–83.
- 55. Larsen PL. Direct and indirect transcriptional targets of DAF-16. Sci Aging Knowl Environ. 2003;2003:PE9.
- Libina N, Berman JR, Kenyon C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. Cell. 2003;115:489–502.
- 57. Liu T, Zimmerman KK, Patterson GI. Regulation of signaling genes by TGFbeta during entry into dauer diapause in *C. elegans*. BMC Dev Biol. 2004;4:11.
- 58. Patterson Gl. Aging: new targets, new functions. Curr Biol. 2003;13:R279-81.
- Gort EH, van Haaften G, Verlaan I, Groot AJ, Plasterk RH, Shvarts A, et al. The TWIST1 oncogene is a direct target of hypoxia-inducible factor-2alpha. Oncogene. 2008;27:1501–10.
- Mohandas N, Hu M, Stroehlein AJ, Young ND, Sternberg PW, Lok JB, et al. Reconstruction of the insulin-like signalling pathway of *Haemonchus contortus*. Parasit Vectors. 2016;9:64.
- 61. Gene class WormBase: Nematode Information Resource. http://www. wormbase.org/resources/gene_class/ (2016). Accessed 2 May 2016.
- Balklava Z, Pant S, Fares H, Grant BD. Genome-wide analysis identifies a general requirement for polarity proteins in endocytic traffic. Nat Cell Biol. 2007;9:1066–73.
- 63. Ceresa BP, Schmid SL. Regulation of signal transduction by endocytosis. Curr Opin Cell Biol. 2000;12:204–10.
- Grant B, Hirsh D. Receptor-mediated endocytosis in the *Caenorhabditis* elegans oocyte. Mol Biol Cell. 1999;10:4311–26.

- Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, et al. *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell. 2001;107:43–54.
- Ma DK, Vozdek R, Bhatla N, Horvitz HR. CYSL-1 interacts with the O₂-sensing hydroxylase EGL-9 to promote H₂S-modulated hypoxia-induced behavioral plasticity in *C. elegans*. Neuron. 2012;73:925–40.
- 67. Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol. 2001;2:0007.1–8.
- Trent C, Tsuing N, Horvitz HR. Egg-laying defective mutants of the nematode *Caenorhabditis elegans*. Genetics. 1983;104:619–47.
- Gasser RB. A perfect time to harness advanced molecular technologies to explore the fundamental biology of *Toxocara* species. Vet Parasitol. 2013; 193:353–64.
- Maizels RM. *Toxocara canis*: molecular basis of immune recognition and evasion. Vet Parasitol. 2013;193:365–74.
- Gasser RB, Korhonen PK, Zhu XQ, Young ND. Harnessing the *Toxocara* genome to underpin toxocariasis research and new interventions. Adv Parasitol. 2016;91:87–110.
- 72. Jex AR, Liu S, Li B, Young ND, Hall RS, Li Y, et al. *Ascaris suum* draft genome. Nature. 2011;479:529–33.
- Jex AR, Nejsum P, Schwarz EM, Hu L, Young ND, Hall RS, et al. Genome and transcriptome of the porcine whipworm *Trichuris suis*. Nat Genet. 2014;46:701–6.
- Hawdon JM, Jones BF, Hoffman DR, Hotez PJ. Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. J Biol Chem. 1996;271:6672–8.
- Hawdon JM, Narasimhan S, Hotez PJ. Ancylostoma secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from Ancylostoma caninum. Mol Biochem Parasitol. 1999; 99:149–65.
- Schwarz EM, Korhonen PK, Campbell BE, Young ND, Jex AR, Jabbar A, et al. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. Genome Biol. 2013;14:R89.
- Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Mol Biochem Parasitol. 2009;167:1–11.
- Moyle M, Foster DL, McGrath DE, Brown SM, Laroche Y, De Meutter J, et al. A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. J Biol Chem. 1994;269:10008–15.
- Rieu P, Sugimori T, Griffith DL, Arnaout MA. Solvent-accessible residues on the metal ion-dependent adhesion site face of integrin CR3 mediate its binding to the neutrophil inhibitory factor. J Biol Chem. 1996;271:15858–61.
- Korhonen PK, Young ND, Gasser RB. Making sense of genomes of parasitic worms: Tackling bioinformatic challenges. Biotechnol Adv. 2016; (In press).
- Tsai IJ, Otto TD, Berriman M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 2010;11:R41.
- Ma GX, Zhou RQ, Song ZH, Zhu HH, Zhou ZY, Zeng YQ. Molecular mechanism of serine/threonine protein phosphatase 1 (PP1cα-PP1r7) in spermatogenesis of *Toxocara canis*. Acta Trop. 2015;149:148–54.
- McCoy CJ, Warnock ND, Atkinson LE, Atcheson E, Martin RJ, Robertson AP, et al. RNA interference in adult *Ascaris suum* - an opportunity for the development of a functional genomics platform that supports organism-, tissue- and cell-based biology in a nematode parasite. Int J Parasitol. 2015;45:673–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

