

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 Mourão¹, Mainá Bitar², Francisco Pereira Lobo³, Ana Paula Peconick⁴, Priscila Grynberg², Francisco Prosdócimi⁵, Michael Waisberg⁶, Gustavo Coutinho Cerqueira⁷, Andréa Mara Macedo², Carlos Renato Machado², Timothy Yoshino⁸, Glória Regina Franco^{2/+}

¹Grupo de Genômica e Biologia Computacional, Centro de Pesquisas René Rachou-Fiocruz, Belo Horizonte, MG, Brasil ²Laboratório de Genética Bioquímica, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil ³Laboratório de Bioinformática Aplicada, Embrapa Informática Aplicada, Campinas, SP, Brasil ⁴Setor de Medicina Veterinária Preventiva, Universidade Federal de Lavras, Lavras, MG, Brasil ⁵Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil ⁶Department of Pathology, University of Virginia, Charlottesville, VA, USA ⁷Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA ⁸Department 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: gfrancofmg@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)+ 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'-CGGTATTTTCAGTCGGTGTTCACCTT9V3' - 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 β -agarase according to Franco et al. (1995) and (ii) precipitation with 15% polyethyleneglycol 8000 (to obtain fragments ≥ 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 α 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 ~ 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-iTTM RNAi Designer tools, available from rnaidesigner.invitrogen.com/rnaiexpress (Life Technologies, Carlsbad, CA, USA). The generated sequences were synthesised by Life Technologies using the StealthTM 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 μ 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 μ 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 μ 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 μ L of cDNA, 12.5 μ L of SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) and 10 μ 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'-GTCACGGTTTACTCTTGT-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* α -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/), 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 5×10^{-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 1×10^{-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

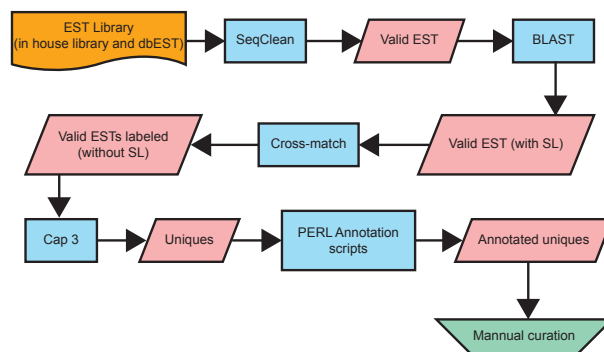


Fig. 1: fluxogram illustrating each step of processing and curation performed with sequences from the dbEST and the spliced leader (SL)-enriched Expressed Sequence Tags library. The fluxogram contains both the names of all programs and methods (blue shaded boxes) used in this study for sequence cleaning,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

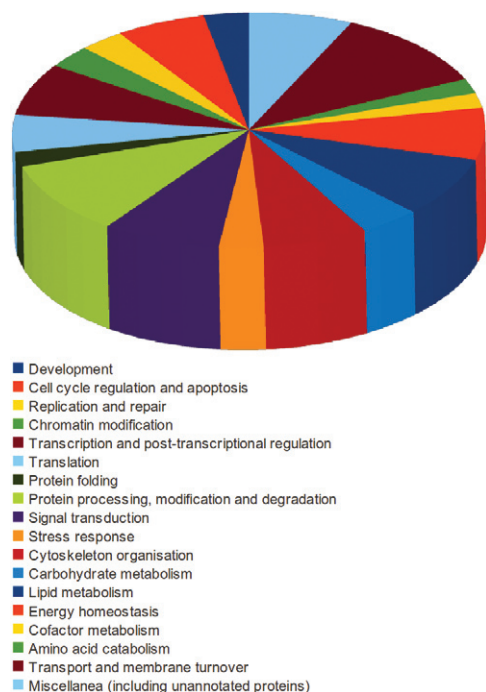


Fig. 2: pie chart illustrating the distribution of annotated proteins in different classes after manual curation and classification.

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^{2+} 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-

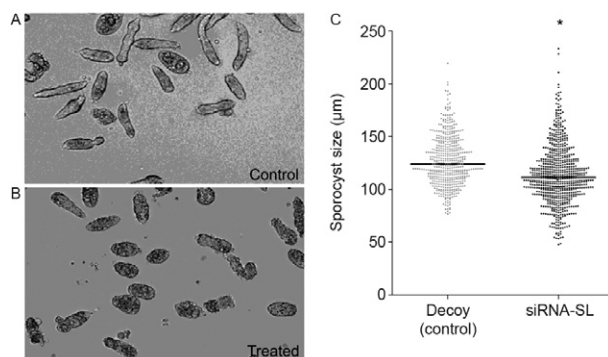


Fig. 3: in vitro cultured *Schistosoma mansoni* larvae seven days post-double stranded RNA treatments. A, B: brightfield photomicrographs of in vitro cultured *S. mansoni* sporocysts after seven days of treatments with spliced leader (SL)-small interfering (siRNA) (A) compared to the control decoy-siRNA (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 $p < 0.0001$.

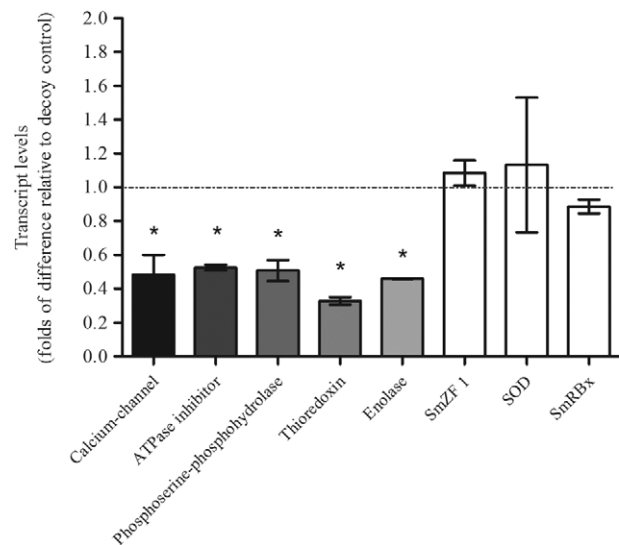


Fig. 4: histogram depicting the relative transcript levels of small interfering (siRNA)-treated sporocysts after seven days of exposure compared to the decoy-siRNA control. For each transcript tested, data is represented as mean fold-differences (± 2 standard error) relative to the decoy-small RNA control (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 $\Delta\Delta\text{Ct}$ 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.

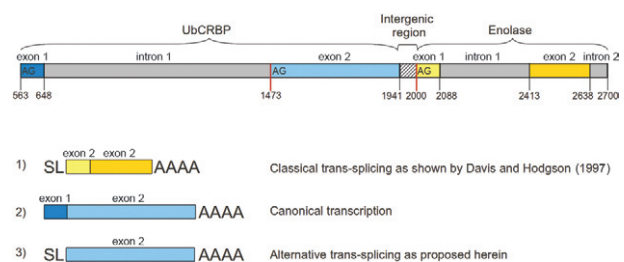


Fig. 5: representative scheme of an alternative spliced leader (SL) trans-splicing, as was observed in the case of ubiquinol-cytochrome C reductase complex ubiquinol binding protein (UbCRBP). While in normal transcription all exons are present in the final transcript, in alternative SL trans-splicing, 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.
- Beranov-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 Parasit* 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.
- Knudsen GM, Medzhradszky 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.
- 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.