

# A directed approach for the identification of transcripts harbouring the spliced leader sequence and the effect of trans-splicing knockdown in *Schistosoma mansoni*

Marina de Moraes Mourão<sup>1</sup>, Mainá Bitar<sup>2</sup>, Francisco Pereira Lobo<sup>3</sup>, Ana Paula Peconick<sup>4</sup>, Priscila Grynberg<sup>2</sup>, Francisco Prosdocimi<sup>5</sup>, Michael Waisberg<sup>6</sup>, Gustavo Coutinho Cerqueira<sup>7</sup>, Andréa Mara Macedo<sup>2</sup>, Carlos Renato Machado<sup>2</sup>, Timothy Yoshino<sup>8</sup>, Glória Regina Franco<sup>2/+</sup>

<sup>1</sup>Grupo de Genômica e Biologia Computacional, Centro de Pesquisas René Rachou-Fiocruz, Belo Horizonte, MG, Brasil <sup>2</sup>Laboratório de Genética Bioquímica, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil <sup>3</sup>Laboratório de Bioinformática Aplicada, Embrapa Informática Aplicada, Campinas, SP, Brasil <sup>4</sup>Setor de Medicina Veterinária Preventiva, Universidade Federal de Lavras, Lavras, MG, Brasil <sup>5</sup>Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil <sup>6</sup>Department of Pathology, University of Virginia, Charlottesville, VA, USA <sup>7</sup>Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA <sup>8</sup>Department of Pathobiological Sciences, University of Wisconsin, Madison, WI, USA

*Schistosomiasis is a major neglected tropical disease caused by trematodes from the genus Schistosoma. Because schistosomes exhibit a complex life cycle and numerous mechanisms for regulating gene expression, it is believed that spliced leader (SL) trans-splicing could play an important role in the biology of these parasites. The purpose of this study was to investigate the function of trans-splicing in Schistosoma mansoni through analysis of genes that may be regulated by this mechanism and via silencing SL-containing transcripts through RNA interference. Here, we report our analysis of SL transcript-enriched cDNA libraries from different S. mansoni life stages. Our results show that the trans-splicing mechanism is apparently not associated with specific genes, subcellular localisations or life stages. In cross-species comparisons, even though the sets of genes that are subject to SL trans-splicing regulation appear to differ between organisms, several commonly shared orthologues were observed. Knockdown of trans-spliced transcripts in sporocysts resulted in a systemic reduction of the expression levels of all tested trans-spliced transcripts; however, the only phenotypic effect observed was diminished larval size. Further studies involving the findings from this work will provide new insights into the role of trans-splicing in the biology of S. mansoni and other organisms. All Expressed Sequence Tags generated in this study were submitted to dbEST as five different libraries. The accessions for each library and for the individual sequences are as follows: (i) adult worms of mixed sexes (LIBEST\_027999: JZ139310 - JZ139779), (ii) female adult worms (LIBEST\_028000: JZ139780 - JZ140379), (iii) male adult worms (LIBEST\_028001: JZ140380 - JZ141002), (iv) eggs (LIBEST\_028002: JZ141003 - JZ141497) and (v) schistosomula (LIBEST\_028003: JZ141498 - JZ141974).*

Key words: spliced leader - trans-splicing - RNA interference - *Schistosoma mansoni*

Schistosomiasis is an important neglected tropical disease caused by species of the parasitic flatworm *Schistosoma*. According to the World Health Organization (WHO 2012), the disease affects more than 230 million people yearly and the resulting morbidity compromises local economies and child development (Fenwick & Webster 2006). Transmission of schistosomiasis has been documented in 77 countries, with more than 90% of cases occurring on the African continent. *Schistosoma mansoni*, which is mainly found in Africa and Brazil, is a major cause of intestinal schistosomiasis. This parasite

exhibits a complex life cycle involving a snail intermediate host and a mammalian definitive host (Pessoa & Martins 1982).

Gene discovery in *S. mansoni* has taken advantage of extensive and well-annotated Expressed Sequence Tags (EST) databases (dbEST, SchistoDB and GeneDB) and, more recently, the Sequence Read Archive, containing next-generation sequencing reads covering the entire transcriptome of this species (Boguski et al. 1993, Zerlotini et al. 2009, Leinonen et al. 2011, Logan-Klumpler et al. 2012). Additionally, reverse genetic approaches have been widely explored in this parasite in order to provide perspectives on the identification of new targets for drug and vaccine development and to develop novel protocols for diagnosis (Skelly et al. 2003, Kalinna & Brindley 2007, Mann et al. 2008, Pearce & Freitas 2008, Mourão et al. 2009a, b, Yoshino et al. 2010).

Schistosomes possess numerous and complex transcriptional and post-transcriptional gene regulatory mechanisms allowing them to maintain their complex life cycle. Because of the prominence of the spliced leader (SL) sequence in a number of *S. mansoni* messen-

doi: 10.1590/0074-0276108062013006

Financial support: CNPq (480576/2010-6), FAPEMIG (CBB-APQ-01715/11), CAPES

MMM and MB contributed equally to this work.

+ Corresponding author: gfrancoufmg@gmail.com

Received 28 February 2013

Accepted 13 June 2013

ger RNA (mRNAs), it is presumed that SL trans-splicing represents an important form of post-transcriptional regulation and could be a potential target for impairing *S. mansoni* development (Davis et al. 1995). All organisms that exhibit trans-splicing display one or more SL-RNAs, which are products of tandemly repeated small intronless genes transcribed by DNA polymerase II (Hastings 2005). SL-RNAs are small non-coding RNAs of 40-140 nucleotides in length, carrying a donor splice site and a hyper-modified cap (Nilsen 1993). The donor splice site divides the SL-RNA into two segments: a 5' leader sequence and an intron-like sequence at the 3' end. Despite a lack of sequence similarity, SL-RNAs from different organisms exhibit an impressive similarity in secondary structure to small nuclear RNAs, which are components of the spliceosome and actively participate in all splicing mechanisms (Hastings 2005).

The function of trans-splicing is still poorly understood. Although the best-documented function of SL trans-splicing is in the generation of monocistronic transcripts from polycistronic operons (Blumenthal & Gleason 2003), trans-splicing has also been implicated in a variety of functions associated with RNA maturation, including (i) providing a 5' cap for RNAs transcribed by RNA polymerase I (Lee & Van der Ploeg 1997, Gunzl et al. 2003), (ii) enhancing translation through the addition of a hyper-modified 5' cap in immature mRNAs and (iii) removing potentially deleterious elements within the 5' UTR (e.g., sequences that could compromise mRNA translation) (Hastings 2005, Matsumoto et al. 2010).

Thus far, SL trans-splicing mechanisms have been identified in cnidarians, primitive chordates, nematodes, platyhelminthes and dinoflagellates (Krause & Hirsh 1987, Rajkovic et al. 1990, Brehm et al. 2000, 2002, Stover & Steele 2001, Vandenberghe et al. 2001, Zayas et al. 2005, Lidie & van Dolah 2007). In contrast, trans-splicing has never been described in plants, vertebrates or fungi, which raises many questions regarding the occurrence of SL trans-splicing in an evolutionary context and its role in post-transcriptional regulation in selected species. In *S. mansoni*, SL trans-splicing does not appear to be associated with any particular tissue, developmental phase or sex (Davis 1996). Moreover, there is no conclusive evidence associating SL trans-splicing with specific genes or gene families (Davis et al. 1995). The main goal of the present study was to identify genes or gene categories that could be targeted by trans-splicing in different schistosome life cycle stages and to provide a better understanding of the importance of the trans-splicing mechanism during *S. mansoni* development.

## MATERIALS AND METHODS

**Biological samples** - The *S. mansoni* life cycle was maintained at the René Rachou Research Centre (CPqRR), Oswaldo Cruz Foundation, and at Interdepartmental Group for Epidemiological Research, Department of Parasitology, Federal University of Minas Gerais (UFMG), Brazil. The LE strain of *S. mansoni* was maintained in the snail intermediate host *Biomphalaria glabrata* (Barreiro de Cima strain). Outbred Swiss Webster mice were housed conventionally in polypropylene

cages with stainless steel screen covers. All animals received laboratory mouse chow and water *ad libitum*. The experimental protocols described herein were reviewed and approved by the Ethical Review Committee for Animal Experimentation (CETEA) of UFMG (185/2006).

Adult worms were obtained via portal perfusion of mice that had been infected for five weeks as previously described (Smithers & Terry 1965). The worms were washed with cold saline solution, carefully separated based on sex with fine forceps under a microscope and immediately frozen at -80°C until further processing. Mechanically transformed schistosomula (7-day-cultured) (Basch 1981) were provided by CPqRR. *S. mansoni* eggs were recovered from the intestinal homogenates of 48-day-infected Swiss Webster mice. The collected tissues and eggs were filtered through a sieve to remove coarse debris and then allowed to settle. The resulting pellet was washed with 1.7% saline and frozen at -80°C for further processing.

**RNA isolation, reverse transcriptase (RT) and SL transcript enrichment** - RNA from male and female adult worms and cultured schistosomula was extracted using the RNagents kit (Promega, Madison, USA) following the manufacturer's protocol. Total RNA from *S. mansoni* eggs was extracted using the TRIzol Reagent (Life Technologies, Carlsbad, CA, USA) according to manufacturer's protocol. Direct isolation of poly(A)<sup>+</sup> mRNA from adult worms was performed using Dynabeads Oligo (dT)25 magnetic beads (Dyna, Life Technologies, Carlsbad, USA). Briefly, following the extraction step, beads containing bound mRNA were re-suspended in 20 µL of 1X RT buffer (Life Technologies, Carlsbad, USA) and used directly in RT-polymerase chain reaction (PCR) assays.

The cDNA synthesis was carried out from all samples, except for those from adult worms, using SuperScript II Reverse Transcriptase (Life Technologies, Carlsbad, USA), according to procedures outlined by the manufacturer. For adult worms, 1 µL of beads containing mRNA was added directly to the reaction. The oligo dT-anchored primer used for cDNA first-strand synthesis (5'CGGTATTTTCAGTCGGTGTTCACCT19V3' - V = A, G, C) was designed to contain a 5' tail that was later employed in a PCR assay to amplify trans-spliced transcripts. The strategy for enriching cDNA libraries in trans-spliced transcripts included a PCR step involving a 5' tailed oligo dT sequence (Brehm et al. 2000) and part of the previously described *S. mansoni* SL sequence (underlined), 5'AACCGTCACGGTTTTACTCTTGTA TTTGTTGCATG3' (Davis et al. 1995).

To prevent the amplification of spurious cDNA sequences, a step-down program was employed for PCR, consisting of one cycle of 5 min at 95°C for DNA denaturation, five cycles of 1 min at 95°C, 1 min at 60°C and 1.5 min at 72°C, five cycles of 1 min at 95°C, 1 min at 59°C and 1.5 min at 72°C, five cycles of 1 min at 95°C, 1 min at 58°C and 1.5 min at 72°C and 19 cycles of 1 min at 95°C, 1 min at 57°C and 1.5 min at 72°C.

**Size selection of cDNAs** - Fragment size selection was performed to prevent over-representation of small PCR

products. To accomplish this, two methodologies were applied: (i) PCR amplicon separation via electrophoresis in a 1% agarose gel and further isolation using the enzyme  $\beta$ -agarase according to Franco et al. (1995) and (ii) precipitation with 15% polyethyleneglycol 8000 (to obtain fragments  $\geq 400$  bp). The purified amplicons were cloned into the pGEM and pCR2.1 vectors using the T-easy System Vector kit (Promega, Madison, USA) and a TA cloning kit (Life Technologies, Carlsbad, USA), respectively, according to manufacturer's specifications.

The recombinant plasmids were used to transform competent *Escherichia coli* of the DH5 $\alpha$  strain. To select clones with large inserts and test library quality, recombinant bacterial clones were subjected to colony PCR using M13 forward and reverse primers. Amplification and insert size estimates were confirmed via electrophoresis in a 1% agarose gel. Selected clones were grown overnight in 2X YT medium (16 g of bacto tryptone, 10 g of bacto yeast extract, 5 g of NaCl per litre, pH 7.0) and recombinant plasmids were purified using a standard protocol. Template preparation and DNA sequencing reactions were conducted through DYEnamic ET dye terminator cycle sequencing (GE Healthcare), following the manufacturer's protocol with a MegaBACE 1000 capillary sequencer (GE Healthcare).

*S. mansoni* *in vitro* culture and small interfering (siRNA) treatment - All RNA interference (RNAi) experiments were performed using the Naval Medical Research Institute strain of *S. mansoni*. Eggs were obtained from the livers of mice that had been infected for seven-eight weeks. Transformation was carried out as previously described (Yoshino & Laursen 1995). Larvae were counted and distributed into either 48 or 96-well polystyrene tissue culture plates (Costar, Corning Incorporated, NY, USA) at concentrations of  $\sim 6,000$  [RT-quantitative PCR (qPCR)] or  $\sim 500$  miracidia/well (Mourão et al. 2009a). All RNAi experiments involved at least two technical replicates of the miracidial treatment and control groups and were repeated in three independent larval cultures. The parasites in culture were exposed on day zero to SL sequence siRNAs (treatment), an irrelevant siRNA (control I; decoy) or medium alone (control II). Cultured larvae were assessed for knockdown effects after seven days of treatment (Mourão et al. 2009a). The RNAi experiments involving mice were pre-approved by the Institutional Animal Care and Use Committee of the University of Wisconsin-Madison, where the experiments were conducted, under assurance A3368-01.

The double-stranded siRNA sequences of 21 nucleotides in length were designed using BLOCK-iT™ RNAi Designer tools, available from rnaidesigner.invitrogen.com/rnaexpress (Life Technologies, Carlsbad, CA, USA). The generated sequences were synthesised by Life Technologies using the Stealth™ proprietary modification. The decoy control was designed with a similar GC content and length as the target SL-siRNA.

*Phenotypic screening* - Cultured *S. mansoni* larvae were plated in 96-well culture plates (Costar) at a density of approximately 500 miracidia per well, which contained a 200 pM concentration of SL-siRNA (experiment

group) or decoy sequence siRNA (control I) diluted in 200  $\mu$ L of CBSS or medium lacking siRNA (control II). The cultures were maintained at 26°C for seven days, after which the sporocysts were monitored for the following phenotypes: failure or delay of transformation, loss of motility, tegumental lysis and granulation (lethality) and changes in larval growth. Parasite viability and morphological changes were monitored daily as previously described (Mourão et al. 2009a). Length measurements were performed in captured images using Metamorph software, version 7.0 (Meta Imaging series, Molecular Devices, Sunnyvale, CA, USA). Larval growth datasets for each experimental replicate were statistically analysed using the Mann-Whitney *U* test (Wilcoxon-Sum of Ranks test) at a significance level of  $p \leq 0.05$ . All treatments were performed in triplicate wells and were independently replicated three times in miracidia isolated from different batches of infected mouse livers.

*Effect of double-stranded RNA treatment on larval gene expression* - qPCR was used to determine steady-state transcript levels in specific ds-siRNA-treated sporocysts. In these experiments  $\sim 6,000$  miracidia were distributed into a 48-well plate (Costar) and treated with 200 nM siRNA diluted in CBSS (500  $\mu$ L/well). The cultures were maintained at 26°C for seven days prior to RNA extraction and isolation.

Following incubation, the sporocysts were extensively washed with CBSS to eliminate unabsorbed siRNAs and shed ciliary epidermal plates, followed by extraction in TRIzol Reagent (Life Technologies, Carlsbad, CA, USA) to isolate total RNA from cultured larvae (Mourão et al. 2009a). Isolated RNA was resuspended in diethylpyrocarbonate-treated water and subjected to DNase treatment using the Turbo DNA-Free kit (Ambion, Austin, TX, USA) to eliminate contaminating genomic DNA. The RNA samples were then quantified and their purity assessed in a Nanodrop Spectrometer, ND-1000 (Nanodrop Technologies, Inc, Wilmington, DE, USA).

