

Developmental Regulation of Genes Encoding Universal Stress Proteins in *Schistosoma mansoni*

Raphael D. Isokpehi^{1,2}, Ousman Mahmud^{1,2}, Andreas N. Mbah¹, Shaneka S. Simmons^{1,2}, Livia Avelar^{3,4}, Rajendram V. Rajnarayanan⁵, Udensi K. Udensi¹, Wellington K. Ayensu^{1,2}, Hari H. Cohly^{1,2}, Shyretha D. Brown^{1,2}, Centdrika R. Dates^{1,2}, Sonya D. Hentz², Shawntae J. Hughes^{1,2}, Dominique R. Smith-McInnis^{1,2}, Carvey O. Patterson², Jennifer N. Sims^{1,2}, Kelisha T. Turner², Baraka S. Williams^{1,2}, Matilda O. Johnson⁶, Taiwo Adubi^{7,8}, Judith V. Mbuh^{7,9}, Chiaka I. Anumudu^{7,10}, Grace O. Adeoye^{7,8}, Bolaji N. Thomas¹¹, Oyekanmi Nashiru¹² and Guilherme Oliveira³

¹Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson, Mississippi, USA. ²Department of Biology, Jackson State University, Jackson, Mississippi, USA. ³Genetics and Computational Biology Group, Centro de Pesquisas René Rachou/FIOCRUZ-MG, Av. Augusto de Lima, 1715, Barro Preto, CEP 30190-002, Belo Horizonte-MG, Brazil. ⁴Universidade Federal de Minas Gerais/Departamento de Genética, Av. Antônio Carlos, 6627, Pampulha, CEP 31270-901, Belo Horizonte-MG, Brazil. ⁵Department of Pharmacology and Toxicology, State University of New York at Buffalo, Buffalo, New York, USA. ⁶Department of Environmental Sciences, University of South Africa, Pretoria, South Africa. ⁷Participant, International Center for Genetic Engineering and Biotechnology/National Biotechnology Development Agency/West African Biotechnology Workshop Series, Ibadan Nigeria. ⁸Department of Zoology, University of Lagos, Akoka, Yaba, Lagos, Nigeria. ⁹Department of Life Sciences, University of Buea, Buea, Cameroon. ¹⁰Department of Zoology, University of Ibadan, Ibadan, Nigeria. ¹¹Department of Biological and Medical Sciences, Rochester Institute of Technology, Rochester, New York, USA. ¹²Department of Molecular Biology and Bioinformatics, National Biotechnology Development Agency (NABDA) Abuja, Nigeria and West African Biotechnology Workshop Series (WABWS). Corresponding author email: raphael.isokpehi@jsums.edu

Abstract: The draft nuclear genome sequence of the snail-transmitted, dimorphic, parasitic, platyhelminth *Schistosoma mansoni* revealed eight genes encoding proteins that contain the Universal Stress Protein (USP) domain. *Schistosoma mansoni* is a causative agent of human schistosomiasis, a severe and debilitating Neglected Tropical Disease (NTD) of poverty, which is endemic in at least 76 countries. The availability of the genome sequences of *Schistosoma* species presents opportunities for bioinformatics and genomics analyses of associated gene families that could be targets for understanding schistosomiasis ecology, intervention, prevention and control. Proteins with the USP domain are known to provide bacteria, archaea, fungi, protists and plants with the ability to respond to diverse environmental stresses. In this research investigation, the functional annotations of the USP genes and predicted nucleotide and protein sequences were initially verified. Subsequently, sequence clusters and distinctive features of the sequences were determined. A total of twelve ligand binding sites were predicted based on alignment to the ATP-binding universal stress protein from *Methanocaldococcus jannaschii*. In addition, six USP sequences showed the presence of ATP-binding motif residues indicating that they may be regulated by ATP. Public domain gene

Gene Regulation and Systems Biology 2011:5 61–74

doi: [10.4137/GRSB.S7491](https://doi.org/10.4137/GRSB.S7491)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



expression data and RT-PCR assays confirmed that all the *S. mansoni* USP genes were transcribed in at least one of the developmental life cycle stages of the helminth. Six of these genes were up-regulated in the miracidium, a free-swimming stage that is critical for transmission to the snail intermediate host. It is possible that during the intra-snail stages, *S. mansoni* gene transcripts for universal stress proteins are low abundant and are induced to perform specialized functions triggered by environmental stressors such as oxidative stress due to hydrogen peroxide that is present in the snail hemocytes. This report serves to catalyze the formation of a network of researchers to understand the function and regulation of the universal stress proteins encoded in genomes of schistosomes and their snail intermediate hosts.

Keywords: expressed sequence tags, gene regulation, gene function, protein domains, *Schistosoma*, serial analysis of gene expression, sequence analysis, universal stress proteins

Introduction

The 363Mb draft nuclear genome sequence of the parasitic flatworm *Schistosoma mansoni* (blood fluke) has revealed at least 11,809 putative genes encoding 13,197 transcripts.^{1,2} Eight *S. mansoni* genes were predicted to encode proteins that contain the Universal Stress Protein (USP) domain (Pfam Accession: PF00582). Proteins with the USP domain are known to provide bacteria, archaea, fungi, protists and plants with the ability to respond to a plethora of environmental stressors such as nutrient starvation, drought, high salinity, extreme temperatures and exposure to toxic chemicals.^{3–10}

Schistosoma species mainly *S. japonicum*, *S. mansoni*, and *S. hematobium* are transmitted by freshwater snails and are the main cause of human schistosomiasis, a Neglected Tropical Disease (NTD) of poverty endemic in at least 76 tropical or sub-tropical countries and territories.¹ Praziquantel is the only widely used and effective drug for schistosomiasis treatment. However, the effectiveness of praziquantel is limited due to its inability to kill schistosomes 2 to 4 weeks post-infection.¹¹ Schistosomiasis is estimated to infect 200 million people worldwide and is second only to malaria in public health significance.¹² The mortality rate of *S. mansoni* in Africa was estimated at 130,000 per year.¹³ In the Americas, Brazil is the most affected country with 4 million to 6 million infections among 25 million people living in endemic areas.¹⁴ The urinary form of schistosomiasis is associated with increased risk of bladder cancer.¹⁵ The availability of the genome sequences of *Schistosoma* species^{16,17} present opportunities for bioinformatics and genomics analyses of its associated gene families that could be used as targets for understanding schistosomiasis ecology, intervention, prevention and control.

The *Schistosoma mansoni* USP genes (Smp_001000, Smp_001010, Smp_031300, Smp_043120, Smp_076400, Smp_097930,

Smp_136870 and Smp_136890) have been analyzed in the context of a phylogenomic analysis of metazoan USPs.¹⁸ Further, the functional annotations integrated into the relational database SchistoDB (<http://www.schistodb.net>) include gene expression data documented on its developmental stages from high-throughput Expressed Sequences Tags (ESTs), Serial Analysis of Gene Expression (SAGE) and microarray methods. These gene expression data present rich genomic resources to elucidate the developmental regulation of *S. mansoni* USP genes. The complex life cycle of schistosomes involves multiple hosts, hostile environments, free-living stages as well as the dimorphic adult worms, which reside in the human host.^{19,20} This life cycle starts with the eggs found in urine and feces, which are deposited in fresh water. The eggs will then hatch to produce miracidia, which penetrates the snail tissue and transforms into sporocysts. The sporocyst in the snail develops into cercariae, which are released into water. Cercariae penetrate the human skin and lose their tail during the process and transform to schistomulae. Schistosomulum enters the blood stream and migrates to the liver, bladder and intestines to mature into adults. The adults mate and produce eggs, which are released from the body through urine and feces to restart the life cycle.²⁰ In addition to developmental stages in the human host, interventions directed to life cycle stages in the environment as well as in the freshwater snail intermediate hosts could provide additional methods to control human schistosomiasis.

The universal stress protein domain can occur as a single domain in small USP proteins (~14–15 kDa), as a tandem domain in larger USP proteins (~30 kDa), or as one or two USP domains together with other functional domains.^{4,5} Examples of functional domains commonly fused to USP domains include antiporter, voltage channels, amino acid permeases, and protein kinase domains. Members of the USP family can be



categorized into two groups based on the presence of ATP binding motif in their amino acid residues.⁸ ATP-binding is a molecular mechanism to regulate the function of universal stress proteins.²¹

Technau et al²¹ identified metazoan USP genes from EST datasets from coral *Acropora millepora* and sea anemone *Nematostella vectensis* as well as related USP sequences from *S. japonicum*. More recently, Forêt et al¹⁸ conducted phylogenomic and in situ expression investigation of USP genes distributed in diverse metazoans genomes and observed (i) a unique phylogenomic pattern that reflects at least five independent losses and multiple independent expansions; (ii) the presence of the USP gene in the common metazoan ancestor; and (iii) the spatial expression of *Hydra* USP genes in the endodermal epithelium, a highly potent chemical barrier for protection against intruding microbes.

Sequences associated with ESTs, SAGE Tags and microarray probes are functional genomic tools appropriate for dissecting gene functions.²² These tools can provide expression data to facilitate forward and reverse genetics approaches including RNA Interference (RNAi)²³ and post-transcriptional regulation using microRNAs.^{24,25} The availability of the draft genome sequence of *S. mansoni* has led to the development of microarray slides based on oligonucleotides from gene sequences suitable for elucidating the developmental expression of *S. mansoni* genes. For example, there is a comprehensive microarray dataset composed of an experimental series of genes expressed during the life cycle stages using a Puerto Rican strain of *S. mansoni*.¹²

The objectives of the reported research were to (i) manually annotate the sequence of the eight *S. mansoni* USPs; (ii) identify amino acid residues that can serve as ligand binding residues including for ATP, a known regulator of USP function; (iii) use Reverse-Transcriptase PCR to assess mRNA expression of USP genes from a strain of *S. mansoni*; and (iv) use public-domain gene expression data on *S. mansoni* to determine evidence of developmental stage expression of USP genes. The research was also conducted to initiate a collaborative network of researchers to characterize the universal stress proteins of schistosomes as potential targets for diverse applications. The results from these objectives have helped to identify research priorities to guide further investigation of *S. mansoni* USP genes.

Methods

Manual annotation of genes

Genes encoding proteins that contain the universal stress protein domain in the *Schistosoma mansoni* genome were obtained from the SchistoDB (www.schistodb.net).²⁶ A query for genes annotated with the Pfam domain accession PF00582 retrieved the expected 8 genes. Each gene was manually annotated to verify automated predictions and make revisions to the amino acid and nucleotide sequences where necessary. The characteristics of these genes verified manually were categorized as follows: (i) database identifiers; (ii) presence in selected databases; (iii) protein domain coordinates; and (iv) protein sequence length. Pfam domain coordinates were generated from Pfam resource at <http://pfam.janelia.org/>.

Ligand binding residues in *Schistosoma mansoni* universal stress proteins

The presence of ATP binding amino acids is a feature for categorizing sequences containing the universal stress protein domain.²⁷ We used multiple sequence alignment generated by ClustalW (<http://www.ebi.ac.uk/clustalw>) to determine the presence of ATP-binding amino acid residues according to those defined for the USP (MJ_0577) of *Methanocaldococcus jannaschii* DSM 2661. Additionally the guide tree generated by ClustalW for all the eight *S. mansoni* sequences was used to infer the similarity of the sequences to each other. The guide tree was visualized using FigTree software (<http://tree.bio.ed.ac.uk/software/figtree>). The possible conserved domain search was carried out using the public server (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) for all the expected 8 USPs genes. The ATP binding residues and other ligand binding residues were identified for all the USPs.

The three-dimensional (3D) structure of a selected *S. mansoni* USP was determined using homology modeling available at the SwissModel server (<http://swissmodel.expasy.org/>). The quality of the model was evaluated with Ramachandran plot data, based on the phi-psi torsion angles of all the residues in the model using DeepView-Swiss-PdbViewer (<http://spdbv.vital-it.ch/>). The Ramachandran plot obtained from DeepView was further assessed using Ramachandran plot 2 assessment server (<http://dicsoft1.physics.iisc.ernet.in/rp/>). The Rasmol tool



(<http://www.openrasmol.org/>) was used in visualizing the modeled 3D structures and to identify any possible “SS” bonds and the distribution of the secondary structures. The Rasmol tool was used to locate the positions of the conserved domains on the 3D structure and particularly the key residues involved in the regulatory mechanism. With the Rasmol tool, the actual locations of all the ATP binding residues were mapped on the protein 3D and folded structure.

Reverse-transcriptase PCR assessment of mRNA expression

Reverse-Transcriptase PCR (RT-PCR) was used to assess the mRNA expression of USP genes in selected stages of the LE strain of *S. mansoni*, which was maintained by passage through albino *Biomphalaria glabrata* snails.²⁸ The assessment was done to help determine the suitability of primers and experimental conditions to detect the 8 USP genes in at least one life cycle stage in the selected strains. Miracidia were obtained from eggs recovered from livers and intestines of Swiss mice infected by a subcutaneous injection of approximately 100 cercariae.²⁹ Adult worms were collected by portal perfusion 8 weeks post-infection, and the males and females were carefully separated by hand. Cercariae were collected from infected snails, maintained for 30 minutes on ice, and mechanically transformed into schistosomula.³⁰ The schistosomula were incubated in M169 culture medium³¹ supplemented with 300 U/ml Penicillin (Gibco), 300 µg/ml Streptomycin (Gibco) and 160 µg/mL Gentamicin (Sigma). The schistosomula

were maintained at 37 °C in a 5% CO₂ atmosphere for 7 days.

Total RNA was extracted from eggs, miracidia, schistosomula, male and female using the Trizol reagent (Sigma) in accordance with the manufacturer’s instructions. Single-stranded cDNA was generated by priming 1 to 2 µg of total RNA with 500 ng of oligo(dT)_{12–18} random primers (Boehringer Mannheim) with Superscript II reverse transcriptase (RT) (Life Technologies). Control reactions (without the addition of RT) were included to monitor for contamination with genomic DNA. Residual RNA was hydrolyzed by the addition of 2 units of *Escherichia coli* RNase H (Promega) and incubated at 37 °C for 20 min. After reverse transcription of the total RNA to single-stranded cDNA, 30 PCR cycles were performed as follows: 94 °C for 30s; 55 °C for 30s; and 72 °C for 1 min. The sequences of the gene-specific oligonucleotide primers were designed from the 5’ and 3’ ends of each CDS (Table 1). Primers targeting a 200-bp region of the *S. mansoni* gene encoding alpha-tubulin (GenBank Accession M80214) were used to amplify the constitutively expressed positive control cDNA. RT-PCR amplification products were electrophoresed through 1% agarose gels containing ethidium bromide, and amplicons were viewed with a UV transilluminator.

Evidence for developmental stage expression of genes encoding universal stress protein in *S. mansoni*

Evidence for developmental stage expression of the eight USP domain-encoding genes of *S. mansoni* as

Table 1. Primer sequences for *Schistosoma mansoni* genes encoding alpha-tubulin and universal stress proteins.