*RT-qPCR analysis* - To evaluate transcript levels between the SL-siRNA-treated sporocysts and control treatment (decoy-siRNA), we performed a qPCR analysis. To this end, 0.5  $\mu$ g of total RNA derived from at least three different extractions was employed to synthesise cDNA using the Superscript III cDNA Synthesis kit (Invitrogen) following the manufacturer's protocol. The RT-qPCR assay mixtures consisted of 2.5  $\mu$ L of cDNA, 12.5  $\mu$ L of SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) and 10  $\mu$ L of 600 or 900 nM primers, determined after primer concentration optimisation following the Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines (Bustin et al. 2009), in 96-Well Optical Reaction Plates (ABI PRISM, Applied Biosystems, Foster City, CA, USA). The reactions were carried out using the AB7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). RT-qPCR validations were performed with the SL forward primer 5'-GTCACGGTTTTACTCTTGT-3' and a gene-specific reverse primer. Specific primers were designed for (i) five previously reported trans-spliced transcripts (Da-

vis et al. 1995), (ii) *S. mansoni*  $\alpha$ -tubulin, used as an endogenous normalisation control in all tested samples and (iii) the *S. mansoni* genes encoding three known non-trans-spliced transcripts, which were used as negative controls (Supplementary data). Each RT-qPCR run was conducted with two internal controls for assessing potential genomic DNA contamination (no RT) and the purity of the reagents used (no cDNA). For each specific set of primers, all individual treatments (including specificity controls) were run in three technical replicates. Each experiment was repeated three times as independent biological replicates and the results were analysed via the  $\Delta\Delta C_t$  method (Livak & Schmittgen 2001). Due to the nonparametric distribution of data, statistical analysis of the  $\Delta\Delta C_t$  values was carried out using the Mann-Whitney *U* test, with significance set at  $p \leq 0.05$ .

**Bioinformatics analysis** - The output files generated from the sequencing reactions (SL-enriched EST library) and from publicly available *S. mansoni* EST data (dbEST) (Boguski et al. 1993) were submitted to a bioinformatics pipeline including algorithms for base-calling, poly-A and vector decontamination, motif searching, similarity-based characterisation, gene ontology (GO) assignment and manually curated annotation and analysis (Fig. 1). All sequence retrieval (except when otherwise stated) was performed within the SchistoDB database. Information on exon content was also obtained from this database.

**Sequence processing** - Phred (Ewing et al. 1998) was employed as a base-calling procedure. Only sequences of at least 150 bases with quality scores higher than 10 were accepted. A multi-Fasta file was generated with all resulting sequences and was used throughout the subsequent analysis. Following Phred processing, the sequences from the SL-enriched EST library were subjected to analysis with the SeqClean program (available from: [compbio.dfc.harvard.edu/tgi/software/](http://compbio.dfc.harvard.edu/tgi/software/)), which consists of a Tiger-developed script capable of analysing EST data. SeqClean was employed to trim ESTs based on informational content, length and vector contamination. We used the vector database UniVec (Kitts et al. 2011) with two additional vectors (pGEM and pCR2.1) to scan our sequences for vector contamination with SeqClean. Poly-A tails and vector adaptors were also removed in this step. Only ESTs longer than 50 bp (following vector cleaning) were further evaluated. *S. mansoni* dbEST data were further submitted to the SeqClean cleaning pipeline and analysed together with the SL-enriched EST library data.

**SL detection and cleaning** - To identify SL-containing sequences in both the SL-enriched EST library and dbEST, we used BLAST algorithms. To classify ESTs as SL-containing ESTs, we defined an e-value cut-off of  $5 \times 10^{-5}$  and considered only sequences exhibiting at least 95% similarity and 25 contiguous nucleotides when aligned to the known SL sequence. After identifying the sequences as SL-containing ESTs, we used the local alignment software `cross_match` (Ewing et al. 1998) to remove the SL region from all SL-containing sequences.

**Sequence clustering and mapping** - Both the SL-enriched EST library and *S. mansoni* dbEST were analysed using the `cap3` program (Huang & Madan 1999) to generate sequence contigs. We employed `cap3` with the default parameters, except for the `-o` and `-p` flags, which were set to 40 and 95, respectively. Therefore, two ESTs were only grouped into a single contig when they shared at least 40 aligned nucleotides with a minimum sequence identity of 95%. After identifying the SL-containing sequences within the SL-enriched EST library and assembling the corresponding contigs, we mapped the sequences in the *S. mansoni* genome. For this purpose, we used BLAST to search for the previously assembled unique sequences in the parasite genome. The defined cut-off for the e-value was  $1 \times 10^{-10}$  and the minimum sequence similarity to be accepted was 90%.

**GO assignment and annotation** - We used SchistoDB (Zerlotini et al. 2009) to automatically annotate unique sequences from the SL-enriched EST library. GO assignment was performed for the set of sequences generated from the previous steps. GO categories were associated with each transcript using GoAnna (McCarthy et al. 2006). GO Slim terms (McCarthy et al. 2006) were also retrieved for all sequences to obtain a more general overview of the GO among the dataset. Manual functional annotation was conducted for all known proteins (excluded sequences were characterised as “expressed proteins” and “hypothetical proteins”). A literature search and homology analysis were carried out to assure correct functional annotation of the gene products. All proteins were clustered according to biological processes categories to provide a better understanding of the trans-splicing function in the cell.

## RESULTS

**Dataset generation and data analysis** - To determine whether the trans-splicing mechanism could target specific functional gene categories in *S. mansoni*, we first generated an SL-enriched EST dataset containing subsets

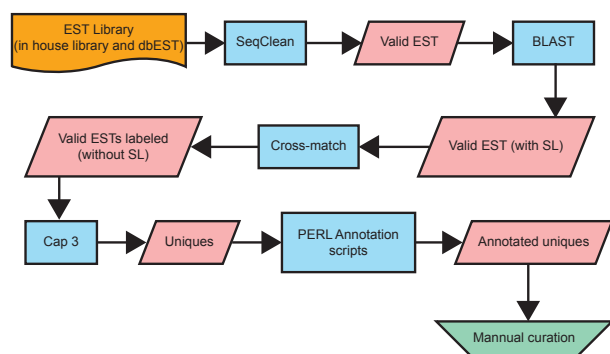


Fig. 1: fluxogram illustrating each step of processing and curation performed with sequences from the dbEST and the spliced leader (SL)-enriched Expressed Sequence Tags library. The fluxogram contains both the names of all programs and methods (blue shaded boxes) used in this study for sequence cleaning, validation, trimming and annotation and the results generated by those (pink shaded boxes).

from diverse parasite life cycle stages. The enriched EST dataset yielded a total of 3,087 sequences, 481 of which were from schistosomula, while 502 were from eggs, 600 from females, 623 from males and 881 from adult worms of mixed sex. After the removal of spurious sequences and vector contamination using SeqClean, 2,781 valid sequences were retained in the dataset. Of these sequences, 1,665 were classified as SL-containing sequences according to the previously described criteria. When the ratio of valid ESTs in the SL-enriched dataset containing the SL sequence vs. the total number of ESTs was compared, we found that 59.8% of the ESTs carried the SL sequence according to our very strict stringency parameters (as reported in the Materials and Methods section). This high percentage of SL sequences confirmed the enrichment of our SL dataset, as the SL/non-SL transcript ratio was only 0.1% when the entire dbEST dataset was analysed using the same protocol (Table I). Furthermore, recent data generated through RNA-Seq analysis (Protasio et al. 2012) indicate that ~11% (1,178 SL transcripts) of all *S. mansoni* transcripts are processed by trans-splicing. Taken together, these data suggest a high enrichment of SL-containing sequences in our dataset.

All 1,665 SL-containing sequences were subjected to BLAST searches in the SchistoDB database to map them into the *S. mansoni* genome and assign annotations and ID numbers according to SchistoDB data. Nine hundred eighty-nine sequences were mapped and further clustered by cap3, resulting in 258 unique sequences (102 singlets and 156 contigs). The remaining 676 sequences corresponded to redundant sequences or to sequences that did not map to the *S. mansoni* genome. The number of unique sequences per life stage is listed in Table I. These 258 unique sequences were mapped to a set of 162 different protein-coding sequences from the parasite genome. Sixty-four unique sequences were classified as “conserved hypothetical proteins” (7), “hypothetical proteins” (10) or “expressed proteins” (47) and were therefore not included in the functional analysis. A final set of 99 unique sequences was defined and was employed in all subsequent analyses (Fig. 2, Supplementary data).

**GO and manual functional annotation** - To identify the biological processes, subcellular localisations and molecular functions associated with the trans-spliced transcripts, GO terms were assigned to 78 of the 99 protein-coding unique sequences (the remaining 21 were not associated with any GO term by GOanna). Of the 78 proteins assessed, only 30 were assigned a particular subcellular localisation based on GOSlim results. Among these proteins, eight were localised to the membrane, eight were cytoplasmic (including 2 cytoskeleton-associated proteins), five were nuclear proteins, five were mitochondrial proteins and another five were classified merely as intracellular proteins, with no specific localisation. Based on analysis of the molecular functions assigned to the sequences through GO annotation, 24 proteins were classified as metal-binding (binding calcium, magnesium, iron, zinc and other metal ions), 12 as nucleotide-binding, 10 as nucleic acid-binding (1

TABLE I  
Number of sequences for each life-cycle stage contained in the spliced leader (SL)-enriched Expressed Sequence Tags (EST) library and derived from dbEST after each step of processing and curation

EST library (dbEST accession)	Detected ESTs (n)	Valid ESTs after sequence trimming (n)	Valid ESTs with SL sequence (n)	Valid ESTs with SL vs. ESTs n (%)	Valid ESTs annotated and mapped in the genome (n)	Valid annotated ESTs vs. valid ESTs n (%)
Egg (028002)	502	499	263	0.527 (53)	153	0.581 (58)
Schistosomula (028003)	481	481	258	0.536 (54)	157	0.608 (61)
Female worms (028000)	600	600	407	0.678 (69)	244	0.599 (60)
Male worms (028001)	623	623	356	0.571 (57)	256	0.719 (72)
Adult worms (027999)	881	578	381	0.659 (66)	179	0.469 (47)
All	3,087	2,781	1,665	0.598 (60)	989	0.594 (59)
dbEST	205,892	201,694	276	0.001 (0.1)	242	0.876 (88)

specifically interacting with RNA and 4 with DNA) and five as adenosine triphosphate (ATP)-binding proteins. Additionally, four proteins were classified as glycolytic enzymes. In the biological processes category, 21 proteins were identified as functioning in metabolic processes, eight of which were associated with biosynthesis, while five proteins were classified as being related to redox mechanisms. The functions of the remainder of the proteins were not specified.

Although several other trends were observed based on GO annotation, there were no clear biases found within the analysed trans-spliced protein dataset regarding cellular location, biological processes or molecular function. This result is in agreement with the current opinion that the trans-splicing mechanism is not associated with any specific gene category or protein feature.

In addition to the GO annotation, manual annotation of all 99 protein-coding unique sequences was performed. This was an important step, based on which additional information about protein functions was retrieved from the literature. The entire set of annotated proteins was subdivided into 19 classes of biological processes: development, cell cycle regulation and apoptosis, replication and repair, chromatin modification, transcription and post-transcriptional regulation, translation, protein folding, protein processing, modification and degradation, signal transduction, stress responses, cytoskeleton organisation, carbohydrate metabolism, lipid metabolism, energy homeostasis, cofactor metabolism, amino acid catabolism, transport and membrane turnover and miscellanea (Supplementary data).

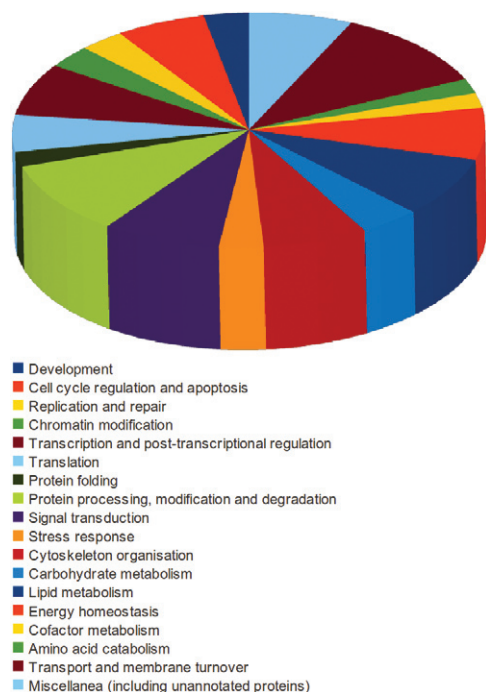


Fig. 2: pie chart illustrating the distribution of annotated proteins in different classes after manual curation and classification.

*Protein length and exon composition of trans-spliced transcripts* - To verify whether the trans-splicing mechanism might be associated with genes containing small exons, as suggested by Davis et al. (1995), we calculated the length of all of the protein-coding sequences obtained from SchistoDB and of all 99 annotated protein-coding genes in our dataset. Based on this survey, we found that the proteins derived from trans-spliced transcripts averaged ~400 amino acids in length. In comparison, the average length of all *S. mansoni* proteins obtained from SchistoDB was only slightly greater (~450 amino acids). We also estimated the number of exons per gene and the average exon length for all protein-coding sequences from the parasite found in both SchistoDB and our dataset. Again, the analysis showed a conserved exon composition in the two sets of proteins, with an average number of six exons per protein and an average length of ~73 amino acids per exonic region being observed.

*SL knockdown in S. mansoni sporocysts* - In an attempt to disrupt the trans-splicing mechanism, we designed an siRNA targeting the *S. mansoni* SL sequence. Over a seven-day period of cultivation in the presence of the siRNA, we monitored cultured sporocysts for various phenotypes, including a decrease in the miracidial/sporocyst transformation rate, mortality during the in vitro cultivation period and larval motility and length. Visual monitoring revealed that SL-siRNA treatment only altered the larval length phenotype, resulting in sporocysts with reduced size (Fig. 3A). To verify that this length alteration represented a significant effect, we measured captured images of live sporocysts and analysed the obtained data using Metamorph software. The average length measurements obtained for the sporocysts from the SL-siRNA treated-groups were significantly reduced compared to larvae from the control decoy-siRNA-treated and blank groups (Fig. 3B).

qPCR was performed to correlate the observed phenotypes and gene expression patterns. Notably, the five target transcripts randomly chosen from known SL-sequence-containing genes in the enriched EST dataset exhibited significant reductions of at least 50% compared to the decoy-siRNA treatment control. The examined calcium channel, ATPase inhibitor, phosphoserine-phosphohydrolase, thioredoxin and enolase transcripts displayed knockdown of 52%, 48%, 50%, 68% and 55%, respectively (Fig. 4). In addition, to check for nonspecific (off-target) knockdown, non-trans-spliced transcripts were assessed to determine expression levels following SL-siRNA treatment. No significant alteration of transcript levels was observed for SmZF1, SmRBx or SOD following SL-siRNA treatment. Thus, all of the tested trans-spliced transcripts analysed via RT-qPCR showed a similar decrease in the transcript expression level following SL-siRNA treatment, suggesting a systemic trans-splicing knockdown effect.

## DISCUSSION

The SL trans-splicing mechanism was first described as a post-transcriptional processing strategy for polycistronic transcripts in trypanosomatids (Agabian 1990). In subsequent years, this mechanism was observed in sev-

eral other organisms, but its functional role was never clearly defined outside the context of polycistronic transcription. One of the first hypotheses put forth to describe this phenomenon was that trans-splicing could be functionally associated with specific genes or gene categories. For example, in *Ciona intestinalis*, trans-splicing was suggested to predominantly regulate the expression of specific functional gene categories, such as the plasma and endomembrane system, Ca<sup>2+</sup> homeostasis and the actin cytoskeleton (Matsumoto et al. 2010). However, this hypothesis was not supported in *S. mansoni* as there was no clear evidence to show that the trans-splicing mechanism was linked to any particular gene category, biological process, molecular function, life-stage, sex, tissue or subcellular localisation of protein-coding transcripts (Davis et al. 1995).

In the present study, we analysed a diverse set of transcripts that are subjected to trans-splicing during different *S. mansoni* life cycle stages. As it appears that approximately 11% of *S. mansoni* transcripts are subjected to trans-splicing (Protasio et al. 2012) and given the complexity of the life cycle of this species, it is plausible that this process is important for the regulation of gene expression associated with parasite development and/or adaptation to different environments. On the other hand, the fraction of SL-containing transcripts reported thus far in *S. mansoni* is considerably lower compared to the high percentage of genes that undergo trans-splicing in organisms such as *C. elegans* and *Ascaris* spp, which can be up to 70% and 90%, respectively (Allen et al. 2011).