Primer name	Length	Forward primer	Reverse primer
Alpha-tubulin	200 bp	TGG TGT CCA GGT TTC AA	ACA AAT GCA CGT TTT GCG TA
Smp_001000	525 bp	ATG GAG AGA AGA CCA TCA GGA TTT TCA	TCA ATT TTT ACA AGA ATT CTT TTG TTC C
Smp_001010	2.748 bp*	ATG GAT AGA AGA CCA TCA GGA TTT	TCA TGA TGT ACT TTG TGT AGT TTC A
Smp_031300	483 bp	ATG GCA TCT AAT TGT ATG CGA CGA	TTA TAA ACT AAT AGA AGG TGG AAT A
Smp_043120	483 bp	ATG GCT GAG TCT AGT GAT CCA	TTA CTG CTT CTC TTG AGG TGG GAC
Smp_076400	555 bp	ATG TCA GAA ACA GAA CAA CCT TCG	TTA TTT TTT CTT TGG ATG TTT AGG TGG T
Smp_097930	480 bp	ATG GGA GTT AAT ACT GAG AAC AAAAA	TTA TCT TAC TGG GGG AGG TAT AAC TAC CA
Smp_136870	873 bp	ATG CCT GGA CTC ATT TCA ACT GAA GAA	TTA ATT ATT CAT AGA TGA TAA AAT GGT
Smp_136890	399 bp	ATG ACT GGA TCC AAT TCA TCA AA	TTA TGA TTC AGA ATC CAC AGC ATT A

Note: *Incorrect assembly. The ORF was adjusted manually.

documented by alignment of Expressed Sequence Tags (ESTs) and Serial Analysis of Gene Expression³² to gene sequences were obtained from SchistoDB. The EST data analyzed in SchistoDB were derived from multiple strains which were deposited in the dbEST resource.³³ The gene expression evidence was defined as the presence of EST tag or SAGE tag for life cycle stages categorized in eight broad categories: egg, miracidium, sporocyst, cercarium, schistosomulum, male worm, female worm and adult worm.

A binary representation strategy combined with visualization of a matrix of binary numbers was used to provide an integrated view of the availability of evidence (presence or absence) of developmental stage expression for the eight USP genes. A 16-digit binary number was generated for each *S. mansoni* USP gene. Further, a matrix of 8 rows (8 genes) by 16 columns (2 methods for 8 stages) was constructed by combining the binary numbers for the 8 genes. The resulting matrix was visualized with matrix2png (<http://www.chibi.ubc.ca/matrix2png/>) with green and red boxes representing absence (0) and presence (1) of gene expression evidence respectively.

The differential or constitutive transcription of *S. mansoni* USP genes in 15 life cycle stages was determined from supplementary datasets provided by Fitzpatrick et al.³⁴ The stages assigned 'Universal Time Codes' (UT) by the data providers were: Egg (UT1), Miracidium (UT2), Mother (2 Day) sporocyst (UT3), Daughter sporocyst (UT4), Cercaria (UT5), 3-Hr schistosomule (UT6), 24-Hr schistosomule (UT7), 3-Day schistosomule (UT8), 6-Day schistosomule (UT9), 2-Wk worm (UT10), 3-Wk worm (UT11), 5-Wk worm (UT12), 7-Wk worm (UT13), 7-Wk male worm (UT14), and 7-Wk female worm

(UT15).³⁴ In the published report by Fitzpatrick et al,³⁴ Dataset S4 spreadsheet file contains expression data [the normalised log₂ AlexaFluor647-dCTP (AF647) intensities] for 35,437 *S. mansoni* oligonucleotide 50-mers and 2,195 controls. Further, Dataset S2 and Dataset S5 spreadsheet files contain a list of differentially expressed transcripts (pairwise comparison of stages) and constitutively expressed transcripts respectively. Constitutively expressed gene product open reading frames (ORFs) were defined as transcripts that displayed the lowest 1% variation in expression across the schistosome life cycle.³⁴ These two datasets also contain protein domain (Pfam) annotation for the oligonucleotides on the microarray. Thus, to identify expression data for *S. mansoni* USP genes, the two spreadsheet files were searched for rows with the code for the universal stress protein signature motif (PF00582) in the Pfam_ID column.

Results

Manual annotation of genes

The database accessions of the eight *S. mansoni* USP genes are presented in Table 2 while Table 3 shows the protein domain coordinates and length as well as entire protein length. The sequence identifiers provided by Forêt et al¹⁸ have been included to extend their findings on the phylogenomic relationships of sequences of metazoan universal stress proteins. Smp_001010 was observed to be an incorrect assembly with a predicted protein sequence of 915 aa. The predicted protein sequence lengths for the other seven USPs ranged from 132 aa to 290 aa. The outlier protein sequence length for Smp_001010 was an indication of a need to verify the annotation. The Open Reading Frame (ORF) was adjusted manually resulting in a 173 aa

Table 2. Accession Identifiers for *Schistosoma mansoni* genes encoding universal stress protein domain.

SchistoDB ID	Scaffold	RefSeq protein	RefSeq mRNA	UniProt ID	Forêt et al 2011*	Entrez gene ID
Smp_001000	Smp_scaff000001	XP_002571572.1	XM_002571526.1	C4PWX3	Sman06	8343749
Smp_001010	Smp_scaff000001	XP_002571573.1	XM_002571527.1	C4PWX4	Sman07	8343750
Smp_031300	Smp_scaff000068	XP_002574582.1	XM_002574536.1	C4Q5U1	Sman01	8350405
Smp_043120	Smp_scaff000103	XP_002575657.1	XM_002575611.1	C4Q8U4	Sman02	8347881
Smp_076400	Smp_scaff000262	XP_002578809.1	XM_002578763.1	C4QH3	Sman03	8354051
Smp_097930	Smp_scaff000612	XP_002570099.1	XM_002570053.1	C1M0Q2	Sman04	8341881
Smp_136870	Smp_scaff000056	XP_002574156.1	XM_002574110.1	C4Q4M5	Sman5b	8355469
Smp_136890	Smp_scaff000056	XP_002574158.1	XM_002574112.1	C4Q4M7	Sman08	8355471

Note: *Source of identifier: Forêt et al.¹⁸

Table 3. Universal stress protein (USP) domains encoded in *Schistosoma mansoni* genome.

SchistoDB ID	USP domain coordinates	Protein length (aa)	Domain length (aa)
Smp_136890	13 to 128	132	115
Smp_097930	7 to 155	159	148
Smp_031300	6 to 155	160	149
Smp_043120	7 to 155	160	148
Smp_001010	15 to 163	173	148
Smp_001000	15 to 163	174	148
Smp_076400	27 to 176	184	149
Smp_136870	12 to 130, 137 to 286	290	118, 149

protein sequence. Further, Smp_136890 (60241 bp to 62619 bp) located in scaffold Smp_scaff000056 was inserted in a gap region from 61195 bp to 62222 bp) of the genome scaffold which can be visualized Artemis Genome Browser and Annotation Tool (<http://www.sanger.ac.uk/resources/software/artemis/>) from the GeneDB³⁵ web page for the gene. Protein motif scanning using MyHits³⁶ predicted two USP domains in only Smp_136870, a 290 aa protein.

In the case of domain length, the following amino acid counts were observed: 115 (Smp_136890), 118 (Smp_136870), 148 (Smp_043120, Smp_001010, Smp_001000) and 149 (Smp_076400). The nucleotide and amino acid sequences for each of the 8 sequences including the re-annotated version of Smp_001010 are available as Supplementary File 1 and Supplementary File 2 respectively. Grouping of the eight sequences by sequence length identified 5 groups of sequences: 132 aa (Smp_136890), 159 aa to 160 aa (Smp_097930, Smp_031300, Smp_043120), 173 aa to 174 aa (Smp_001010, Smp_001000), 184 aa (Smp_076400) and 290 aa (Smp_136870).