As previously stated, the EST dataset generated in this work was highly enriched in SL-containing sequences, showing a 6,000-fold increase in SL-containing sequences compared to the total number of *S. man-*

*soni* ESTs obtained from dbEST (60% SL-containing sequences were observed in our dataset vs. only 0.01% in dbEST). This group of SL-sequence-enriched transcripts represents a highly informative set of genes that could potentially shed light on several features of the trans-splicing mechanism. An interesting result obtained in this study comes from comparisons between our set of annotated trans-spliced sequences and those reported by Protasio et al. (2012). Approximately half of the protein-coding transcripts in our annotated dataset of 99 unique sequences were also included in the larger set generated by these authors. Two observations can be made from this comparison: (i) the strategy employed in the present study was appropriate for the investigation of SL transcripts, as demonstrated by the fact that 50 SL transcripts were identified in both datasets and (ii) because our dataset contained unique protein-coding genes that are not found in the broader Protasio dataset, it can be inferred that our approach allowed the retrieval of different SL transcripts. The differences in the contents of these datasets could be explained by the applied methodologies, in that Protasio et al. (2012) used a “whole-

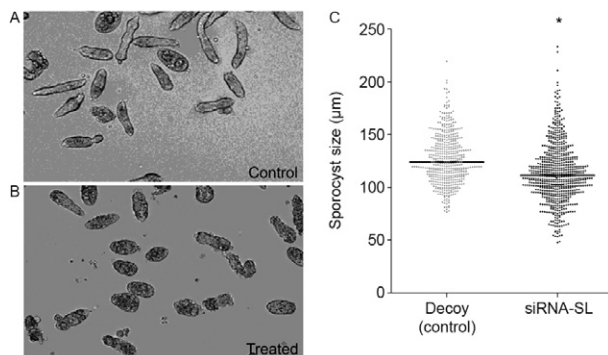


Fig. 3: in vitro cultured *Schistosoma mansoni* larvae seven days post-double stranded RNA treatments. **A, B:** brightfield photomicrographs of in vitro cultured *S. mansoni* sporocysts after seven days of treatments with spliced leader (SL)-small interfering (siRNA) (A) compared to the control decoy-siRNA (B), illustrating the effects of the exposure to SL-siRNA on sporocyst lengths; **C:** graphic representation of sporocyst length measurements (µm) after seven days of siRNA treatment by scatter plots with the calculated median values indicated by the horizontal bars. The median values for siRNA treatments were compared to decoy-siRNA (grey plots) treatment control. All measurements were statistically analysed using Mann-Whitney *U* test within each experiment. Asterisk means  $p < 0.0001$ .

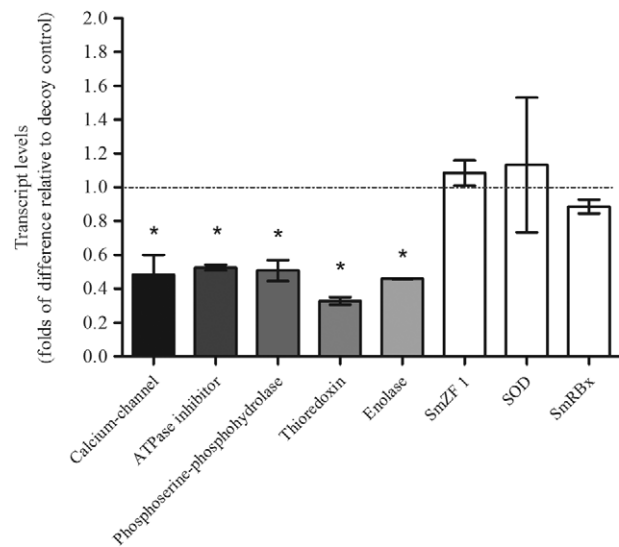


Fig. 4: histogram depicting the relative transcript levels of **small interfering (siRNA)-treated sporocysts after seven days of exposure compared to the decoy-siRNA control**. For each transcript tested, data is represented as mean fold-differences ( $\pm 2$  standard error) relative to the decoy-small RNA control (1.00). Gray bars represent sporocyst messenger RNA levels showing consistent and statistically significant decrease of known trans-spliced transcripts [calcium channel/decoy,  $p = 0.0056$ ; adenosine triphosphate (ATP)ase inhibitor/decoy,  $p = 0.0358$ ; phosphoserine-phosphohydrolase/decoy,  $p = 0.0136$ ; thioredoxin/decoy,  $p = 0.0358$ ; enolase/decoy,  $p = 0.0189$ ]. White bars represent relative transcript levels for non-trans-spliced transcripts in siRNA-treated sporocysts that showed no differences when compared to decoy-siRNA treated controls (SmZF1/decoy,  $p = 0.0755$ ; SOD/decoy,  $p = 0.8969$ ; SmRBx/decoy,  $p = 0.0765$ ). Transcript levels were determined by reverse transcriptase-quantitative polymerase chain reaction and data analysed using the  $\Delta\Delta C_t$  method followed by statistical analysis using the Mann-Whitney *U* test. Significance levels were set at  $p < 0.05$  (\*). Data were generated from three independent experiments. The gray line represents the decoy-control level.

transcriptome” approach, whereas we employed a more selective protocol involving the capture and enrichment of SL transcripts prior to cloning and sequencing.

Some of the transcripts in our *S. mansoni* dataset were previously described as SL-containing sequences, such as enolase and an ATPase inhibitor (Davis et al. 1995, Davis 1996). Although other protein-coding sequences described by Davis et al. (1995), such as synaptobrevin, a guanine nucleotide-binding protein and HMG-CoA reductase, were not identified within our dataset, sequences related to these genes or the pathways in which they are involved (e.g., Golgi Snare bet1, small GTPases or mevalonate pathway enzymes) were represented herein. Although a trans-spliced form of the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was not previously found in *S. mansoni*, this enzyme has been observed to undergo trans-splicing in *Caenorhabditis* spp. As reported herein, we also found a trans-spliced GAPDH transcript. This discrepancy between studies may be explained by the hypothesis that a transcript regulated by trans-splicing is not necessarily always processed by this mechanism. In other words, any given transcript may or may not be subjected to transplicing, which could account for the suggested role of trans-splicing as a mechanism for gene expression modulation and coordination (Davis et al. 1995).

In this work, we found *S. mansoni* orthologues of genes from different functional categories that had previously been described to undergo trans-splicing in other organisms. These genes encode ribosomal proteins, small nuclear ribonucleoproteins, members of the solute carrier family, GAPDH, thioredoxin, a mitochondrial ribosomal protein component, a WD-repeat containing protein, a peptidyl prolyl cis-trans isomerase, serine/threonine kinases and a cAMP-dependent protein kinase (Davis et al. 1995, Davis 1996). These findings support the hypothesis that there is some conservation among the genes regulated by trans-splicing, indicating that some genes may have maintained trans-splicing as a form of post-transcriptional regulation throughout evolution.

Bachvaroff and Place (2008) showed that the SL trans-splicing of dinoflagellate transcripts is correlated with their expression levels, suggesting that highly expressed genes are more likely to be SL trans-spliced. This correlation was made by comparing the levels of SL trans-spliced transcripts with the abundance of the corresponding proteins, as estimated through proteomic analyses (Beranova-Giorgianni 2003). Accordingly, we observed that at least 25% of the protein-coding transcripts that we classified as undergoing trans-splicing in *S. mansoni* encode proteins identified in previous proteomic studies (Curwen et al. 2004, Knudsen et al. 2005, Cass et al. 2007, Wu et al. 2009, Mathieson & Wilson 2010, Castro-Borges et al. 2011). These proteins include some glycolytic enzymes and many ribosomal proteins that we identified as TS transcripts in the present dataset. This observation is an indication of the importance of the trans-splicing mechanism among different organisms, as it could contribute to increasing protein abundance.

One of the trans-spliced transcripts of particular interest identified in this study is ubiquinol-cytochrome C

reductase complex ubiquinol binding protein (UbCRBP), which has been previously described as the first cistron of a trans-spliced polycistronic transcript, in which only the second gene (enolase) undergoes trans-splicing (Agabian 1990). The UbCRBP transcript also has been reported to undergo trans-splicing in *Echinococcus multilocularis* (Brehm et al. 2000), reinforcing the idea that trans-splicing is a conserved mechanism among selected orthologous genes. Our findings also revealed that the insertion of the SL in the *S. mansoni* UbCRBP sequence occurred before the second exon of the gene, which also contains an upstream AG acceptor splicing signal (Fig. 5). This result suggests that trans-splicing may generate alternatively spliced products, in which different exons could potentially receive the SL sequence. Alternative splice sites were previously observed in the 3-hydroxy-3-methyl-glutaryl CoA reductase transcript of *S. mansoni*, in which the third exon accepts the SL sequence (Rajkovic et al. 1990). Thus, alternative trans-splicing appears to be a conserved mechanism in this parasite, suggesting a unique means of expanding the protein repertoire of this organism. Whether alternative trans-splicing is confined to polycistronic transcripts is unknown.

Previous studies have suggested that short exons in pre-mRNAs are more prone to undergo trans-splicing (Davis et al. 1995). However, we did not observe any differences in exon size when we compared our dataset to the whole set of *S. mansoni* genes. Additionally, the observed protein lengths and the number of exons per sequence were equivalent when our dataset was compared to the set of all protein-coding sequences from SchistoDB.

Interestingly, many transcripts encoding proteins involved in the spliceosome machinery appear to undergo trans-splicing. This observation indicates that the trans-splicing mechanism may be self-regulated, which would represent a unique characteristic of this mechanism. In addition to transcripts encoding spliceosome proteins, the eukaryotic translation initiation factor 4e-binding protein and a subunit of the eukaryotic translation initiation factor 3 transcripts were also shown to undergo trans-splicing. These proteins are involved in a mechanism that enables the efficient translation of trimethyl-guanosine-capped mRNAs in nematodes (Wallace et al.

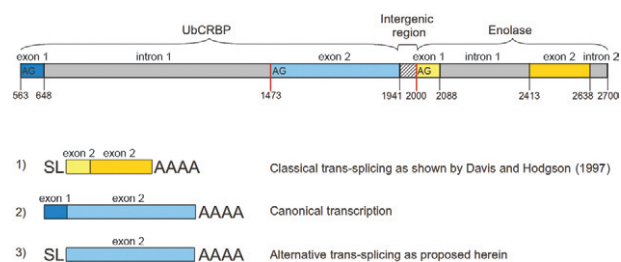


Fig. 5: representative scheme of an alternative spliced leader (SL) trans-splicing, as was observed in the case of ubiquinol-cytochrome C reductase complex ubiquinol binding protein (UbCRBP). While in normal transcription all exons are present in the final transcript, in alternative SL trans-splicing, the SL insertion is between the first and second exons, yielding a shortened transcript with a missing exon.



2010) and they are of great importance for trans-splicing because one of the functions attributed to this mechanism is facilitation of the translation of transcripts containing this modified cap.

To our knowledge, this is the first report describing an attempt to disrupt the trans-splicing mechanism in a metazoan using RNAi to assess its regulatory function. The introduction of SL-siRNA to in vitro-cultured sporocysts resulted in a phenotype characterised by a reduction of larval size. Because a large variety of SL-containing genes may have been affected by RNAi knockdown, it is difficult to speculate how this size reduction phenotype occurred. This phenotype may have resulted from a metabolic imbalance caused by a decrease in a large number of different trans-spliced transcripts. Proteins associated with crucial metabolic processes may have been affected by the knockdown of the trans-splicing mechanism, thereby resulting in a systemic decrease in metabolism, leading to possible parasite starvation and decreased larval length. Apart from the previous discussion of the occurrence of trans-splicing in glycolytic transcripts, other affected processes could also account for the diminished size of sporocysts following knockdown, for example, involving proteins associated with the cell cycle, metabolic pathways other than glycolysis and morpho-proteins, such as those described in our results. This phenotype may also reflect a type of stress caused by decreased activity of the trans-splicing mechanism. Taking these findings together, we can infer that the parasites subjected to knockdown were not in physiological equilibrium and that growth impairment is a common consequence of systemic stress and starvation, which could be caused by the reduced expression of transcripts under trans-splicing control in *S. mansoni*.

Although lethality was not observed after seven days of SL knockdown, our attempt to silence the trans-splicing machinery decreased the expression of SL-containing transcripts by 60%. It is likely that this partial knockdown at the mRNA expression level may have exerted only a minor effect on intact larvae, not only because transcripts were still present in these parasites, but also because their encoded proteins may have persisted for an extended period of time, depending on their turnover rate. Thus, the remaining transcript levels and residual protein pools were most likely sufficient to maintain larval viability, even though the larvae appeared to be morphologically stunted. Another possible explanation is that only a fraction of the molecules produced from a given transcript undergo trans-splicing. Interestingly, all of the tested trans-spliced transcripts exhibited a similar decrease at the transcript expression level, suggesting a systemic trans-splicing knockdown effect following SL-siRNA treatment. Because 11% of the *S. mansoni* transcript population appears to be trans-spliced, a hypothesis explaining the limited transcript knockdown observed could include saturation of the components of the RNAi machinery.

In this study, we generated and analysed a diverse set of *S. mansoni* ESTs that were highly enriched in transcripts bearing the SL sequence. In agreement with the literature, the SL sequence-containing transcripts were not found to be associated with specific gene categories,

subcellular localisations or life cycle stages within the transcript dataset we analysed. We also investigated protein lengths, the number of exons and exon length among the SL-containing transcripts and found no differences compared to the entire set of *S. mansoni* transcripts. Disruption of the SL trans-splicing mechanism in *S. mansoni* sporocysts through RNAi resulted in a reduction of larval size. This result provides evidence of the importance of this mechanism for the development of this organism and suggests a crucial role for the regulation of metabolic processes by SL trans-splicing. To determine whether the SL trans-splicing mechanism has a unique ancestral origin or multiple unrelated origins, we searched for homologous proteins in other organisms in which the transcripts were also trans-spliced. This search provided support for the hypothesis of the origin of this mechanism in a common ancestor, although further analyses are needed. The association of SL transcripts with a wide range of different genes suggests that this mechanism plays an important regulatory role, influencing the expression levels of different proteins as well as the protein repertoires observed in different life stages and under distinct environmental conditions. To our knowledge, this is the most comprehensive survey of SL transcripts conducted in schistosomes to date and our results provide a valuable resource for further studies addressing the mechanisms of SL trans-splicing at both the biological and molecular levels.

#### ACKNOWLEDGEMENTS

To Neuza Antunes, for contribution to experimental procedures analysis.

#### REFERENCES

- Agabian N 1990. Trans splicing of nuclear pre-mRNAs. *Cell* 61: 1157-1160.
- Allen MA, Hillier LW, Waterston RH, Blumenthal T 2011. A global analysis of *C. elegans* trans-splicing. *Genome Res* 21: 255-264.
- Bachvaroff TR, Place AR 2008. From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS ONE* 3: e2929.
- Basch PF 1981. Cultivation of *Schistosoma mansoni* in vitro. I. Establishment of cultures from cercariae and development until pairing. *J Parasitol* 67: 179-185.
- Beranova-Giorgianni S 2003. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strengths and limitations. *Trends Anal Chem* 22: 9.
- Blumenthal T, Gleason KS 2003. *Caenorhabditis elegans* operons: form and function. *Nat Rev Genet* 4: 112-120.
- Boguski MS, Lowe TM, Tolstoshev CM 1993. dbEST - database for "expressed sequence tags". *Nat Genet* 4: 332-333.
- Brehm K, Hubert K, Sciuotto E, Garate T, Frosch M 2002. Characterization of a spliced leader gene and of trans-spliced mRNAs from *Taenia solium*. *Mol Biochem Parasitol* 122: 105-110.
- Brehm K, Jensen K, Frosch M 2000. mRNA trans-splicing in the human parasitic cestode *Echinococcus multilocularis*. *J Biol Chem* 275: 38311-38318.
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT 2009. The MIQE guidelines: minimum information

- for publication of quantitative real-time PCR experiments. *Clin Chem* 55: 611-622.
- Cass CL, Johnson JR, Califf LL, Xu T, Hernandez HJ, Stadecker MJ, Yates JR 3rd, Williams DL 2007. Proteomic analysis of *Schistosoma mansoni* egg secretions. *Mol Biochem Parasitol* 155: 84-93.
- Castro-Borges W, Simpson DM, Dowle A, Curwen RS, Thomas-Oates J, Beynon RJ, Wilson RA 2011. Abundance of tegument surface proteins in the human blood fluke *Schistosoma mansoni* determined by QconCAT proteomics. *J Proteomics* 74: 1519-1533.
- Curwen RS, Ashton PD, Johnston DA, Wilson RA 2004. The *Schistosoma mansoni* soluble proteome: a comparison across four life-cycle stages. *Mol Biochem Parasitol* 138: 57-66.
- Davis RE 1996. Spliced leader RNA trans-splicing in metazoa. *Parasitol Today* 12: 33-40.
- Davis RE, Hardwick C, Tavernier P, Hodgson S, Singh H 1995. RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. *J Biol Chem* 270: 21813-21819.
- Davis RE, Hodgson S 1997. Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in *Schistosoma mansoni*. *Mol Biochem Parasitol* 89: 25-39.
- Ewing B, Hillier L, Wendl MC, Green P 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185.
- Fenwick A, Webster JP 2006. Schistosomiasis: challenges for control, treatment and drug resistance. *Curr Opin Infect Dis* 19: 577-582.
- Franco GR, Adams MD, Soares MB, Simpson AJ, Venter JC, Pena SD 1995. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene* 152: 141-147.
- Gunzl A, Bruderer T, Laufer G, Schimanski B, Tu LC, Chung HM, Lee PT, Lee MG 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell* 2: 542-551.
- Hastings KE 2005. SL trans-splicing: easy come or easy go? *Trends Genet* 21: 240-247.
- Huang X, Madan A 1999. CAP3: A DNA sequence assembly program. *Genome Res* 9: 868-877.
- Kalinna BH, Brindley PJ 2007. Manipulating the manipulators: advances in parasitic helminth transgenesis and RNAi. *Trends Parasitol* 23: 197-204.
- Kitts PA, Madden TL, Sicotte H, Black L, Ostell JA 2011. UniVec Database. Available from: [ncbi.nlm.nih.gov/VecScreen/UniVec.html](http://ncbi.nlm.nih.gov/VecScreen/UniVec.html).
- Knudsen GM, Medzihradzky KF, Lim KC, Hansell E, McKerrow JH 2005. Proteomic analysis of *Schistosoma mansoni* cercarial secretions. *Mol Cell Proteomics* 4: 1862-1875.
- Krause M, Hirsh D 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49: 753-761.
- Lee MG, Van der Ploeg LH 1997. Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol* 51: 463-489.
- Leinonen R, Sugawara H, Shumway M 2011. The sequence read archive. *Nucleic Acids Res* 39: D19-D21.
- Lidie KB, van Dolah FM 2007. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. *J Eukaryot Microbiol* 54: 427-435.
- Livak KJ, Schmittgen TD 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-408.
- Logan-Klumpler FJ, de Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, Carver T, Aslett M, Olsen C, Subramanian S, Phan I, Farris C, Mitra S, Ramasamy G, Wang H, Tivey A, Jackson A, Houston R, Parkhill J, Holden M, Harb OS, Brunk BP, Myler PJ, Roos D, Carrington M, Smith DF, Hertz-Fowler C, Berriman M 2012. GeneDB - an annotation database for pathogens. *Nucleic Acids Res* 40: D98-D108.
- Mann VH, Morales ME, Kines KJ, Brindley PJ 2008. Transgenesis of schistosomes: approaches employing mobile genetic elements. *Parasitology* 135: 141-153.
- Mathieson W, Wilson RA 2010. A comparative proteomic study of the undeveloped and developed *Schistosoma mansoni* egg and its contents: the miracidium, hatch fluid and secretions. *Int J Parasitol* 40: 617-628.
- Matsumoto J, Dewar K, Wasserscheid J, Wiley GB, Macmil SL, Roe BA, Zeller RW, Satou Y, Hastings KE 2010. High-throughput sequence analysis of *Ciona intestinalis* SL trans-spliced mRNAs: alternative expression modes and gene function correlates. *Genome Res* 20: 636-645.
- McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, Luthe DS, Bridges SM, Burgess SC 2006. AgBase: a functional genomics resource for agriculture. *BMC Genomics* 7: 229.
- Mourão MM, Dinguirard N, Franco GR, Yoshino TP 2009a. Phenotypic screen of early-developing larvae of the blood fluke, *Schistosoma mansoni*, using RNA interference. *PLoS Negl Trop Dis* 3: e502.
- Mourão MM, Dinguirard N, Franco GR, Yoshino TP 2009b. Role of the endogenous antioxidant system in the protection of *Schistosoma mansoni* primary sporocysts against exogenous oxidative stress. *PLoS Negl Trop Dis* 3: e550.
- Nilsen TW 1993. Trans-splicing of nematode premessenger RNA. *Annu Rev Microbiol* 47: 413-440.
- Pearce EJ, Freitas TC 2008. Reverse genetics and the study of the immune response to schistosomes. *Parasite Immunol* 30: 215-221.
- Pessôa SB, Martins AV 1982. *Parasitologia médica*, 11th ed., Guanabara Koogan, Rio de Janeiro, 872 pp.
- Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, de Silva N, Velarde GS, Anderson TJ, Clark RC, Davidson C, Dillon GP, Holroyd NE, LoVerde PT, Lloyd C, McQuillan J, Oliveira G, Otto TD, Parker-Manuel SJ, Quail MA, Wilson RA, Zerlotini A, Dunne DW, Berriman M 2012. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis* 6: e1455.
- Rajkovic A, Davis RE, Simonsen JN, Rottman FM 1990. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci USA* 87: 8879-8883.
- Skelly PJ, Da'dara A, Harn DA 2003. Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *Int J Parasitol* 33: 363-369.
- Smithers SR, Terry RJ 1965. The infection of laboratory hosts with cercariae of *Schistosoma mansoni* and the recovery of the adult worms. *Parasitology* 55: 695-700.
- Stover NA, Steele RE 2001. Trans-spliced leader addition to mRNAs in a cnidarian. *Proc Natl Acad Sci USA* 98: 5693-5698.
- Vandenbergh AE, Meedel TH, Hastings KE 2001. mRNA 5'-leader trans-splicing in the chordates. *Genes Dev* 15: 294-303.
- Wallace A, Filbin ME, Veo B, McFarland C, Stepinski J, Jankowska-Anyszka M, Darzynkiewicz E, Davis RE 2010. The nematode eukaryotic translation initiation factor 4E/G complex works with a trans-spliced leader stem-loop to enable efficient translation of trimethylguanosine-capped RNAs. *Mol Cell Biol* 30: 1958-1970.

- WHO - World Health Organization 2012. Schistosomiasis. Fact sheet 115. Available from: [who.int/mediacentre/factsheets/fs115/en](http://who.int/mediacentre/factsheets/fs115/en).
- Wu XJ, Sabat G, Brown JF, Zhang M, Taft A, Peterson N, Harms A, Yoshino TP 2009. Proteomic analysis of *Schistosoma mansoni* proteins released during in vitro miracidium-to-sporocyst transformation. *Mol Biochem Parasitol* 164: 32-44.
- Yoshino TP, Dinguirard N, Mourão MM 2010. In vitro manipulation of gene expression in larval *Schistosoma*: a model for postgenomic approaches in Trematoda. *Parasitology* 137: 463-483.
- Yoshino TP, Laursen JR 1995. Production of *Schistosoma mansoni* daughter sporocysts from mother sporocysts maintained in synxenic culture with *Biomphalaria glabrata* embryonic (Bge) cells. *J Parasitol* 81: 714-722.
- Zayas RM, Bold TD, Newmark PA 2005. Spliced-leader trans-splicing in freshwater planarians. *Mol Biol Evol* 10: 2048-2054.
- Zerlotini A, Heiges M, Wang H, Moraes RL, Dominitini AJ, Ruiz JC, Kissinger JC, Oliveira G 2009. SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic Acids Res* 37: D579-D582.

Primers used in reverse transcriptase-quantitative polymerase chain reaction experiments for testing the levels of trans-spliced transcripts after RNA interference treatment

Protein	Protein ID	Forward primer spliced leader	Reverse primer
Trans-spliced transcripts (positive trans-splicing control)			
High voltage-activated calcium channel-Cav B2	AY277532.1	GTCACGGTTTTACTCTTGT	GGCAAATCGCTACCTGAAT
Phosphoserine phosphohydrolase-like	SMU30262	GTCACGGTTTTACTCTTGT	GTCCAAACAGACGCATTTGA
ATPase inhibitor-like	U30266.1	GTCACGGTTTTACTCTTGT	GAACCACCAGCTTCACGAAT
Thioredoxin precursor	AY948980.1	GTCACGGTTTTACTCTTGT	AAGCAGACGTATTGAAGGGAAA
Enolase	SMU30175	GTCACGGTTTTACTCTTGT	TCCACCTCAACTGTGGGATT
Constitutive transcripts (endogenous normalisation control)			
$\alpha$ -tubulin	XM_002579987.1	CAAATGGGAAATGCTTGTTG	TGAACGAGTCATCACCACT
Non-transpliced transcripts (negative trans-splicing control)			
SmZF1	AF316827.1	ACTTCTCTCAGAAATCCAGCCT	TGGAGAGGATTATACAATCTGGTT
Superoxide dismutase 1	M27529.1	GATCCGGCTATTGCTTCATT	CTGCCACGCTTCCATTAAC
SmRing box	DQ466078.1	GGCATCCCTCAGTGAGAATAA	TCCGGCAGATAGCACACAATTA

## Functional classification and description of trans-spliced transcripts in the final dataset

Class	Protein name (SchistoDB ID)	Protein function
Development	bola-like protein my016 (012050)	Morphogenesis, cell growth
	cactin (038640)	Morphogenesis, cell growth and survival
	chibby protein pkd2 interactor (035010)	Morphogenesis, cell growth
Cell cycle regulation and apoptosis	chl1 helicase (130210)	Chromosome segregation
	enhancer of rudimentary protein (175210)	Pyrimidine synthesis and cell cycle
	jun activation domain binding protein (190100)	Activation of mitotic checkpoint
	programmed cell death protein (067490)	Cell regulation and apoptosis
	pten-related phosphatase (012400)	Cell regulation and apoptosis
	regulator of chromosome condensation-related (074990)	Chromosome condensation
Replication and repair	cript-related (132860)	DNA double strand break repair
	endonuclease III (006710)	DNA repair of damaged pyrimidine
	pol-related (186960.2)	DNA replication
Chromatin modification	histone h1/h5 (003770)	Nucleosome formation
	histone deacetylase hda2 (138770)	Histone deacetylation
	mrg-binding protein (130630)	Histone acetylation
Transcription and post-transcriptional regulation	DNA-directed rna polymerase I (072800)	rRNA synthesis
	lsm1, putative (175550)	RNA splicing (snRNP formation)
	nucleotide binding protein 2 (154750)	RNA processing, traffic and stability (snRNP formation)
	pre-mRNA splicing factor (080110)	RNA splicing
	ribonucleic acid binding protein S1(070200)	RNA traffic and surveillance
	small nuclear ribonucleoprotein U1a, U2b (069880)	RNA splicing (spliceosome protein)
	serine/arginine rich splicing factor (113620.4)	RNA splicing (spliceosome protein)
Translation	eukaryotic translation initiation factor 4e-binding protein (074390)	Inhibition of protein translation
	eukaryotic translation initiation factor 3 subunit (158500)	mRNA-ribosome interaction
	mitochondrial ribosomal protein L24 (035410)	Mitochondrial protein synthesis
	mitochondrial ribosomal protein S33 (039980)	Mitochondrial protein synthesis
	translation machinery-associated protein (169920)	Translation
Protein folding	peptidyl-prolyl cis-trans isomerase-like 3 (080150)	Acceleration of protein folding
	rotamase (160790)	Acceleration of protein folding
Protein processing, modification and degradation	26S proteasome non-atpase regulatory subunit (181380)	Protein degradation via proteasome
	ammd-like (peptidase) (074940)	Protein degradation
	aspartic proteinase (013040.1)	Protein degradation
	beta-1,4-galactosyltransferase (153110)	Glycosyltransferase activity
	glycosyltransferase (046880.2)	Glycosyltransferase activity
	methionine aminopeptidase (011120)	Removal of initial methionine
	o-sialoglycoprotein endopeptidase (012450)	Sialoglycoprotein degradation
	protein-l-isoaspartate o-methyltransferase (058140.2)	Proteins repair and degradation
	peptidase (165910)	Protein degradation
ring finger (028430)	Protein ubiquitination	



Class	Protein name (SchistoDB ID)	Protein function
Signal transduction	hybrid protein kinase, other group, WNK family (154440)	Chloride ion traffic
	protein phosphatase 2C (133060.3)	Serine/threonine phosphatase
	ral guanine nucleotide dissociation stimulator (154920)	GTPase activation
	rab 15, 13, 10, 1, 35, 5 (072290)	Small GTPase
	rho GDP-dissociation inhibitor-related (045610)	Interaction with Rho GTPases
	serine/threonine kinase, CAMK group, CAMKL family, AMPK subfamily (142990)	Signal transduction pathways for cellular energy homeostasis
Stress response	serine/threonine kinase, CMGC group, RCK family, MAK subfamily (132890)	Signal transduction pathways for cell cycle regulation
	tbcl domain family member (152820)	Small GTPase activation
	glutaredoxin (006550.2)	Antioxidant defence
Cytoskeleton organisation	hsp70-interacting protein (064860)	Protection against physiologic stress
	thioredoxin mitochondrial type (037530)	Antioxidant defence
	beta-parvin-related (133200)	Regulation of cell adhesion and cytoskeleton organisation
	cofilin, actophorin (120700.2)	Actin binding
	dynactin subunit 3 (dynactin light chain p24) (136990)	Microtubule and dynein binding
Carbohydrate metabolism	nuclear movement protein nudc (103320)	Dynein binding
	tektin (059600)	Structural component of microtubules
	tropomodulin (003710)	Actin growth regulation
	tropomyosin, putative (031770.4)	Actin mechanics regulation
	enolase (024110)	Glycolysis
Lipid metabolism	glyceraldehyde-3-phosphate dehydrogenase (056970.1)	Glycolysis
	l-lactate dehydrogenase (038950)	Glycolysis
	ribulose-5-phosphate-3-epimerase (104580.2)	Pentose phosphate pathway
	1-acylglycerol-3-phosphate o-acyltransferase (000070)	Lipid metabolism
	acetyl-CoA C-acyltransferase (129330)	Lipid catabolism
Energy homeostasis	acyl-coenzyme A binding domain containing (147930)	Lipid metabolism
	arv1 (021280)	Sterol homeostase
	serine palmitoyltransferase I (028080)	Sphingolipids biosynthesis
	3-dehydroecdysone 3alpha-reductase-related (042680)	Ecdysteroids inactivation
	isopentenyl-diphosphate delta isomerase (130430)	Biosynthesis of isoprenoids
	sterol reductase-related (124300)	Biosynthesis of steroids
	atpase inhibitor (023010)	Inhibition of ATP synthesis
	ATP synthase delta chain (082120)	Catalysis of ATP synthesis
	cytochrome C oxidase copper chaperone (029150)	Involved in ATP synthesis
	cytochrome C subunit I	Involved in ATP synthesis
ectonucleotide pyrophosphatase/phosphodiesterase (104270)	ATP hydrolysis	
hexaprenyldihydroxybenzoate methyltransferase (029060)	Ubiquinone and ATP synthesis	
ubiquinone biosynthesis protein (106190)	Ubiquinone and ATP synthesis	
ubiquinol-cytochrome C reductase complex ubiquinol binding protein (024110)	Involved in ATP synthesis	



Class	Protein name (SchistoDB ID)	Protein function
Cofactor metabolism	molybdopterin-binding (126650)	Molybdopterin cofactor synthesis
	uroporphyrinogen-III synthase (079840)	Haeme synthesis
Amino acid catabolism	4-hydroxyphenylpyruvate dioxygenase (007960)	Tyrosine catabolism
	valacyclovir hydrolase (193750)	Alpha-amino acid catabolism
Transport and membrane turnover	fatty acid binding protein (174440.4)	Fatty acids transport
	mitochondrial carrier protein-related (058130)	Transport across mitochondrial membrane
	nuclear transport factor (037700)	Proteins transport
	peripheral-type benzodiazepine receptor, putative (102510)	Transport of cholesterol, porphyrin and anions
	solute carrier family 35 member C1, putative (155830)	Transport of GDP-fucose
	B-cell receptor-associated protein-like protein (175660)	Vesicular transport of proteins between the ER and the Golgi
	dynamamin-associated protein (135790)	Vesicular transport
	golgi snare bet1-related (130940)	Vesicular transport between the ER and the Golgi
	rabb and c (169460)	Vesicular transport of proteins between the membrane and the Golgi
	small VCP/p97-interacting protein (194560)	ER integrity maintenance
	vacuolar ATP synthase proteolipid subunit 1, 2, 3 (004310.1)	Transport across membranes and regulation of internal pH
Miscellanea (including unannotated proteins)	amyloid beta A4 protein related (144990)	Involved in several physiological processes
	autophagy protein 16-like (135430)	Involved in cellular autophagy
	metal dependent hydrolase-related (024490)	Presents hydrolase activity
	protein C14orf153 precursor (147820)	Unannotated protein
	Sj-Ts1 (123270)	Unannotated transmembrane protein
	tetraspanin D76 (041460)	Involved in cell adhesion
	WD repeat protein SL1-17 (049520)	Involved in several physiological processes

transcripts were classified according to their biological function after manual annotation. ATP: adenosine triphosphate; ER: endoplasmic reticulum; GDP: guanidine 5'-diphosphate-beta-1; mRNA: messenger RNA; rRNA: ribosomal RNA; snRNP: small nuclear ribonucleoprotein.