Ligand binding residues in *Schistosoma mansoni* universal stress proteins

A guide tree generated by ClustalW for all the eight *S. mansoni* sequences was used to infer the similarity of the sequences to each other (Fig. 1). Combined with the protein length information in Table 2, the guide tree could also provide insights into genes that may have arisen by gene duplication in *S. mansoni*. In the fragment B of the universal stress protein MJ_0577 of *Methanocaldococcus jannaschii*, the 12 binding site residues were identified as Pro11, Tyr12, Asp13, Val41, Met126, Gly127, His129, Gly130, Gly140, Ser141, Val142 and Thr143.²⁷ Further, the Asp13 and Val41 bind to adenosine. The amino acid sequences from

Met126 to Thr143 contain the motif G2xG9xG(S/T), which includes binding sites for the phosphoryl and ribosyl groups of ATP. The protein sequences of six *S. mansoni* USPs (Smp_001000, Smp_001010, Smp_031300, Smp_043120, Smp_076400, and Smp_097930) have the conserved aspartate as well as the conserved ATP-binding motif (Fig. 2).

Using the 12 ligand binding site residues in fragment B of MJ_5077 (Protein DataBank Accession: 1MJH_B) sequence as a reference set, the residues that align to these binding sites in the 8 *S. mansoni* USP were identified using the National Center for Biotechnology Information (NCBI) Conserved Domain Search tool (Table 4). Their identity (UniProt, SchistoDB and Forêt et al)¹⁸ and count for the eight *S. mansoni* USP were visualized using Tableau

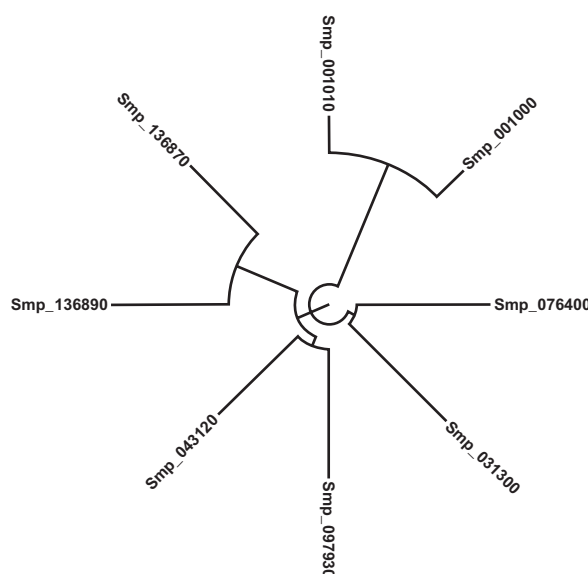


Figure 1. Clustal guide tree of the 8 *Schistosoma mansoni* Universal Stress Proteins. Guide tree of the similarity of the protein sequences to each other as constructed from a multiple sequence alignment of the *S. mansoni* universal stress protein sequences and visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree>).



Smp_001010	-----MRRPSGFSKPPPIIGSRSLIIAIDGSEHSKKAFFDYLLKWLQRPDDSVT	48
Smp_001000	-----MERRPSGFSKIPPIIGSRSVLIAIDGSEHSKKAFFDYLLKWLQRPDDSVT	48
Smp_031300	-----MASNCMRRVLLPIDGSEHSKRAVNWYLTEFSRPDDFAY	38
Smp_097930	-----MGVNTENKVVFLPVDASDHSARAFQWYLDNLRGKNDLH	40
Smp_043120	-----MAESSDPSRVILIPIDGSDHCDRAFRWYLENMRKRDTCIT	40
Smp_076400	MSETEQPSTSDGLDIGETKGTISMTDATRKVLMVPVDGSEHSERAFNWYMDNVMKITDGLY	60
MJ_0577	-----MSVMYKKILYPTDFSETAEIALKHVKAFKTLKAEVI	37
	: : : . * * : . * . :	
ATP Contacts	=== A	
Smp_001010	IYHAVE--PVSLPTISLSNPISIPSE--WSNIVQTNVQKRVREL----EKDYSADCLAHNL	101
Smp_001000	IYHAVG--PVSLPTISSNPNISIPSEE--WSNLVQTNVQKRVREL----ENDYSADCLAHNL	101
Smp_031300	FFHVVE--AHYS--KSTANESYDHGKE--LNTNLDKNIKMYSELGKILGDKLHDDLKNSNI	93
Smp_097930	FVYVIK--PIFTPTPTIELAMASSPITD--IIQSTQENIENAKKLL----QKYLKAKRFGI	93
Smp_043120	FVHVIE--PVYNTPAIGMTMESPPIDP--MTRVMEESIEQKGLG----QYMHEAKSYKL	93
Smp_076400	LVHIVE--PLSQGLNYNLASKSPSIKDDFSKHLNSLVESGRALR----AKFFTRCEDSGL	114
MJ_0577	LLHVIDEREIKKRDI F S L L L G V A G L N K S V E E F E N E L K N K L T E E A K N K M E N I K K E L E D V G F	97
	: : : . : : . : : . : :	
ATP Contacts	= A	
Smp_001010	TYQFLYESVDHIGASIIIEKAEKYNACLIIIGSRGLGAIKRTIMGSVSDYVHHANTTICV	161
Smp_001000	TYQFLYESVDHIGAAIVQNAEKYNVHLLIVGSRGLGAIKRTFMGVSVDYVIHHANTAVCV	161
Smp_031300	QMEYVMQIGNKPGELIINVAKERSVDVILIGNRGLGAFRRFLGVSSEYILHHCNVFPII	153
Smp_097930	SCQAFVHVNAKPGPTLVKFAEEQKADIIIGPRGLGLIRRTLLGSVTNYVMHHTKTPLVV	153
Smp_043120	NAKAFLHVDTKPGSSLVKAISDHKANVILMGNRGLGAIIRRTFLGVSVDYVLLHSHIPVVI	153
Smp_076400	SARFTIHVGTKPGENIVRIAHEHGVLDLVIIGNRGIGTVKRTFLGVSVDYVLLHSHANVPVVI	174
MJ_0577	KVKDIIIVG--IPHEEIVKIAEDEGVDI I IMGSHGKTNLKEILLGSVTENVIKKS NKPVLV	156
	. : : . . : : * * . . : : * * : : : : . . :	
ATP Contacts	===== R, P P	
Smp_001010	VPWIDETTQSTS-	173
Smp_001000	IPSITEQKNSCKN	174
Smp_031300	IPPSISL-----	160
Smp_097930	IPPPVR-----	159
Smp_043120	VPPQEKQ-----	160
Smp_076400	IPPPKHPKKK---	184
MJ_0577	VKRKNS-----	162
	:	

Figure 2. Multiple Sequence Alignment of the *Schistosoma mansoni* and *Methanocaldococcus jannaschii* universal stress protein sequences. Six USP of *S. mansoni* and USP (MJ_0577) of *Methanocaldococcus jannaschii* were aligned using ClustalW. Alignment shows the motif G2xG9xG(S/T) (indicated by series of #) that contain ATP binding residues was present in the six *S. mansoni* USPs (Smp_001000, Smp_001010, Smp_031300, Smp_043120, Smp_076400) as well as in MJ_0577, a known ATP-binding universal stress protein. Residues making contacts with ATP are indicated with “=” according annotation to Zarembinski et al.²⁷

Abbreviations: A, Adenine; R, Ribosyl; P, Phosphoyl groups.

Table 4. Amino acid residues in ligand binding sites of *Schistosoma mansoni* universal stress proteins.