**Detailed annotation of all proteins derived from trans-spliced transcripts  
classified in this study**

<b>Protein Name</b> bola-like protein my016	<b>Life-cycle Phase</b> Adult worms of mixed gender, Schistosomula, Female worm, Male worm	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> 0004129 cytochrome-c oxidase activity		
<b>Description</b> The morpho-protein BolA from <i>Escherichia coli</i> is related to round morphology and may be involved in switching the cell between elongation and septation systems during cell division. The expression of BolA is growth rate regulated and is induced during the transition into the stationary phase. BolA is also induced by stress during early stages of growth and may have a general role in stress response (InterPro IPR002634).		
<b>Protein Name</b> cactin-related	<b>Life-cycle Phase</b> Male worms	<b>Localization</b> Nucleus (Psort and GO)
<b>GO annotation</b> 0000398 nuclear mRNA splicing, via spliceosome 0005634 nucleus 0005681 spliceosomal complex 0006397 mRNA processing 0008380 RNA splicing 0071013 catalytic step 2 spliceosome		
<b>Description</b> Cactin specifically interacts with the cactus protein, having a possible function in the Rel pathway controlling the formation of dorsal-ventral embryonic polarity. In mammals, Rel transcription factors control the immune and inflammatory responses, cell growth and survival, differentiation of vertebrate limbs and muscle development. In <i>Drosophila</i> it is involved with the establishment of dorsal-ventral polarity in the early embryo. Cactin seems to act through the degradation of cactus, liberating the dorsal protein to translocate to the nucleus and regulating transcription of target genes. A cactin homologous in <i>Caenorhabditis elegans</i> is required for correct migration of distal tip cells and for gonad morphogenesis.		
<b>Protein Name</b> chibby protein pkd2 interactor, putative	<b>Life-cycle Phase</b> Adult worms of mixed genders, Female worms, Male worms	<b>Localization</b> Mitochondria (Psort)
<b>GO annotation</b> null		
<b>Description</b> Chibby is a nuclear protein that directly interacts with beta-catenin and seems to inhibit beta-catenin-mediated transcriptional activation. Beta-catenin is a subunit of the cadherin complex and has been implicated as a component of the Wnt signaling pathway.		
<b>Protein Name</b> chl1 helicase homolog, putative	<b>Life-cycle Phase</b> Schistosomula	<b>Localization</b> Nucleus (Psort and GO)



<b>GO annotation</b>		
0000166	nucleotide binding	
0003676	nucleic acid binding	
0003677	DNA binding	
0004003	ATP-dependent DNA helicase activity	
0004386	helicase activity	
0005524	ATP binding	
0005634	nucleus	
0006139	nucleobase-containing compound metabolic process	
0008026	ATP-dependent helicase activity	
0016787	hydrolase activity	
0016817	hydrolase activity, acting on acid anhydrides	
0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	
<b>Description</b>		
Chl1 helicases are nuclear proteins that present an ATP binding motif, which is essential for its function in chromosome segregation. ATPases/DNA helicases may alter DNA topology to allow loading of proteins involved in maintaining sister chromatid cohesion. <i>Saccharomyces cerevisiae chl1</i> mutants have a significant increase in the rate of chromosome missegregation.		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
enhancer of rudimentary protein, putative	All	Cytoplasm (Psort)
<b>GO annotation</b>		
0007049	cell cycle	
<b>Description</b>		
The <i>Drosophila</i> protein 'enhancer of rudimentary' is a small protein whose function is not yet clear. From an evolutionary point of view, it is highly conserved and has been found to exist in probably all multicellular eukaryotic organisms. It has been proposed that this protein has functions implicated in pyrimidine biosynthesis and the cell cycle (InterPro IPR000781).		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
jun activation domain binding protein, putative	Male worms	Extracellular (Psort)
<b>GO annotation</b>		
null		
<b>Description</b>		
Jun activation domain binding protein (Jab) 1 is required for the hyperphosphorylation of 53BP1 upon mitotic stress conditions and is involved in proper activation of the mitotic checkpoint mechanism. Jun proteins are a component of the activator protein 1 (AP1) complex, which regulates gene expression in response to a variety of stimuli, including cytokines growth factors, stress, and bacterial and viral infections. AP1 in turn controls a number of cellular processes including differentiation, proliferation and apoptosis.		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
programmed cell death protein, putative	Adult worms of mixed gender, Egg	Mitochondria (Psort)
<b>GO annotation</b>		
0005509	calcium ion binding	
<b>Description</b>		
Involved in the programmed cell death (apoptosis) pathways.		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
pten-related phosphatase (plip)	Adult worms of mixed	Cytoplasm,

EC: 3.1.3.16 3.1.3.48	gender, Female worms	Extracellular (Psort)
<b>GO annotation</b> 0004725 protein tyrosine phosphatase activity 0006470 protein dephosphorylation 0008138 protein tyrosine/serine/threonine phosphatase activity 0016311 dephosphorylation 0016787 hydrolase activity 0016791 phosphatase activity 0035335 peptidyl-tyrosine dephosphorylation		
<b>Description</b> PTEN (Phosphatase and tensin homolog) is related to the cell cycle regulation and apoptosis. The loss of just one copy of this gene has been shown to trigger uncontrolled cell growth. The protein encoded by this gene is a phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase. It contains a tensin-like domain as well as a catalytic domain similar to that of the dual specificity protein tyrosine phosphatases. Unlike most of the protein tyrosine phosphatases, this protein preferentially dephosphorylates phosphoinositide substrates. It negatively regulates intracellular levels of phosphatidylinositol-3,4,5-trisphosphate in cells and functions as a tumor suppressor by negatively regulating the Akt/PKB signaling pathway.		
<b>Protein Name</b> regulator of chromosome condensation-related	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Nucleus (Psort)
<b>GO annotation</b> null		
<b>Description</b> This protein is involved in chromossome condensation, which is a crucial step in the preparation of cells for division.		
<b>Protein Name</b> cript-related	<b>Life-cycle Phase</b> Female worms, male worms	<b>Localization</b> Nucleus (GO)
<b>GO annotation</b> 0005634 nucleus 0006355 regulation of transcription, DNA-dependent 0043189 H4/H2A histone acetyltransferase complex		
<b>Description</b> This protein contains a 5'-tyrosyl DNA phosphodiesterase domain (TDP2). In human, TDP2 is required for the efficient repair of topoisomerase II-induced DNA double strand breaks. The topoisomerase is covalently linked by a phosphotyrosyl bond to the 5'-terminus of the break. TDP2 cleaves the DNA 5'-phosphodiester bond and restores 5'-phosphate termini, needed for subsequent DNA ligation, and hence repair of the break. This family belongs to the large exonuclease/endonuclease/phosphatase superfamily that contains functionally diverse enzymes that share a common catalytic mechanism of cleaving phosphodiester bonds.		
<b>Protein Name</b> endonuclease III, putative EC: 3.2.2	<b>Life-cycle Phase</b> Schistosomula	<b>Localization</b> Nucleus, Cytoplasm (Psort)
<b>GO annotation</b> 0003677 DNA binding 0003824 catalytic activity 0004519 endonuclease activity 0006281 DNA repair		

0006284	base-excision repair		
0051539	4 iron, 4 sulfur cluster binding		
<b>Description</b>			
Endonuclease III plays a role in the release of damaged pyrimidines from double-stranded DNA and the subsequent cleavage of the 3' from the apurinic site.			
<b>Protein Name</b> pol-related		<b>Life-cycle Phase</b> Adult worms of mixed gender	<b>Localization</b> Nucleus, Cytoplasm (Psort)
<b>GO annotation</b>			
0006259	DNA metabolic process		
0008150	biological process		
0008152	metabolic process		
0009058	biosynthetic process		
<b>Description</b>			
Pol, is a reverse transcriptase, a DNA polymerase enzyme that transcribes single-stranded RNA into single-stranded DNA. It is also a DNA-dependent DNA polymerase which synthesizes the second strand of DNA of the double helix DNA once the RNA has been reverse transcribed.			
<b>Protein Name</b> histone h1/h5, putative		<b>Life-cycle Phase</b> Egg	<b>Localization</b> Nucleus (Psort and GO)
<b>GO annotation</b>			
0000786	nucleosome		
0003677	DNA binding		
0005634	nucleus		
0005694	chromosome		
0006334	nucleosome assembly		
<b>Description</b>			
Histone proteins have central roles in both chromatin organization (as structural units of the nucleosome) and gene regulation (as dynamic components that have a direct impact on DNA transcription and replication). The core histone octamer is formed by a central H3-H4 tetramer and two flanking H2A-H2B dimers. In addition, there is a linker histone, called H1 (or H5 in avian species) which is crucial to the higher-order organization of the chromatin. This enzyme, present in all multicellular organisms is the most divergent one, with cell-type and stage-specific variation (InterPro IPR005818).			
<b>Protein Name</b> histone deacetylase hda2, putative EC: 3.5.1.98		<b>Life-cycle Phase</b> Male worms	<b>Localization</b> Nucleus (Psort)
<b>GO annotation</b>			
0006464	protein modification process		
0006810	transport		
0006950	response to stress		
0006996	organelle organization		
0007010	cytoskeleton organization		
0007154	cell communication		
0007165	signal transduction		
0008150	biological process		
0008152	metabolic process		
0008219	cell death		

0009056	catabolic process
0009058	biosynthetic process
0009605	response to external stimulus
0015031	protein transport
0016043	cellular component organization
0019538	protein metabolic process
0044238	primary metabolic process
0050789	regulation of biological process

**Description**

HDA2 and HDA3 associate with HDA1 to generate the activity of the HDA1 histone deacetylase complex. HDA1 interacts with itself and with the HDA2-HDA3 subcomplex to form a probable tetramer and these interactions are necessary for catalytic activity. The HDA1 histone deacetylase complex is responsible for the deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4). Histone deacetylation gives a tag for epigenetic repression and plays an important role in transcriptional regulation, cell cycle progression and developmental events (InterPro IPR021006).

Protein Name	Life-cycle Phase	Localization
mrg-binding protein, putative	Female worms, Male worms	Nucleus (Psort)

**GO annotation**

0006139	nucleobase-containing compound metabolic process
0006351	transcription, DNA-dependent
0008150	biological process
0008152	metabolic process
0009058	biosynthetic process
0044238	primary metabolic process
0050789	regulation of biological process
0006139	nucleobase-containing compound metabolic process

**Description**

Component of a histone acetyltransferase complex which is involved in transcriptional activation of specific genes by acetylation of nucleosomal histones H4 and H2A. This modification may both alter nucleosome-DNA interactions and promote interaction of the modified histones with other proteins which positively regulate transcription. This complex may be required for the activation of transcriptional programs associated with oncogene and proto-oncogene mediated growth induction, tumor suppressor mediated growth arrest and replicative senescence, apoptosis, and DNA repair. The complex may also play a direct role in DNA repair when recruited to sites of DNA damage (UniProt Q9NV56).

Protein Name	Life-cycle Phase	Localization
DNA-directed rna polymerase I, putative	Egg, Male worms	Nucleus / Extracellular (Psort)

**GO annotation**

0006351	transcription, DNA-dependent
---------	------------------------------

**Description**

RNA polymerase I, one of three nuclear DNA-directed RNA polymerases found in all eukaryotes, is a multisubunit complex which typically produces rRNAs. Two large subunits comprise the most conserved portion including the catalytic site and share similarity with other eukaryotic and bacterial multisubunit RNA polymerases. The remainder of the complex is composed of smaller subunits (generally ten or more), some of which are also found in RNA polymerase III and others of which are also found in RNA polymerases II and III. Although the core is competent to mediate

ribonucleic acid synthesis, it requires additional factors to select the appropriate template.		
<b>Protein Name</b> lsm1, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender, Egg, Schistosomula, Female worms, Male worms	<b>Localization</b> Mitochondria (Psort)
<b>GO annotation</b> null		
<b>Description</b> U6 snRNA-associated Sm-like protein Lsm1 is part of the Sm-like proteins class, which were identified in a variety of organisms based on the presence of the Sm sequence motif. The Sm-like proteins are thought to form a stable heteromer present in tri-snRNP particles, which are important for pre-mRNA splicing.		
<b>Protein Name</b> nucleotide binding protein 2 (nbp 2), putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Nucleus (Psort)
<b>GO annotation</b> 0000166 nucleotide binding 0003676 nucleic acid binding 0005524 ATP binding		
<b>Description</b> This protein contains two domains, the RRM and the MRP domains. RRM (RNA recognition motif) is a highly abundant eukaryotic domain, found in proteins involved in post-transcriptional gene expression processes including mRNA and rRNA processing, RNA export, and RNA stability. The RRM domain is found in a variety heterogeneous nuclear ribonucleoproteins (hnRNPs), proteins implicated in regulation of alternative splicing, and protein components of small nuclear ribonucleoproteins (snRNPs). MRP (Multiple Resistance and pH adaptation) is a homologue of the Fer4_NifH superfamily. Like the other members of the superfamily, MRP contains an ATP-binding domain at the N-termini. It is found in bacteria as a membrane-spanning protein and functions as a Na <sup>+</sup> /H <sup>+</sup> antiporter.		
<b>Protein Name</b> pre-mRNA splicing factor, putative	<b>Life-cycle Phase</b> Egg, Schistosomula, Female worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> null		
<b>Description</b> Pre-mRNA splicing is a predominantly co-transcriptional event which involves a large number of essential splicing factors.		
<b>Protein Name</b> ribonucleic acid binding protein S1, putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoplasm, Nucleus (Psort)
<b>GO annotation</b> 0006139 nucleobase-containing compound metabolic process 0008150 biological process 0008152 metabolic process		
<b>Description</b> This gene encodes a protein that is part of a post-splicing multiprotein complex involved in both mRNA nuclear export and surveillance. mRNA surveillance detects exported mRNAs with		

truncated open reading frames and initiates nonsense-mediated mRNA decay (NMD). When translation ends upstream from the last exon-exon junction, this triggers NMD to degrade mRNAs containing premature stop codons. This serine-rich protein binds to the mRNA and remains bound after nuclear export, acting as a nucleo-cytoplasmic shuttling protein.

<b>Protein Name</b> small nuclear ribonucleoprotein U1a, U2b, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender, Schistosomula, Female worms	<b>Localization</b> Nucleus (Psort)
---	---	--

<b>GO annotation</b>	
0000166	nucleotide binding
0003676	nucleic acid binding
0030529	ribonucleoprotein complex

**Description**  
Small nuclear ribonucleoproteins are RNA-protein complexes that combine with unmodified pre-mRNA and various other proteins to form a spliceosome.

<b>Protein Name</b> serine/arginine rich splicing factor, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender, Egg, Schistosomula	<b>Localization</b> Mitochondria (Psort)
---	--	---

<b>GO annotation</b>	
0000166	nucleotide binding
0003676	nucleic acid binding

**Description**  
Serine/arginine-rich splicing factors play an important role in constitutive and alternative splicing as well as during several steps of RNA metabolism.

<b>Protein Name</b> eukaryotic translation initiation factor 4e-binding protein, putative	<b>Life-cycle Phase</b> Egg	<b>Localization</b> Cytoplasm, Nucleus (Psort)
--	--------------------------------	---

<b>GO annotation</b>	
0006412	translation
0008150	biological process
0008152	metabolic process
0009058	biosynthetic process
0019538	protein metabolic process
0044238	primary metabolic process
0050789	regulation of biological process

**Description**  
This protein interacts with the eukaryotic translation initiation factor 4e, which binds to the cap structure of mRNAs during an early step of protein synthesis initiation and facilitates ribosome binding by inducing the unwinding of mRNA secondary structures. The binding protein prevents the translation factor to assemble its complex and therefore inhibits protein synthesis (UniProt Q13541).

<b>Protein Name</b> eukaryotic translation initiation factor 3 subunit, putative	<b>Life-cycle Phase</b> Female worms, Male worms	<b>Localization</b> Plasma membrane (Psort)
---	---	--

<b>GO annotation</b>	
null	

<b>Description</b> Eukaryotic initiation factor 3 is a multi-subunit complex that is required for binding of mRNA to 40S ribosomal subunits, stabilization of ternary complex binding to 40S subunits, and dissociation of 40S and 60S subunits (InterPro IPR007783).		
<b>Protein Name</b> mitochondrial ribosomal protein L24, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender, Female worms	<b>Localization</b> Mitochondria, Cytoplasm (Psort)
<b>GO annotation</b> 0006412 translation		
<b>Description</b> Mammalian mitochondrial ribosomal proteins are encoded by nuclear genes and help in protein synthesis within the mitochondria. Among different species, the proteins comprising the mitoribosome differ greatly in sequence, and sometimes in biochemical properties. Mitochondrial ribosomal protein L24 sequence analysis identified two transcript variants that encode the same protein in humans.		
<b>Protein Name</b> mitochondrial ribosomal protein S33, putative	<b>Life-cycle Phase</b> Egg, Male worms	<b>Localization</b> Nucleus (Psort)
<b>GO annotation</b> 0005840 ribosome		
<b>Description</b> Mammalian mitochondrial ribosomal proteins are encoded by nuclear genes and help in protein synthesis within the mitochondria. Among different species, the proteins comprising the mitoribosome differ greatly in sequence, and sometimes in biochemical properties.		
<b>Protein Name</b> Translation machinery-associated protein, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender	<b>Localization</b> Nucleus (Psort)
<b>GO annotation</b> null		
<b>Description</b> Involved in protein translation.		
<b>Protein Name</b> peptidyl-prolyl cis-trans isomerase-like 3, putative (ppiase) (rotamase) (cyclophilin-like protein ppi3) (cyp10l) EC 5.2.1.8	<b>Life-cycle Phase</b> Egg, Schistosomula, Female worms, Male worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> 0006139 nucleobase-containing compound metabolic process 0006464 protein modification process 0008150 biological process 0008152 metabolic process 0019538 protein metabolic process		
<b>Description</b> This enzyme acts accelerating protein folding by catalysing the cis-trans isomerisation of proline imidic peptide bonds in oligopeptides. It also has chaperone-like functions and is the major high-affinity binding proteins for the immunosuppressive drug cyclosporin A (CSA) in vertebrates. Cyclophilins are found in all prokaryotes and eukaryotes, and have been structurally conserved throughout evolution, implying their importance in cellular function. They share a common		

domain with isomerase and additional domains unique to each member of the family, which are important for selection of protein substrates and subcellular compartmentalization (InterPro IPR002130).

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
rotamase, putative EC: 5.2.1.8	Adult worms of mixed gender, Egg, Female worms	Membrane (Psort)

<b>GO annotation</b>	
0006464	protein modification process
0008150	biological process
0008152	metabolic process
0019538	protein metabolic process

**Description**  
This protein is involved in the interconversion of cis and trans isomers of peptide bonds involving proline residues.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
26S proteasome non-atpase regulatory subunit, putative	Adult worms of mixed gender, Female worms, Egg	Nucleus (Psort)

<b>GO annotation</b>	
0000502	proteasome complex

**Description**  
In eukaryotic cells, most proteins in the cytosol and nucleus are degraded via the ubiquitin-proteasome pathway. The 26S proteasome is a molecular machine built from approximately 31 different subunits, which catalyzes protein degradation.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
Ammd-like protein EC: 3.4.14.4	Female worms	Cytoplasm, Cytoskeleton (Pfam)

<b>GO annotation</b>	
0008152	metabolic process
0016787	hydrolase activity

**Description**  
Annotated as “family S9 non-peptidase homologue”, it is homologous to *Clonorchis sinensis* dipeptidyl-peptidase III. The enzyme is classified as a hydrolase acting on peptide bonds (peptidase), involved in the release of an N-terminal dipeptide from a peptide comprising four or more residues. Human dipeptidyl-peptidase III is a member of the S9B family in clan SC of the serine proteases. This cytoplasmic protein binds a single zinc ion and has post-proline dipeptidyl aminopeptidase activity. Both *Schistosoma mansoni* and *C. sinensis* enzymes have a conserved a esterase/lipase domain, the presence of which could be explained from the close relationship between lipases and peptidases.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
aspartic proteinase, putative	Schistosomula	Extracellular / ER (Psort)

<b>GO annotation</b>	
0004190	aspartic-type endopeptidase activity
0006508	proteolysis
0008233	peptidase activity



0016787 hydrolase activity		
<p><b>Description</b></p> <p>Many parasites which cause serious human or livestock diseases produce aspartic proteinases which perform critical functions for parasite viability. A <i>Schistosoma japonicum</i> aspartic proteinase was characterized and shown to be closely related to the cathepsin D, a lysosomal aspartic protease of mammals. The main function of cathepsin D is the degradation of endocytosed proteins. Aspartic proteinase activity is present in extracts of adult <i>S. japonicum</i> and <i>S. mansoni</i> and in culture media in which schistosomes were maintained and was capable of digesting hemoglobin. The schistosome aspartic proteinase may play a pivotal role in the catabolism of hemoglobin obtained from host erythrocytes.</p>		
<p><b>Protein Name</b> beta-1,4-galactosyltransferase, putative EC: 2.4.1</p>	<p><b>Life-cycle Phase</b> Female worms</p>	<p><b>Localization</b> Golgi /Cytoplasmic (Psort)</p>
<p><b>GO annotation</b></p> <p>0005975 carbohydrate metabolic process 0016740 transferase activity 0016757 transferase activity, transferring glycosyl groups</p>		
<p><b>Description</b></p> <p>Beta-1,4-galactosyltransferase is a trans-Golgi glycosyltransferase that participates in the synthesis of Gal beta1-4-GlcNAc-disaccharide (type 2 disaccharide determinants) unit of glycoconjugates. Glycosyltransferases acts in the transfer of a monosaccharide unit from an activated nucleotide sugar to another molecule. The result of a glycosyl transfer can be a carbohydrate, a glycoside, an oligo- or a polysaccharide. Some glycosyltransferases catalyse transfer to Pi or water. Glycosyl transfer can also occur to protein residues or lipids (forming glycolipids). The glycosylation of proteins can help during its folding and increase structural stability. Glycosylation may also play a role in cell-cell adhesion. Conformational changes in beta4Gal-T1 also creates the binding site for a mammary gland-specific protein, alpha-lactalbumin, which changes the acceptor specificity of the enzyme toward glucose to synthesize lactose during lactation in mammals.</p>		
<p><b>Protein Name</b> glycosyltransferase, putative EC: 2.4.1</p>	<p><b>Life-cycle Phase</b> Schistosomula</p>	<p><b>Localization</b> Membrane (Psort)</p>
<p><b>GO annotation</b></p> <p>0006464 protein modification process 0006629 lipidmetabolic process 0008150 biological process 0008152 metabolic process 0009058 biosynthetic process</p>		
<p><b>Description</b></p> <p>Glycosyltransferases acts in the transfer of a monosaccharide unit from an activated nucleotide sugar to another molecule. The result of a glycosyl transfer can be a carbohydrate, a glycoside, an oligo- or a polysaccharide. Some glycosyltransferases catalyse transfer to Pi or water. Glycosyl transfer can also occur to protein residues or lipids (forming glycolipids). The glycosylation of proteins can help during its folding and increase structural stability. Glycosylation may also play a role in cell-cell adhesion.</p>		
<p><b>Protein Name</b> methionine aminopeptidase, putative EC: 3.4.11.18</p>	<p><b>Life-cycle Phase</b> Egg</p>	<p><b>Localization</b> Cytoplasm (Psort)</p>

<b>GO annotation</b>		
0004177	aminopeptidase activity	
0006508	proteolysis	
0008233	peptidase activity	
0008235	metalloexopeptidase activity	
0009987	cellular process	
0016787	hydrolase activity	
0046872	metal ion binding	
<b>Description</b>		
This enzyme is responsible for the removal of the amino-terminal (initiator) methionine from nascent eukaryotic cytosolic and cytoplasmic prokaryotic proteins if the penultimate amino acid is small and uncharged (InterPro IPR001714).		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
o-sialoglycoprotein endopeptidase, putative EC: 3.4.24.57	Male worms	Extracellular (Psort)
<b>GO annotation</b>		
0004222	metalloendopeptidase activity	
0006508	proteolysis	
<b>Description</b>		
This enzyme catalyzes the hydrolysis of O-sialoglycoproteins. A sialoglycoprotein is a combination of sialic acid and glycoprotein, which is, itself, a combination of sugar and protein.		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
protein-l-isoaspartate o-methyltransferase, putative EC: 2.1.1.77	Adult worms of mixed gender	Cytoplasm (Psort)
<b>GO annotation</b>		
0006464	protein modification process	
0008150	biological process	
0008152	metabolic process	
<b>Description</b>		
Is an enzyme that catalyses the transfer of a methyl group from S-adenosylmethionine to the free carboxyl groups of D-aspartyl or L-isoaspartyl residues in a variety of peptides and proteins. The enzyme does not act on normal L-aspartyl residues L-isoaspartyl and D-aspartyl are the products of the spontaneous deamidation and/or isomerisation of normal L-aspartyl and L-asparaginyl residues in proteins. This enzyme plays a role in the repair and/or degradation of these damaged proteins; the enzymatic methyl esterification of the abnormal residues can lead to their conversion to normal L-aspartyl residues (InterPro IPR000682).		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
peptidase, putative	Adult worms of mixed gender, Female worms	Nucleus (Psort)
<b>GO annotation</b>		
null		
<b>Description</b>		
This enzyme is involved in protein degradation pathways, acting in the cleavage of peptide bonds.		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
ring finger, putative	Male worms	Extracellular (Psort)
<b>GO annotation</b>		

0008270	zinc ion binding		
0046872	metal ion binding		
<b>Description</b>			
RING (Really Interesting New Gene) finger domain is a protein structural domain of zinc finger type which contains a Cys <sub>3</sub> HisCys <sub>4</sub> amino acid motif which binds two zinc cations. Many proteins containing a RING finger play a key role in the ubiquitination pathway.			
<b>Protein Name</b>		<b>Life-cycle Phase</b>	<b>Localization</b>
Hybrid Protein Kinase_Other Group_WNK Family EC: 2.7.1.37		Schistosomula	Cytoplasm, Nucleus (Psort)
<b>GO annotation</b>			
0004672	protein kinase activity		
0004674	protein serine/threonine kinase activity		
0005524	ATP binding		
0006468	protein phosphorylation		
0016301	kinase activity		
0016310	phosphorylation		
0016772	transferase activity, transferring phosphorus-containing groups		
0016881	acid-amino acid ligase activity		
<b>Description</b>			
WNKs are highly conserved serine/threonine kinases that comprise a unique branch of the kinome. Members of this family are components of a signaling pathway that coordinates Cl <sup>-</sup> influx and efflux to dynamically regulate intracellular Cl <sup>-</sup> activity.			
<b>Protein Name</b>		<b>Life-cycle Phase</b>	<b>Localization</b>
protein phosphatase 2c, putative EC: 3.1.3.16		Adult worms of mixed gender	Cytoplasm (Psort)
<b>GO annotation</b>			
0000287	magnesium ion binding		
0003824	catalytic activity		
0004721	phosphoprotein phosphatase activity		
0004722	protein serine/threonine phosphatase activity		
0006470	protein dephosphorylation		
0008287	protein serine/threonine phosphatase complex		
0016787	hydrolase activity		
0030145	manganese ion binding		
0046872	metal ion binding		
<b>Description</b>			
Protein phosphatase 2C (PP2C) is one of the four major classes of mammalian serine/threonine specific protein phosphatases. Its exact physiological role is still unclear. Despite its evolutionary conservation, with isozymes in <i>C. Elegans</i> , <i>Arabidopsis thaliana</i> , <i>Leishmania</i> , among other organisms, PP2C does not seem to be evolutionary related to the main family of serine/threonine phosphatases. However, it is significantly similar to the catalytic subunit of pyruvate dehydrogenase phosphatase. Phosphatases remove a phosphate group from its substrate by hydrolysing phosphoric acid monoesters into a phosphate ion and a molecule with a free hydroxyl group. Phosphatases act in opposition to kinase/phosphorylases, which add phosphate groups to proteins. The addition of a phosphate group may activate/de-activate an enzyme or enable a protein-protein interaction to occur. Therefore phosphatases are integral to many signal transduction pathways.			

<b>Protein Name</b> ral guanine nucleotide dissociation stimulator, ralgds, putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoplasm, Nucleus (Psort)
<b>GO annotation</b> 0005085      guanyl-nucleotide exchange factor activity 0005622      intracellular 0007264      small GTPase mediated signal transduction 0051056      regulation of small GTPase mediated signal transduction		
<b>Description</b> Stimulates the dissociation of GDP from the Ras-related Ras-like protein A (RalA) and RalB GTPases which allows GTP binding and activation of the GTPases (UniProt Q12967). RalA and RalB function as the critical GTP sensors for GTP-dependent exocytosis.		
<b>Protein Name</b> rab15, 13, 10, 1, 35, 5, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender, Female worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> 0000166      nucleotide binding 0005488      binding 0005525      GTP binding 0007264      small GTPase mediated signal transduction 0015031      protein transport		
<b>Description</b> This enzyme is a member of the small GTPase superfamily (Ras superfamily), which comprises monomeric GTP-binding proteins with molecular masses usually in the range of 20-25 kDa. The small GTPase superfamily has been found in all eukaryotic organisms. Small GTPases function as molecular switches in intracellular signaling to control a wide variety of cellular functions. In general the GDP-bound form is the inactive configuration of the molecular switch, while the GTP-bound form is active. The switch is activated by the exchange of GDP for GTP, catalyzed by guanine nucleotide exchange factors, in response to a variety of upstream signal signals. Activated GTPases then interact with one or more effector proteins, leading to activation of downstream signaling pathways. GTPase activating proteins stimulate an otherwise low intrinsic GTPase activity, reverting the conformation back to the inactive GDP-bound form. The process of activation and inactivation is repeated as long as the upstream stimulatory signal is present.		
<b>Protein Name</b> rho GDP-dissociation inhibitor-related	<b>Life-cycle Phase</b> Male worms	<b>Localization</b> Cytoplasm (Psort and GO)
<b>GO annotation</b> 0005094      Rho GDP-dissociation inhibitor activity 0005737      cytoplasm		
<b>Description</b> Rho GDP-dissociation inhibitor (RhoGDI) plays an essential role in control of a variety of cellular functions through interactions with Rho family GTPases, including Rac1, Cdc42, and RhoA. RhoGDI might play a role in the development of drug resistance in cancer cells. Overexpression of RhoGDI in cancer cells increased resistance to apoptosis induction, probably due to the inactivation of Rac1 degradation, since this protein is required for apoptosis in response to cytotoxic drugs. Taken together, the data show that RhoGDI is an antiapoptotic molecule that mediates cellular resistance to chemotherapy agents.		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>

Serine/Threonine Kinase_CAMK Group_CAMKL Family_AMPK Subfamily EC: 2.7.11.1	Egg, Schistosomula, Female worms	Cytoplasm, Mitochondria (Psort)
<b>GO annotation</b>		
0000166	nucleotide binding	
0003824	catalytic activity	
0004672	protein kinase activity	
0004674	protein serine/threonine kinase activity	
0005524	ATP binding	
0006468	protein phosphorylation	
0008081	phosphoric diester hydrolase activity	
0016301	kinase activity	
0016310	phosphorylation	
0016772	transferase activity, transferring phosphorus-containing groups	
0046872	metal ion binding	
<b>Description</b>		
The role of AMPK in regulating cellular energy charge places this enzyme at a central control point in maintaining energy homeostasis. More recent evidence has shown that AMPK activity can also be regulated by physiological stimuli, independent of the energy charge of the cell, including hormones and nutrients. Once activated, AMPK-mediated phosphorylation events switch cells from active ATP consumption to active ATP production.		
<b>Protein Name</b> Serine/Threonine Kinase_CMGC Group_RCK Family_MAK Subfamily EC: 2.7.11.1	<b>Life-cycle Phase</b> Male worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b>		
0000166	nucleotide binding	
0001750	photoreceptor outer segment	
0003677	DNA binding	
0004672	protein kinase activity	
0004674	protein serine/threonine kinase activity	
0004693	cyclin-dependent protein kinase activity	
0005524	ATP binding	
0006366	transcription from RNA polymerase II promoter	
0006468	protein phosphorylation	
0007049	cell cycle	
0016301	kinase activity	
0016310	phosphorylation	
0016740	transferase activity	
0016772	transferase activity, transferring phosphorus-containing groups	
0045494	photoreceptor cell maintenance	
<b>Description</b>		
CMGC kinases include cyclin-dependent, MAP, glycogen synthase 3 and CK2-related kinases. MAK are serine/threonine protein kinases related to kinases involved in cell cycle regulation.		
<b>Protein Name</b> tbc1 domain family member, putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b>		
0005097	Rab GTPase activator activity	