Protein	Ligand binding position and amino acid residue*											
	L1 11	L2 12	L3 13	L4 41	L5 126	L6 127	L7 129	L8 130	L9 140	L10 141	L11 142	L12 143
1MJH_B	P	T	D	V	M	G	H	G	G	S	V	T
Smp_001000	A	I	D	A	V	G	R	G	G	S	V	S
Smp_001010	A	I	D	A	I	G	R	G	G	S	V	S
Smp_031300	P	I	D	V	I	G	R	G	G	S	V	T
Smp_043120	P	I	D	V	M	G	R	G	G	S	V	S
Smp_076400	P	V	D	I	I	G	R	G	G	S	V	S
Smp_097930	P	V	D	V	I	G	R	G	G	S	V	T
Smp_136870A	P	V	Y	V	M	L	R	S				
Smp_136870B				V	M	L	R	S	G	S	V	S
Smp_136890	P	L	D	V								

Note: *Amino acid residue positions are based on the sequence of 1MJH_B (<http://pdbe.org/1mjh/>). Table shows the amino acids in *Schistosoma mansoni* that were aligned to the 12 ligand binding sites (L1 to L12) of fragment B of the universal protein MJ_0577 of *Methanocaldococcus jannaschii* (1MJH_B). A visualization of Table 4 is presented in Figure 4.

Public Software (<http://www.tableausoftware.com/public>) and presented in Figure 3. The amino acid residues in *S. mansoni* USPs found aligned to the ligand binding positions of 1MJH_B were Ala, Arg, Asp, Gly, Ile, Leu, Met, Pro, Ser, Thr, Tyr and Val. The patterns in the alignment of amino acid residues to the ligand binding residues of 1MJH_B could provide insights into the regulation of *S. mansoni* proteins (Fig. 3, Table 4).

Smp_031300 protein sequence was selected for homology modeling based on the length of the USP domain, completeness of the USP domain and the shared identity of certain ligand binding residues with the sequence of 1MJH_B. The USP domain of Smp_031300 was predicted to be 149 aa. In addition, the valine and threonine in position 41 and 143 respectively of 1MJH_B are conserved in Smp_031300 protein sequence. The distribution of the predicted binding residues of Smp_031300 and particularly its constituted ATP binding residues are elaborated in Figure 4a.

The crystal structure of Smp_031300 has not yet been determined; here we present the 3D homology model structure of the protein. The Rasmol tool was used for visualizing the tertiary structures of the Smp_031300 protein. The structure is composed of 10 α -helices, 7 β sheets (stands) 7 turns (loops) and 107 hydrogen bonds. The α -helices are represented in red; β sheets are yellow and loops are depicted

in blue. The white or greyish colored regions near the loops are other secondary structures (Fig. 4a). The ATP binding residues (Magenta) are located on the loop and part of β sheet, while the other binding residues (Green) are located on the α -helices, β sheets and the loops. The Ser residue (Black), which is also part of the ATP binding motif is also located on the loop. The interface location of these residues on the folded structure (Fig. 4b) shows that the ATP binding motif (green) residues are located in the fold (core) of the protein in a top-bottom orientation, while the other binding residues (magenta) were more close to the surface of the protein. The Ser (black) residue was shown to be located very close to the other ATP binding residues, but it is orientated towards the surface of the protein molecule. The position of serine indicates that it bridges the link between the ATP binding motif residues and the other ligand binding residues and also as a site for phosphorylating the protein molecule.

Reverse-transcriptase PCR assessment of mRNA expression

The transcription of the eight *S. mansoni* USP encoding genes was determined for the following life cycle stages: egg, miracidium, schistosomulum, male worm, and female worm. In the RT-PCR mRNA expression experiments using LE strain of *S. mansoni*, only two genes (Smp_043120

UniProt ID	Schisto DB ID	Gene symbol	Ligand binding residue (symbol)											
			Ala	Arg	Asp	Gly	Ile	Leu	Met	Pro	Ser	Thr	Tyr	Val
C1M0Q2	Smp_097930	Smans04		1	1	3	1			1	1	1		3
C4PWX3	Smp_001000	Smans06	2	1	1	3	1				2			2
C4PWX4	Smp_001010	Smans07	2	1	1	3	2				2			1
C4Q4M5	Smp_136870A	Smans5a		1				1	1	1	1		1	2
	Smp_136870B	Smans5b		1		1		1	1		3			2
C4Q4M7	Smp_136890	Smans08			1			1		1				1
C4Q5U1	Smp_031300	Smans01		1	1	3	2			1	1	1		2
C4Q8U4	Smp_043120	Smans02		1	1	3	1		1	1	2			2
C4QHX3	Smp_076400	Smans03		1	1	3	2			1	2			2

Number of records

Figure 3. Ligand binding residues in *Schistosoma mansoni* universal stress protein sequences. Using the 12 binding site residues in 1MJH_B as a reference set, the residues that align to the binding sites in the 8 *Schistosoma mansoni* universal stress proteins (USPs) were identified. The following 12 binding amino acid residues (Ala, Arg, Asp, Gly, Ile, Leu, Met, Pro, Ser, Thr, Tyr and Val) from the *S. mansoni* USP sequences were aligned to the ligand binding sites of 1MJH_B. A website that allows for interaction with the dataset presented in Table 4 can be accessed at http://public.tableausoftware.com/views/schisto_mansoni/ligand_residues.

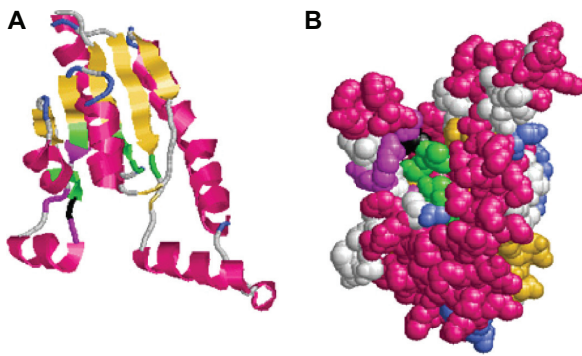


Figure 4. Homology model of *Schistosoma mansoni* universal stress protein Smp_031300. On the 3-dimensional structure, the ligand binding residues are indicated in different colors. Magenta represents location for the ATP binding residues (Gly124, Arg126, Gly127, Gly137, Ser138) and the green represent other binding residues (Pro12, Ile13, Asp14, Val42, Ile123 Val139, Ser140). Ser138 (black) indicates the phosphorylation site.

and Smp_001000) were amplified using RT-PCR. Smp_043120 was amplified in schistosomulum and female stages while Smp_001000 was amplified in schistosomulum, male worm and female worm stages (Fig. 5).

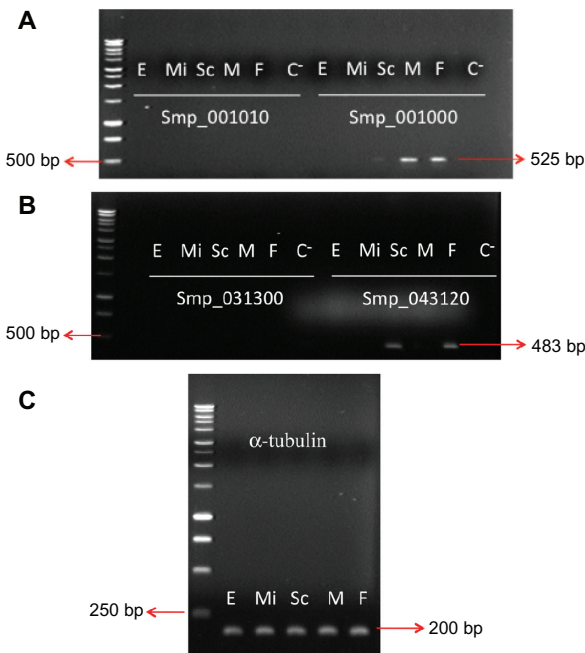


Figure 5. Reverse Transcriptase-PCR assessment of mRNA expression for genes encoding universal stress proteins in strain LE of *Schistosoma mansoni*. (A) and (B): Primers were constructed to assess mRNA expression for genes encoding the 8 *S. mansoni* USP genes (four shown, A and B) in strain LE of *S. mansoni*. Sizes of the amplified bands are shown relative to a 1kb DNA Ladder Promega. Life cycle developmental stages examined were Egg (E), Miracidium (Mi), Schistosomula (Sc), Male (M), Female (F). The lane labeled C⁻ is negative control. (C) The alpha-tubulin (α -tubulin) mRNA was used as a constitutively expressed positive control for the RT-PCR. In the case of the USP genes, only Smp_001000 and Smp_043120 were amplified in at least one stage.