0005622	intracellular		
0032313	regulation of Rab GTPase activity		
0032851	positive regulation of Rab GTPase activity		
<b>Description</b>			
This is a widespread domain present in Gyp6 and Gyp7, thereby giving rise to the notion that it performs a GTP-activator activity on Rab-like small GTPases. The mouse <i>TBC1</i> gene is presumed to have a role in regulating cell cycle control and differentiation.			
<b>Protein Name</b> glutaredoxin, grx, putative EC: 1.8.4.1		<b>Life-cycle Phase</b> Adult worms of mixed gender, Female worms, Male worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b>			
0009055	electron carrier activity		
0015035	protein disulfide oxidoreductase activity		
0045454	cell redox homeostasis		
<b>Description</b>			
Glutaredoxins are electron carriers in the glutathione-dependent synthesis of deoxyribonucleotides. Moreover, it has an important role in the antioxidant defence.			
<b>Protein Name</b> heat shock protein 70 (hsp70)-interacting protein, putative		<b>Life-cycle Phase</b> Egg	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b>			
0005488	binding		
0006950	response to stress		
<b>Description</b>			
CHIP (C-terminus of Hsp70-interacting protein) is a cochaperone and ubiquitin ligase that interacts with Hsp70 through an amino-terminal tetratricopeptide repeat (TPR) domain. Prominent among the cellular functions mediated by CHIP is protection against physiologic stress and inhibition of apoptosis.			
<b>Protein Name</b> thioredoxin m(mitochondrial)-type, putative EC: 1.11.1.15		<b>Life-cycle Phase</b> Female worms, Male worms	<b>Localization</b> Cytoplasm, Mitochondria (Psort)
<b>GO annotation</b>			
0006662	glycerol ether metabolic process		
0009055	electron carrier activity		
0015035	protein disulfide oxidoreductase activity		
0045454	cell redox homeostasis		
<b>Description</b>			
Thioredoxin is a class of small redox proteins known to be present in all organisms. Although not entirely understood, thioredoxin plays a central role in humans as antioxidants, through their response to reactive oxygen species. It has also recently been found to play a role in cell-to-cell communication. Mitochondrial thioredoxin systems are critical in protecting cells from mitochondria-dependent ROS and apoptosis.			
<b>Protein Name</b> beta-parvin-related		<b>Life-cycle Phase</b> Adult worms of mixed gender, Female worms	<b>Localization</b> Cytoskeleton, Mitochondria (Psort)
<b>GO annotation</b>			

0004252	serine-type endopeptidase activity		
0006508	proteolysis		
<b>Description</b>			
Probably plays a role in the regulation of cell adhesion and cytoskeleton organization (UniProt Q9HBI1). Beta-parvin, a focal adhesion and cytoskeleton associated protein, is a member of the parvin-multigene family. It consists of an $\alpha$ -actinin-like actin binding domain. Beta-parvin functions as a linker between important cytoskeletal proteins and is involved in many crucial cellular processes ranging from cell adhesion, spreading and proliferation, to deposition of extracellular matrix proteins.			
<b>Protein Name</b>	cofilin, actophorin, putative	<b>Life-cycle Phase</b>	<b>Localization</b>
		Adult worms of mixed gender	Nucleus (Psort)
<b>GO annotation</b>			
0003779	actin binding		
0005622	intracellular		
<b>Description</b>			
Actophorin is a member of the ADF/cofilin family that binds F-actin.			
<b>Protein Name</b>	Dynactin subunit 3 (Dynactin light chain p24), putative	<b>Life-cycle Phase</b>	<b>Localization</b>
		All	Mitochondria (Psort)
<b>GO annotation</b>			
null			
<b>Description</b>			
Dynactin binds to both microtubules and cytoplasmic dynein. It is involved in a diverse array of cellular functions, including ER-to-Golgi transport, the centripetal movement of lysosomes and endosomes, spindle formation, cytokinesis, chromosome movement, nuclear positioning, and axonogenesis.			
<b>Protein Name</b>	nuclear movement protein nudc, putative	<b>Life-cycle Phase</b>	<b>Localization</b>
		Adult worms of mixed genders, Egg, Male worms	Cytoplasm (Psort)
<b>GO annotation</b>			
null			
<b>Description</b>			
NudC is a nuclear movement protein that associates with dynein forming a biochemical complex with components of the dynein/dynactin complex. NudC plays multiple roles in mitosis and cytokinesis, in part by regulating microtubule organization.			
<b>Protein Name</b>	tektin, putative	<b>Life-cycle Phase</b>	<b>Localization</b>
		Female worms	Cytoplasm (Psort)
<b>GO annotation</b>			
0000226	microtubule cytoskeleton organization		
0005874	microtubule		
<b>Description</b>			
Tektins are cytoskeletal proteins found in cilia and flagella as structural components of outer doublet microtubules. They are also present in centrioles and basal bodies. They are polymeric in nature, and form filaments. The amino acid sequences of tektins are well conserved in nature, from			

<i>C. elegans</i> to <i>Homo sapiens</i> .		
<b>Protein Name</b> tropomodulin, putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoskeleton (GO)
<b>GO annotation</b> 0005523 tropomyosin binding 0005856 cytoskeleton		
<b>Description</b> Tropomodulin binds and caps the minus end of actin, regulating the length of actin filaments.		
<b>Protein Name</b> tropomyosin, putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> 0005523 tropomyosin binding 0005856 cytoskeleton		
<b>Description</b> Tropomyosin is an actin-binding protein that regulates actin mechanics.		
<b>Protein Name</b> Enolase (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase), putative	<b>Life-cycle Phase</b> Adult worms of mixed gender, Egg, Female worms, Male worms	<b>Localization</b> Cytoplasm (GO and Psort)
<b>GO annotation</b> 0000015 phosphopyruvate hydratase complex 0000287 magnesium ion binding 0004634 phosphopyruvate hydratase activity 0005737 cytoplasm 0006096 glycolysis 0016829 lyase activity 0046872 metal ion binding		
<b>Description</b> In the glycolysis, it catalyses the conversion of phosphoenolpyruvate in to 2-phosphoglycerate.		
<b>Protein Name</b> Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Major larval surface antigen) (P-37), putative EC: 1.2.1.12	<b>Life-cycle Phase</b> Egg	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> 0000166 nucleotide binding 0004365 glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity 0005737 cytoplasm 0006006 glucose metabolic process 0006096 glycolysis 0016491 oxidoreductase activity 0050661 NADP binding 0051287 NAD binding 0055114 oxidation-reduction process 0016620 oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor		
<b>Description</b>		



This enzyme catalyzes the sixth step of glycolysis, the breakdown of glucose yielding energy and carbon molecules. GAPDH has recently been implicated in several non-metabolic processes, including transcription activation, initiation of apoptosis and ER-Golgi vesicle shuttling.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
l-lactate dehydrogenase, putative EC: 1.1.1.27	Egg	Cytoplasm (Psort)

**GO annotation**

0000166	nucleotide binding
0003824	catalytic activity
0004459	L-lactate dehydrogenase activity
0005737	cytoplasm
0005975	carbohydrate metabolic process
0006096	glycolysis
0016491	oxidoreductase activity
0044262	cellular carbohydrate metabolic process
0055114	oxidation-reduction process
0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor

Lactate dehydrogenase is a highly conserved enzyme that catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD<sup>+</sup>. It converts pyruvate, the final product of glycolysis, to lactate when oxygen is absent or in short supply.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
ribulose-5-phosphate-3-epimerase, putative EC: 5.1.3.1	Female worm	Cytoplasm (Psort)

**GO annotation**

0003824	catalytic activity
0004750	ribulose-phosphate 3-epimerase activity
0005975	carbohydrate metabolic process
0008152	metabolic process
0016853	isomerase activity
0046872	metal ion binding

**Description**

This is the enzyme that converts D-ribulose 5-phosphate into D-xylulose 5-phosphate in the pentose phosphate pathway which generates NADPH and pentoses. The primary outcomes of this pathway are NADPH used in reductive biosynthetic reactions and building blocks for nucleotides, amino acids and nucleic acids.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
1-acylglycerol-3-phosphate o-acyltransferase, putative EC: 2.3.1.51	Male worms	Membrane (GO)

**GO annotation**

0003841	1-acylglycerol-3-phosphate O-acyltransferase activity
0008152	metabolic process
0008654	phospholipid biosynthetic process
0016020	membrane
0016740	transferase activity
0016746	transferase activity, transferring acyl groups

**Description**

This transferase participates in three metabolic pathways: glycerolipid metabolism, glycerophospholipid metabolism and ether lipid metabolism. The animal enzyme is specific for the transfer of unsaturated fatty acyl groups.  
([http://www.llamp.net/sites/default/files/pictures/recycling\\_of\\_phospholipids.svg](http://www.llamp.net/sites/default/files/pictures/recycling_of_phospholipids.svg)).

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
acetyl-CoA C-acyltransferase, putative EC: 2.3.1.16	Male worms	Mitochondria or Peroxisome, Membrane (Psort)

**GO annotation**

0003824	catalytic activity
0003985	acetyl-CoA C-acetyltransferase activity
0003988	acetyl-CoA C-acyltransferase activity
0008152	metabolic process
0008270	zinc ion binding
0016740	transferase activity
0016746	transferase activity, transferring acyl groups
0016747	transferase activity, transferring acyl groups other than amino-acyl groups
0046872	metal ion binding

**Description**

This enzyme catalyzes the final step of the beta oxidation, which is the cleavage of 3-ketoacyl CoA. Beta-oxidation is the process by which fatty acids are broken down in the mitochondria and/or peroxisomes to generate acetyl-CoA. Acetyl-CoA is the entry molecule for the Citric Acid Cycle, which is used by all aerobic organisms to generate energy from carbohydrates, fats and proteins. In addition, the cycle provides biosynthetic precursors for some amino acids and NADH, which is used in numerous biochemical reactions.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
acyl-coenzyme A binding domain containing, putative	Female worms	Cytoskeleton (Psort)

**GO annotation**

0000062	fatty-acyl-CoA binding
0005488	binding

**Description**

Acyl-coenzyme A binding proteins (ACBP) bind thiol esters of long fatty acids and coenzyme A in a one-to-one binding mode with high specificity and affinity. Acyl-CoAs are important intermediates in fatty lipid synthesis and fatty acid degradation and play a role in regulation of intermediary metabolism and gene regulation. The suggested role of ACBP is to act as a intracellular acyl-CoA transporter and pool former. ACBPs are present in a large group of eukaryotic species and several tissue-specific isoforms have been detected.  
This enzyme is closely related to ACBD6 proteins, which preferentially bind unsaturated long-chain acyl-CoAs (NCBI CDD).

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
arv1, putative	Female worms	Membrane (Psort)

**GO annotation**

0006355	regulation of transcription, DNA-dependent
0008270	zinc ion binding

**Description**

Arv1 is a transmembrane protein, with potential zinc-binding motifs, that mediates sterol (a class of lipids) homeostasis, preventing free sterol toxicity. In yeast, Arv1p is important for the delivery

of an early GPI (glycosylphosphatidylinositol) intermediate to the first mannosyltransferase of GPI synthesis in the ER lumen. It is important for the traffic of sterol in yeast and in humans. In eukaryotic cells, it may function in the sphingolipid metabolic pathway as a transporter of ceramides between the ER and Golgi (InterPro IPR007290). Sterol homeostasis in eukaryotic cells relies on the reciprocal interconversion of free sterols and sterol esters, which constitute an important storage form for fatty acids and sterols that are deposited in intracellular lipid particles (Eukaryotic sterol homeostasis: Sterol ester hydrolases in *Saccharomyces cerevisiae*).

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
serine palmitoyltransferase I, putative EC: 2.3.1.50	Egg	Cytoplasm, Peroxisome (Psort)

#### **GO annotation**

0003824	catalytic activity
0009058	biosynthetic process
0016740	transferase activity
0016769	transferase activity, transferring nitrogenous groups
0030170	pyridoxal phosphate binding

#### **Description**

This enzyme catalyzes a key step in the biosynthesis of sphingosine which is a precursor of many other sphingolipids. Ceramides, sphingomyelins, glycosphingolipids and gangliosides are the main types. Sphingolipids are commonly believed to protect the cell surface against harmful environmental factors by forming a mechanically stable and chemically resistant outer leaflet of the lipid bilayer. Certain complex glycosphingolipids were found to be involved in specific functions, such as signal transmission and cell recognition. Cell recognition depends mainly on the physical properties of the sphingolipids, whereas signaling involves specific interactions of the glycan structures of glycosphingolipids with similar lipids present on neighboring cells or with proteins. Recently, relatively simple sphingolipid metabolites have been shown to be important mediators in the signaling cascades involved in apoptosis, proliferation and stress. Ceramide-based lipids self-aggregate in cell membranes and form microdomains known as lipid rafts, which were proposed to sort membrane proteins along the cellular pathways of membrane transport. Sphingolipids synthesis begins in the ER and is completed in the Golgi.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
3-dehydroecdysone 3alpha-reductase-related	Adult worms of mixed gender, Female worms, Male worms	Membrane (Psort)

#### **GO annotation**

0000166	nucleotide binding
0004316	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
0008152	metabolic process
0016491	oxidoreductase activity
0055114	oxidation-reduction process

#### **Description**

One route of ecdysteroids inactivation in insects involves the conversion of ecdysone into 3-dehydroecdysone, followed by its reduction to 3-epicdysone, catalyzed by 3-dehydroecdysone 3 alpha reductase. This enzyme is a novel eukaryotic member of the dehydrogenase/reductases superfamily. Ecdysteroids have been detected in trematode parasites and in several species of gastropod snails. There are potential roles of these and other hormones as regulators of host/parasite interactions between *Schistosoma mansoni* and *Biomphalaria glabrata*. Beta-ecdysterone was found to be effective in stimulating host location activities in *S. mansoni* miracidia and in enhancing growth and egg production in *B. glabrata*. However, plant extracts rich

in phytoecdysteroids were not attractive to miracidia and did not affect growth and egg production in snails. The potential role of ecdysteroids in the process of parasitism and snail physiological processes is discussed. Plants rich in phytoecdysteroids do not appear to be of value in process which may interfere with the host parasite interactions in schistosomes (Regulation of ecdysteroid signalling: molecular cloning, characterization and expression of 3-dehydroecdysone 3 alpha-reductase, a novel eukaryotic member of the short-chain dehydrogenases/reductases superfamily from the cotton leafworm, *Spodoptera littoralis* and Ecdysteroids as regulators of host and parasite interactions: a study of interrelationships between *Schistosoma mansoni* and the host snail, *Biomphalaria glabrata*).

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
isopentenyl-diphosphate delta isomerase, putative EC: 5.3.3.2	Adult worms of mixed gender, Female worms	Cytoplasm, Nucleus (Psort and GO)

**GO annotation**

0000287	magnesium ion binding
0004452	isopentenyl-diphosphate delta-isomerase activity
0004743	pyruvate kinase activity
0005634	nucleus
0006096	glycolysis
0006352	transcription initiation, DNA-dependent
0008299	isoprenoid biosynthetic process
0016787	hydrolase activity
0016853	isomerase activity
0030955	potassium ion binding

**Description**

This enzyme catalyzes the conversion of the relatively un-reactive isopentenyl pyrophosphate (IPP) to the more-reactive electrophile dimethylallyl pyrophosphate. This reaction is a key step in the biosynthesis of isoprenoids through the highly conserved mevalonate pathway. Isoprenoids play widely varying roles in the physiological processes of plants and animals. Their function includes pigments, fragrances, vitamins and precursors of sex hormones.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
sterol reductase-related EC: 1.3.1.71	Egg	Membrane (Psort)

**GO annotation**

0004872	receptor activity
0016020	membrane

**Description**

This oxidoreductase acts on the CH-CH group of donor with NAD<sup>+</sup> or NADP<sup>+</sup> as acceptor. This enzyme participates in the biosynthesis of steroids such as cholesterol, estradiol and testosterone.

<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
atpase inhibitor, putative	All	Mitochondria (GO) / Cytoplasm (Psort)

**GO annotation**

0004857	enzyme inhibitor activity
0005739	mitochondrion
0043086	negative regulation of catalytic activity
0045980	negative regulation of nucleotide metabolic process

**Description**

It is thought to inhibit ATP synthesis by preventing the release of ATP. The ATP synthase inhibitor prevents the enzyme from switching to ATP hydrolysis during collapse of the electrochemical gradient, for example during oxygen deprivation. The inhibitor has two oligomeric states, dimer (the active state) and tetramer. At low pH, the inhibitor forms a dimer via antiparallel coiled coil interactions between the C-terminal regions of two monomers. At high pH, the inhibitor forms tetramers and higher oligomers by coiled coil interactions involving the N terminus and inhibitory region, thus preventing the inhibitory activity (InterPro IPR007648).

<b>Protein Name</b> ATP synthase delta chain, mitochondrial, putative EC: 3.6.3.14	<b>Life-cycle Phase</b> Egg	<b>Localization</b> Mitochondria
---	--------------------------------	-------------------------------------

<b>GO annotation</b>		
0015986	ATP synthesis coupled proton transport	
0045261	proton-transporting ATP synthase complex, catalytic core F(1)	
0046933	hydrogen ion transporting ATP synthase activity, rotational mechanism	
0046961	proton-transporting ATPase activity, rotational mechanism	

**Description**  
Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation.

<b>Protein Name</b> cytochrome C oxidase copper chaperone, putative	<b>Life-cycle Phase</b> All	<b>Localization</b> Mitochondria, Extracellular (Psort)
---	--------------------------------	---

<b>GO annotation</b>		
0004129	cytochrome-c oxidase activity	

**Description**  
This metallochaperone is essential for the assembly of functional cytochrome c oxidase (CCO) and for delivery of copper ions to the mitochondrion for insertion into the enzyme (InterPro IPR007745). Cytochrome c oxidase is a large transmembrane protein complex found in bacteria and, in eukaryotic mitochondria that plays a fundamental role in the electron transport chain. Therefore, the metallochaperone is involved with the generation of a proton gradient that culminates with ATP formation.

<b>Protein Name</b> ectonucleotide pyrophosphatase/phosphodiesterase, putative EC: 3.1.4	<b>Life-cycle Phase</b> Adult worms of mixed gender	<b>Localization</b> Membrane (Psort)
---	---	---

<b>GO annotation</b>		
0003824	catalytic activity	
0008152	metabolic process	

**Description**  
Involved primarily in ATP hydrolysis at the plasma membrane. Plays a role in regulating pyrophosphate levels, and functions in bone mineralization and soft tissue calcification (Uniprot P22413).

<b>Protein Name</b> hexaprenyldihydroxybenzoate methyltransferase, putative EC: 2.1.1.114	<b>Life-cycle Phase</b> Male worms	<b>Localization</b> Mitochondria (Psort)
--	---------------------------------------	---

<b>GO annotation</b>		
----------------------	--	--

0006744	ubiquinone biosynthetic process		
0008152	metabolic process		
0008168	methyltransferase activity		
0008425	2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity		
0016740	transferase activity		
0032259	methylation		
<b>Description</b>			
This transferase participates in the biosynthesis of ubiquinone (coenzyme Q10). This oil-soluble, vitamin-like substance is present in most eukaryotic cells, primarily in the mitochondria. It is a component of the electron transport chain and participates in aerobic cellular respiration, generating energy in the form of ATP.			
<b>Protein Name</b> ubiquinone biosynthesis protein, putative		<b>Life-cycle Phase</b> All	<b>Localization</b> Cytoplasm, Extracellular (Psort) / Mitochondria (GO)
<b>GO annotation</b>			
0005739	mitochondrion		
0005743	mitochondrial inner membrane		
0006744	ubiquinone biosynthetic process		
0016020	membrane		
<b>Description</b>			
This protein is required for the biosynthesis of ubiquinone (InterPro IPR012762). This oil-soluble, vitamin-like substance is present in most eukaryotic cells, primarily in the mitochondria. It is a component of the electron transport chain and participates in aerobic cellular respiration, generating energy in the form of ATP.			
<b>Protein Name</b> ubiquinol-cytochrome C reductase complex ubiquinol binding protein (UbcRBP)		<b>Life-cycle Phase</b> Adult worms of mixed gender, Egg, Female worms, Male worms	<b>Localization</b> Mitochondria
<b>GO annotation</b>			
null			
<b>Description</b>			
This protein is involved in redox-linked proton pumping as a constituent of the mitochondrial respiratory chain in the inner mitochondrial membrane.			
<b>Protein Name</b> molybdopterin-binding, putative EC: 2.7.7.2		<b>Life-cycle Phase</b> Egg	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b>			
0003824	catalytic activity		
0003919	FMN adenylyltransferase activity		
0006747	FAD biosynthetic process		
0006777	Mo-molybdopterin cofactor biosynthetic process		
0008152	metabolic process		
0016740	transferase activity		
0016779	nucleotidyltransferase activity		
<b>Description</b>			
Eukaryotic and prokaryotic molybdoenzymes require a molybdopterin cofactor (MoCF) for their activity. The biosynthesis of this cofactor involves a complex multistep enzymatic pathway. One of			

the eukaryotic proteins involved in this pathway is highly similar to a microtubule-associated protein (InterPro IPR001453). Molybdenum is essential for most biological systems as it is required by enzymes catalyzing reactions in carbon, sulfur and nitrogen metabolism. ([www.imoa.info/HSE/environmental\\_data/biology/reviews\\_of\\_molybdoenzymes.php](http://www.imoa.info/HSE/environmental_data/biology/reviews_of_molybdoenzymes.php)).

<b>Protein Name</b> uroporphyrinogen-III synthase, putative EC: 4.2.1.75	<b>Life-cycle Phase</b>	<b>Localization</b> Cytoplasm (GO) / Extracellular (Psort)
--	-------------------------	--

<b>GO annotation</b>	
0004852	uroporphyrinogen-III synthase activity
0016829	lyase activity
0033014	tetrapyrrole biosynthetic process

**Description**  
Uroporphyrinogen III synthase is an enzyme involved in the metabolism of porphyrin. This enzyme catalyses the conversion of hydroxymethyl bilane into uroporphyrinogen III through the inversion of the final pyrrole unit (ring D) of the linear tetrapyrrole molecule, linking it to the first pyrrole unit (ring A), thereby generating a large macrocyclic structure, uroporphyrinogen III. Porphyrins are a class of organic compounds including heme. Earlier studies from the Padmanaban group had shown that the malarial parasite synthesizes heme *de novo*, despite acquiring heme from the host red cell hemoglobin in the intraerythrocytic stage. In *Plasmodium falciparum*, the uroporphyrinogen-III synthase is not present and its activity is probably supplied by porphobilinogen deaminase, which may have a dual function (porphobilinogen → hydroxymethyl bilane → uroporphyrinogen III) (Unique Properties of *Plasmodium falciparum* Porphobilinogen Deaminase).

<b>Protein Name</b> 4-hydroxyphenylpyruvate dioxygenase, putative EC: 1.13.11.27	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoplasm (Psort)
--	---	--

<b>GO annotation</b>	
0003868	4-hydroxyphenylpyruvate dioxygenase activity
0009072	aromatic amino acid family metabolic process
0055114	oxidation-reduction process
0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen
0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen

**Description**  
This Fe-containing enzyme catalyses the second reaction in the catabolism of tyrosine.

<b>Protein Name</b> valacyclovir hydrolase, putative	<b>Life-cycle Phase</b> Adult worms of mixed gender	<b>Localization</b> Mitochondria (Psort)
---	---	---

<b>GO annotation</b> null	
------------------------------	--

**Description**  
Serine hydrolase that catalyzes the hydrolytic activation of amino acid ester prodrugs of nucleoside analogs such as valacyclovir and valganciclovir. Activates valacyclovir to acyclovir. May play a role in detoxification processes. It is a specific alpha-amino acid ester hydrolase that prefers small, hydrophobic, and aromatic side chains and does not have a stringent requirement for the leaving group other than preferring a primary alcohol (Uniprot Q86WA6).

<b>Protein Name</b> fatty acid binding protein, putative	<b>Life-cycle Phase</b> Adult worms of mixed	<b>Localization</b> Cytoplasm (Psort and
---	---	---

	gender	GO)
<b>GO annotation</b>		
0005215	transporter activity	
0005488	binding	
0005737	cytoplasm	
0006810	transport	
0008289	lipid binding	
<b>Description</b>		
<p>The fatty-acid-binding proteins are a family of carrier proteins for fatty acids and other lipophilic substances such as eicosanoids and retinoids. These proteins are thought to facilitate the transfer of fatty acids between extra- and intracellular membranes. Some family members are also believed to transport lipophilic molecules from outer cell membrane to certain intracellular receptors. Levels of fatty-acid-binding protein have been shown to decline with ageing in the mouse brain, possibly contributing to age-associated decline in synaptic.</p>		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
mitochondrial carrier protein-related	Egg	Mitochondria, Cytoplasm (Psort and GO)
<b>GO annotation</b>		
0005488	binding	
0005743	mitochondrial inner membrane	
0006810	transport	
0016020	membrane	
0016021	integral to membrane	
0055085	transmembrane transport	
<b>Description</b>		
<p>Mitochondrial carriers are proteins from the solute carrier family which transfers molecules across the membranes of the mitochondria. These proteins are usually involved in energy transfer. Common substrates include: citrate, ornithine, phosphate, glutamate, ADP and ATP.</p>		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
nuclear transport factor, putative	Egg	Nucleus, Extracellular (Psort)
<b>GO annotation</b>		
0005622	intracellular	
0006810	transport	
<b>Description</b>		
<p>These proteins are involved in the transport of other proteins to the cellular nucleus. The transport of macromolecules between the nucleus and the cytoplasm is crucial for controlling cellular processes such as gene expression, cell division, and signal transduction. (Mapping Interactions between Nuclear Transport Factors in Living Cells Reveals Pathways through the Nuclear Pore Complex)</p>		
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>
peripheral-type benzodiazepine receptor, putative	Adult worms of mixed gender, Egg, Female worms	Extracellular (Psort) / Membrane (GO)
<b>GO annotation</b>		
0004872	receptor activity	
0016021	integral to membrane	



<b>Description</b> The peripheral-type benzodiazepine receptor or recognition site (PBR) is a widely distributed protein, located mainly in the outer mitochondrial membrane. The protein binds to high-affinity drug ligands and cholesterol. Many functions are associated directly or indirectly with the protein, including the regulation of cholesterol transport and the synthesis of steroid hormones, porphyrin transport and heme synthesis, apoptosis, cell proliferation, anion transport, regulation of mitochondrial functions and immunomodulation.		
<b>Protein Name</b> solute carrier family 35 member C1, putative	<b>Life-cycle Phase</b> Male worms	<b>Localization</b> Cytoplasm (Psort) / Membrane (GO)
<b>GO annotation</b> 0016020 membrane 0016021 integral to membrane		
<b>Description</b> In humans, this gene encodes a GDP-fucose transporter that is found in the Golgi apparatus. Mutations in this gene result in congenital disorder of glycosylation type IIc.		
<b>Protein Name</b> B-cell receptor-associated protein-like protein	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> ER-Golgi, Extracellular (Psort) / ER, Membrane (GO)
<b>GO annotation</b> 0004872 receptor activity 0005783 endoplasmic reticulum 0006886 intracellular protein transport 0016021 integral to membrane		
<b>Description</b> Its human homologous is believed to have a role in the anterograde transport of membrane proteins from the endoplasmic reticulum to the Golgi. It may also be involved in CASP8-mediated apoptosis, since it cleaved by CASP8 and other caspases (UniProt P51572).		
<b>Protein Name</b> dynamin-associated protein, putative	<b>Life-cycle Phase</b> All	<b>Localization</b> Cytoplasm (Psort) / Membrane (GO)
<b>GO annotation</b> 0004872 receptor activity 0005085 guanyl-nucleotide exchange factor activity 0005089 Rho guanyl-nucleotide exchange factor activity 0005509 calcium ion binding 0005515 protein binding 0005622 intracellular 0006897 endocytosis 0007264 small GTPase mediated signal transduction 0012505 endomembrane system 0016020 membrane 0016021 integral to membrane 0019209 kinase activator activity 0019717 synaptosome 0030027 lamellipodium 0030054 cell junction 0030139 endocytic vesicle		

0035023	regulation of Rho protein signal transduction		
0035556	intracellular signal transduction		
0042327	positive regulation of phosphorylation		
0042995	cell projection		
0043524	negative regulation of neuron apoptosis		
0045202	synapse		
0051897	positive regulation of protein kinase B signaling cascade		
<b>Description</b>			
This protein is associated to Dynamin, which is an eukaryotic GTPase responsible for endocytosis.			
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>	
golgi snare bet1-related	Female worms	Plasma membrane (Psort)	
<b>GO annotation</b>			
null			
<b>Description</b>			
The Bet1 protein is involved in the vesicular transport between the ER and the Golgi complex.			
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>	
rabb and c, putative	Schistosomula	Cytoplasm (Psort)	
<b>GO annotation</b>			
0000166	nucleotide binding		
0005525	GTP binding		
0007264	small GTPase mediated signal transduction		
0015031	protein transport		
<b>Description</b>			
The Rab family is a member of the Ras superfamily of monomeric G proteins. Rab GTPases regulate many steps of membrane traffic, including vesicle formation, vesicle movement along actin and tubulin networks, and membrane fusion. These processes make up the route through which cell surface proteins are trafficked from the Golgi to the plasma membrane and are recycled. Surface protein recycling returns proteins to the surface whose function involves carrying another protein or substance inside the cell, such as the transferrin receptor, or serves as a means of regulating the number of a certain type of protein molecules on the surface.			
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>	
Small VCP/p97-interacting protein, putative	Egg, Schistosomula, Female worm, Male worm	Cytoplasm (Psort)	
<b>GO annotation</b>			
null			
<b>Description</b>			
VCP/p97 is involved in a variety of cellular processes, including membrane fusion and ubiquitin-dependent protein degradation. Small VCP/p97-interacting protein is a novel VCP/p97 adaptor whose function is related to the integrity of the endoplasmic reticulum (SVIP Is a Novel VCP/p97-interacting Protein Whose Expression Causes Cell Vacuolation). Overexpression causes the formation of large vacuoles that seemed to be derived from the endoplasmic reticulum.			
<b>Protein Name</b>	<b>Life-cycle Phase</b>	<b>Localization</b>	
vacuolar ATP synthase proteolipid subunit 1, 2, 3,, putative	Egg	Membrane (GO) / Membrane (Psort)	
<b>GO annotation</b>			

0006810	transport		
0006811	ion transport		
0015078	hydrogen ion transmembrane transporter activity		
0015991	ATP hydrolysis coupled proton transport		
0015992	proton transport		
0016020	membrane		
0016021	integral to membrane		
0033177	proton-transporting two-sector ATPase complex, proton-transporting domain		
0033179	proton-transporting V-type ATPase, V <sub>0</sub> domain		
<b>Description</b>			
Vacuolar-type ATPases (V-ATPases) are ubiquitous proton pumps that occur in the endomembrane system of all eukaryotic cells and in plasma membranes of many animal cells. They have various functions including the energization of transport processes across membranes and the regulation of the intracellular or intraorganellar pH (Beyenbach and Wieczorek, 2006; Forgac, 2007). V-ATPases are heteromultimeric enzymes consisting of a cytosolically oriented catalytic V <sub>1</sub> complex, and a membrane bound proton translocating V <sub>0</sub> complex (Inhibitors of V-ATPases: old and new players).			
<b>Protein Name</b>		<b>Life-cycle Phase</b>	<b>Localization</b>
amyloid beta A4 protein related		Egg	Cytoplasmic (Psort)
<b>GO annotation</b>			
null			
<b>Description</b>			
Beta amyloid proteins are potentially involved with several physiological processes within the cell, such as: the activation of kinases, protection against oxidative stress, regulation of cholesterol transport, transcription factor and anti-microbial activity.			
<b>Protein Name</b>		<b>Life-cycle Phase</b>	<b>Localization</b>
autophagy protein 16-like		Egg, Schistosomula, Female worm, Male worm	Cytoplasm (Psort)
<b>GO annotation</b>			
null			
<b>Description</b>			
This protein is part of a large complex that plays a role in autophagy, the major process by which intracellular components are targeted to lysosomes for degradation. Several transcript variants encoding different isoforms have been found for this gene.			
<b>Protein Name</b>		<b>Life-cycle Phase</b>	<b>Localization</b>
metal dependent hydrolase - related		Adult worms of mixed gender	Cytoskeleton (Psort)
<b>GO annotation</b>			
0016787	hydrolase activity		
<b>Description</b>			
This is a protein with hydrolase activity, that acts on the catalysis of chemical bond hydrolysis.			
<b>Protein Name</b>		<b>Life-cycle Phase</b>	<b>Localization</b>
Protein C14orf153 precursor, putative		Female worms	Cytoplasm (Psort)
<b>GO annotation</b>			
null			
<b>Description</b>			

This proteins contains an uncharacterized conserved domain (DUF2315); This is a family of small conserved proteins found from worms to humans. The function is not known (Pfam).		
<b>Protein Name</b> Sj-Ts1 , putative	<b>Life-cycle Phase</b>	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> null		
<b>Description</b> Sj-Ts1 is a predicted <i>Schistosoma japonicum</i> protein that contains one transmembrane helix, one N-myristoylation site, two phosphorylation sites for protein kinase C and one for tyrosine kinase.		
<b>Protein Name</b> tetraspanin D76 , putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Plasma membrane (Psort and GO)
<b>GO annotation</b> 0016020      membrane 0016021      integral to membrane		
<b>Description</b> This protein is involved in mechanisms that act during hemocyte phagocytosis, controlling the abrupt transition of hemocytes from resting, nonadherent cells to activated, adherent cells during these cell-mediated responses.		
<b>Protein Name</b> WD repeat protein SL1-17, putative	<b>Life-cycle Phase</b> Female worms	<b>Localization</b> Cytoplasm (Psort)
<b>GO annotation</b> null		
<b>Description</b> WD-repeat proteins are involved in a variety of cellular functions, from signal transduction and transcription regulation to cell cycle control, autophagy and apoptosis. One common function of all WD-repeat proteins seems to be the coordination of multi-protein complex assemblies, where the repeating units serve as a rigid scaffold for protein interactions.		