Evidence for developmental stage expression of genes encoding universal stress protein in *S. mansoni*

Combining the evidence for developmental expression based on public-domain EST and SAGE data revealed the transcription of USP genes in at least one of the life cycle stages of the helminth (Fig. 6, Table 5). Gene expression evidence for Smp_097930, Smp_136890 and Smp_136870 were available for only the SAGE method. Smp_097930 and Smp_136890 transcripts were documented for only the adult worm and sporocyst respectively. In the case of Smp_136870, SAGE transcript evidence was documented in both miracidium and the sporocyst stages. The data integration also revealed that Smp_001010, and Smp_043120 had evidence of gene expression using both methods for the Female and Adult stages.

Differential or constitutive expressed USP genes were identified in the supplementary dataset from Fitzpatrick et al³⁴ using the keyword “PF00582”. Pair-wise statistical comparison at $P < 0.05$ by the authors showed that Smp_031300 transcript was up-regulated in the following stages; miracidium (compare to egg); cercaria (compared to daughter sporocysts) and the 7-wk male worm (compared to 7-wk female worm) (Supplementary File 3). The Smp_076400 gene was down-regulated in the intra-snail stages compared to the intra-mouse stages (Supplementary File 3). Further, intra-snail stages/intra-mouse stages comparisons represent those genes that are differentially expressed

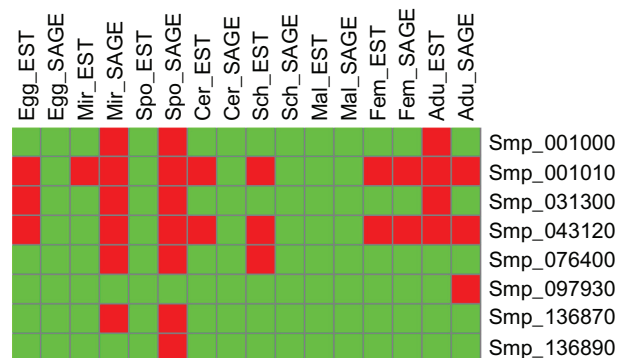


Figure 6. Integration of evidence of gene expression from Expressed Sequence Tags (EST) and Serial Analysis of Gene Expression (SAGE) of the 8 *Schistosoma mansoni* USP genes in developmental stages. Red represents expression being detected and green represents no evidence of expression. The SchistoDB IDs are indicated on the right side.

Abbreviations: Mir, Miracidia; Spo, Sporocyst; Cer, Cercariae; Sch, Schistosomula; Mal, Male; Fem, Female; Adu, Adult.

Table 5. Summary of gene expression evidence for *Schistosoma mansoni* universal stress proteins.*

Genes for universal stress proteins	Frequency of evidence**	
	EST	SAGE
Smp_001000	1	2
Smp_001010	6	4
Smp_031300	2	2
Smp_043120	5	4
Smp_076400	1	2
Smp_097930	0	1
Smp_136870	0	2
Smp_136890	0	1
Total	15	18

Note: *Table was constructed from Figure 6. ** EST: Expressed Sequence Tags; SAGE: Serial Analysis of Gene Expression. The frequency provided for each gene is count of evidence for gene expression from public domain EST and SAGE data on *Schistosoma mansoni*. The maximum frequency is the number of stages (in this case, eight).

as detected by the following formula: \log_2 gene expression ratios of (Miracidium + Daughter sporocyst + Mother Sporocyst + Cercariae)/4 compared to \log_2 gene expression ratios of (3-Hr schistosomules + 24-Hr schistosomules + 3-Day schistosomules + 6-Day schistosomules + 2-Wk worms + 3-Wk worms + 5-Wk worms + 7-Wk worms + 7-Wk Male worms + 7-Wk Female worms/10) when egg and miracidium expression data were compared. Smp_043120 was constitutively expressed in the life cycle stages (Supplementary File 3). The average expression intensities from the three USP genes were visualized with a Heat map (Fig. 7) to provide a summary of the relative expression profiles of the genes. Clearly, Smp_043120 is expressed in all stages, while peak expression for Smp_031300 was observed in 3-Hr schistosomulum

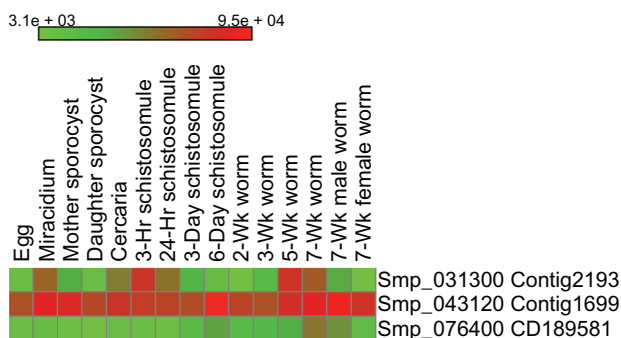


Figure 7. Heat map showing expression of Smp_031300, Smp_043120, and Smp_076400. Heat map visualization of differentially and constitutively expressed transcripts for Smp_031300 (Contig2193), Smp_043120 (Contig1699), and Smp_076400 (CD189581) in the life cycle stages of *Schistosoma mansoni*.

and 5-Wk worm. The dataset and graphs of the expression intensities for each gene are available as Supplementary File 3.

Discussion

In 1992, Nyström and Neidhardt⁷ reported the cloning, mapping and nucleotide sequencing of a monocistronic gene in *Escherichia coli* encoding a small (13.5 kDa) cytoplasmic protein with increased synthesis during growth inhibition or presence of toxic agents. The gene was designated UspA and subsequent mutant-based analysis of the gene led to the proposal that the encoded protein may have a general protective function related to the growth arrest state.^{8,9} Genomic data have been used to identify genes encoding the USP domain in the archaea, bacteria and eukaryotes including *S. mansoni*.^{4,5,18,21,22} The research investigation reported here represents a useful application of data mining and integration to generally underutilized bioinformatics, genomic and functional genomic resources on the parasitic helminth *S. mansoni*. Genomic data from a strain of *Biomphalaria glabrata*, a major snail vector of *S. mansoni* has enabled microarray-based time series analysis of internal defense and stress-related genes of the vector in response to parasite infection.³⁷ There is also evidence from the dbEST³³ for the presence of a USP gene with mapped expressed sequence tag EE723402 in the genome of the M-line of *B. glabrata*. This USP transcript (<http://www.ncbi.nlm.nih.gov/nucest/EE723402>) was expressed in response to injection with the bacteria *Micrococcus leteus* after 12 hours exposure. The USP genes in schistosomes and their intermediate hosts could be of particular value to researchers investigating *Schistosoma* biology, disease ecology and intervention targets. The prioritized information from these investigations could be used for comparative evaluation of stress response among schistosomes.

As of January 23, 2011, version 4 is the current annotation version of the draft *S. mansoni* genome available in GeneDB.³⁵ Manual re-annotation of published automated gene annotation and prediction is recognized as a mechanism to reduce the propagation of errors that may hamper progress in functional genomics and translational research.³⁸ The manual annotation of the eight USP genes led to adjustment of the gene prediction for Smp_001010 (Supplementary



File 1 and Supplementary File 2). Since the genome is draft, it is expected that incomplete contig assemblies as well as incorrect gene structures will exist. For example, the comments section of the GeneDB page of Smp_001010 (http://www.genedb.org/gene/Smp_001010:mRNA) indicates that the coding sequence (CDS) was manually annotated in the v3.1 set, but the gene structure is different or does not exist in the auto predicted v4 set. Database curators at SchistoDB²⁶ are implementing the manual curation of incorrect gene predictions directly in the GeneDB to help improve the overall annotation of the *S. mansoni* genome.

The USP sequences were grouped by length to facilitate studies on the evolutionary history and constraints of the *S. mansoni* USP genes. Additionally, the number and type of protein domains were also determined. It is noteworthy that none of the eight proteins had any other significant protein domains such as the protein kinase found in bacteria³⁹ and plants USPs.⁴ Literature curation of *S. mansoni* USPs¹⁸ and a motif prediction search conducted during this research confirmed that the Smp_136870 gene which encodes a 290 aa protein sequence is the only gene coding for tandem USP domains. In *E. coli*, mutants of UspE (which contains tandem USP domains) were unable to form cell-cell interactions and cell aggregates in stationary phase.⁶ In *Mycobacterium tuberculosis*, which has 8 of its 10 USPs having tandem domains, Rv2623 has growth-regulating capability linked to ATP-binding.⁴⁰

There was evidence from SAGE profiling for the transcription of Smp_136870 in miracidium and sporocyst stages. These are the intramolluscan stages exposed to oxidative stress within the snail haemocytes.⁴¹ This observation may be related to the fact that defense against oxidative damage is crucial for the developmental stages in the snail intermediate host. Interestingly, hydrogen peroxide is the main Reactive Oxygen Species (ROS) responsible for killing sporocysts.⁴² In *E. coli*, cells forming aggregates are protected from hydrogen peroxide killing.⁴³ In microarray experiments to determine the mode of action of praziquantel (PZQ) by Aragon et al,¹¹ the induction of genes annotated with the Gene Ontology 'antioxidant category' suggest that treatment with PZQ may result in a molecular response similar to that observed when schistosomes undergo oxidative

stress. Further, a microarray experiment observed that induction of the antioxidant thioredoxin occurs during egg and adult male stages of *S. japonicum*.¹⁹ These results support the need to further evaluate *S. mansoni* USP transcription regulation in response to oxidative stress during its developmental stages. Gene Smp_136870 may be a prime candidate that is up-regulated in response to oxidative stress during intramolluscan development.

The USP domain containing proteins can be divided into two classes based on the presence or absence of ATP binding motifs.²⁷ Six of the eight *S. mansoni* USP genes (Smp_001000, Smp_001010, Smp_031300, Smp_043120, Smp_076400, Smp_097930) encoded ATP binding motif residues as well as residues predicted to binds other ligands and phosphorylation site (Figs. 2 and 3). The main residues in the ATP binding motif (Gly-2x-Gly-9x-Gly (Ser/Thr) observed among these USPs were Gly, Arg, and Ser as shown by all the proteins with this motif. ATP binding generally involves hydrogen bond formation with charged or polar side chains.⁴⁴ The key residue regulating ATP binding could be the Gly residue. The Gly residues have been shown to contribute to the formation of hydrogen bonds with the gamma-phosphate of nucleotide triphosphates.⁴⁵ Using the Smp_031300 as a structural model for the *S. mansoni* USPs, the Ras-mol visualization tool revealed a total of 107 hydrogen bonds. Also the interface localization of the ATP binding residues positioned these hydrogen bonds in the core of the protein in a top-bottom orientation (Fig. 4b). This configuration might be a functional adaptation for strong bonding with the ATP molecule. The main residues including the ATP binding residues that could regulate the functioning of these USP genes were Ala, Asp, Gly, Ile, Leu, Met, Pro, Arg, Ser, Thr, Val, and Tyr. The amino acids Ser, Thr and Tyr which are known to function as phosphorylation site and protein kinase A (PKA), had been shown to be developmental regulated in *S. mansoni*.⁴⁶ The other residues could be responsible for binding other molecules in the stress overhauling mechanism. The orientation of these non-ATP binding residues to the protein surface (Fig. 4b) suggests an adaptive mechanism to bind incoming molecules within the cellular network.

A major contribution of our research is the validation of two USP genes (Smp_043120 and



Smp_001010) in strain LE of *S. mansoni*. These two genes also had the highest frequency of gene transcripts observed when EST and SAGE evidence were visualized (Table 5, Fig. 6). The developmental regulation of Smp_043120 and Smp_001010 in both intra-mammalian and intra-molluscan stages suggests a crucial role in environmental stress response in *S. mansoni*. The USP genes expressed in the miracidium could be induced in response to environmental stresses encountered in the snail habitat as well as during penetration of the snail tissue. Gene transcript evidence for Smp_001010 was observed with EST and SAGE methods (Fig. 6) providing a further justification for additional studies on the function of this gene in the response to environmental stressors.

Analysis of microarray datasets covering 15 key schistosome stages identified at a significance of $P < 0.05$ showed that Smp_031300 was up-regulated in three pairwise comparisons of the development stages, two of which are intramolluscan stages (Fig. 6). Smp_043120 was amplified here and was among the 355 constitutively-expressed transcripts identified by Fitzpatrick et al.³⁴ Thus, Smp_043120 represent a prime candidate for additional investigation including RNA interference, in-situ hybridization and sequence diversity and gene expression in field isolates. Future research will help to clarify the role of Smp_043120 in *S. mansoni* development and host-schistosome interactions. The utility of focused analysis of the expression of members of this gene family has been demonstrated by Fitzpatrick et al³⁴ with analysis of three major gene families: fucosyltransferases (PF00852); tetraspanins (PF00335) and G-protein Coupled Receptors (PF0001, PF0002, PF0003).

The functions, regulation and evolutionary history of the *S. mansoni* USP genes remain to be elucidated. The role of USP genes in other organisms includes response to diverse stresses such as oxidative stress and toxic substances. Based on the sequence length, two groups of sequences with multiple members were identified as follows: 159 aa to 160 aa (Smp_097930, Smp_031300, Smp_043120) and 173 aa to 174 aa (Smp_001010, Smp_001000). Genes that share sequence length may have resulted from a gene duplication event and may have redundant functions. In *Mycobacterium tuberculosis* (Mtb), knock-out mutants of six USP genes did not affect long-term survival of Mtb in oxygen starvation.¹⁰

A binary representation of data from EST and SAGE gene transcript detection revealed evidence of expression of all the 8 genes in at least one of the life cycle stages examined (Fig. 6). The purpose of the representation was primarily to view the evidence of transcription from public domain data that was compared. Thus the purpose was not comparing the efficiency of the methods for transcript detection. The life cycle stages sampled in the EST and SAGE datasets are not necessary identical and processed under same conditions. However, there were three stages (miracidium, female worm and adult worm) where the agreement of evidence could be further discussed. Transcripts of Smp_001010 were documented in these three stages by EST and SAGE. Instances where EST evidence was observed without SAGE evidence are Smp_001000 and Smp_031300 for Adult Worm. In the miracidium stage, six genes had SAGE transcripts compared to one gene (Smp_001010) for EST (Fig. 6). An explanation for this is that the SAGE method is useful for detecting low-abundant transcripts.⁴⁷ It is possible that *S. mansoni* USP transcripts in the miracidium stage are low-abundant transcripts, which are induced to perform specialized functions triggered by environmental stressors during the complex life cycle of the parasite. Smp_001000 and Smp_031300 did not have SAGE tag evidence at the adult worm stage.

The Reverse Transcriptase-PCR assessment of mRNA expression for USP genes in strain LE of *S. mansoni* detected transcription in only 2 USP genes (Smp_001000 and Smp_043120) (Fig. 5). The expectation was that mRNA expression of the eight genes would be detected by RT-PCR in at least one of the developmental stages. Thus, additional assays will need to be conducted to optimize the detection and measure expression levels of the *S. mansoni* USP genes under diverse conditions and developmental stages.

Conclusion

Manual re-annotation, sequence analysis, data mining and experimental assays were used to determine distinctive sequence features and developmental regulation of eight *Schistosoma mansoni* genes encoding the universal stress protein domain. One of the genes was manually adjusted for the coding sequence. The research investigation provided evidence from public



domain gene expression data for differential and constitutive expression of members of the gene family. The egg, miracidia and cercaria stages are known to undergo rapid transition between environments in and outside the host that are accompanied by temperature, osmotic stresses and other stressors. Therefore, during development in vertebrate blood and snail haemocytes, schistosome USP genes may function in defense against hydrogen peroxide induced oxidative stress. In settings where there is risk of human and animal diseases caused by schistosomes, it could be of interest to determine seasonal, geographic, strain and species differences in USP gene expression especially in the miracidium and cercaria free-living environmental forms. Such research could elucidate schistosomiasis ecology and be useful for surveillance, intervention and risk assessment. This report serves to catalyze the formation of a network of researchers to understand the function and regulation of the universal stress proteins encoded in genomes of schistosomes and their snail intermediate hosts.

Acknowledgements

National Science Foundation (EPS-0903787; NSF-DBI-0958179; DBI-1062057); National Institutes of Health (NIH-NCRR G12RR13459; NIH-NIGMS T36GM095335; NIH-NIMHD 1P20MD002725-01; NIH-NCRR-P20RR016476; NIH-NCRR-P20RR-016460); US Department of Homeland Security Science and Technology Directorate (2007-ST-104-000007; 2009-ST-062-000014; 2009-ST-104-000021); the International Center for Genetic Engineering and Biotechnology, Trieste, Italy/National Biotechnology Development Agency/West African Biotechnology Workshop Series Protozoan Pathogens Comparative Genomics Workshop (December 2009), Nigeria; FAPEMIG (PPM-00439-10) and CNPq (306879/2009-3) Brazil.

Disclosures

Author(s) have provided signed confirmations to the publisher of their compliance with all applicable legal and ethical obligations in respect to declaration of conflicts of interest, funding, authorship and contributorship, and compliance with ethical requirements in respect to treatment of human and animal test subjects. If this article contains identifiable human subject(s) author(s) were required to supply signed

patient consent prior to publication. Author(s) have confirmed that the published article is unique and not under consideration nor published by any other publication and that they have consent to reproduce any copyrighted material. The peer reviewers declared no conflicts of interest.

References

1. Brindley PJ, Mitreva M, Ghedin E, Lustigman S. Helminth genomics: The implications for human health. *PLoS Negl Trop Dis*. 2009;3:e538.
2. Berriman M, Haas BJ, LoVerde PT, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*. 2009;460:352–8.
3. Isokpehi RD, Simmons SS, Cohly HH, et al. Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinform Biol Insights*. 2011;5:41–58.
4. Kerk D, Bulgrien J, Smith DW, Gribskov M. *Arabidopsis* proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiol*. 2003;131:1209–19.
5. Kvint K, Nachin L, Diez A, Nystrom T. The bacterial universal stress protein: function and regulation. *Curr Opin Microbiol*. 2003;6:140–5.
6. Nachin L, Nannmark U, Nystrom T. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J Bacteriol*. 2005;187:6265–72.
7. Nystrom T, Neidhardt FC. Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. *Mol Microbiol*. 1992;6:3187–98.
8. Nystrom T, Neidhardt FC. Isolation and properties of a mutant of *Escherichia coli* with an insertional inactivation of the *uspA* gene, which encodes a universal stress protein. *J Bacteriol*. 1993;175:3949–56.
9. Nystrom T, Neidhardt FC. Expression and role of the universal stress protein, *UspA*, of *Escherichia coli* during growth arrest. *Mol Microbiol*. 1994;11:537–44.
10. O'Toole R, Williams HD. Universal stress proteins and *Mycobacterium tuberculosis*. *Res Microbiol*. 2003;154:387–92.
11. Aragon AD, Imani RA, Blackburn VR, Cunningham C. Microarray based analysis of temperature and oxidative stress induced messenger RNA in *Schistosoma mansoni*. *Mol Biochem Parasitol*. 2008;162:134–41.
12. Borch M, Kiernan M, Rust K, et al. Schistosomiasis: a case study. *Urol Nurs*. 2009;29:26–9.
13. van der Werf MJ, de Vlas SJ, Brooker S, et al. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop*. 2003;86:125–39.
14. Lambertucci JR. Acute schistosomiasis mansoni: revisited and reconsidered. *Mem Inst Oswaldo Cruz*. 2010;105:422–35.
15. Shiff C, Naples JM, Isharwal S, Bosompem KM, Veltri RW. Non-invasive methods to detect schistosome-based bladder cancer: is the association sufficient for epidemiological use? *Trans R Soc Trop Med Hyg*. 2010;104:3–5.
16. Webster JP, Oliveira G, Rollinson D, Gower CM. Schistosome genomes: a wealth of information. *Trends Parasitol*. 2010;26:103–6.
17. Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature*. 2009;460:345–51.
18. Foret S, Seneca F, de JD, et al. Phylogenomics reveals an anomalous distribution of USP genes in metazoans. *Mol Biol Evol*. 2011;28:153–61.
19. Gobert GN, Moertel L, Brindley PJ, McManus DP. Developmental gene expression profiles of the human pathogen *Schistosoma japonicum*. *BMC Genomics*. 2009;10:128.
20. Jolly ER, Chin CS, Miller S, et al. Gene expression patterns during adaptation of a helminth parasite to different environmental niches. *Genome Biol*. 2007;8:R65.
21. Technau U, Rudd S, Maxwell P, et al. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet*. 2005;21:633–9.



22. Han ZG, Brindley PJ, Wang SY, Chen Z. *Schistosoma* genomics: new perspectives on schistosome biology and host-parasite interaction. *Annu Rev Genomics Hum Genet.* 2009;10:211–40.
23. Mourao MM, Dinguirard N, Franco GR, Yoshino TP. Phenotypic screen of early-developing larvae of the blood fluke, *Schistosoma mansoni*, using RNA interference. *PLoS Negl Trop Dis.* 2009;3:e502.
24. Simoes MC, Lee J, Djikeng A, et al. Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics.* 2011;12:47.
25. Huang J, Hao P, Chen H, et al. Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. *PLoS One.* 2009;4:e8206.
26. Zerlotini A, Heiges M, Wang H, et al. SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic Acids Res.* 2009;37:D579–82.
27. Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, et al. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc Natl Acad Sci U S A.* 1998;95:15189–93.
28. Pellegrino J, Katz N. Experimental chemotherapy of Schistosomiasis mansoni. *Adv Parasitol.* 1968;6:233–90.
29. Dalton JP, Day SR, Drew AC, Brindley PJ. A method for the isolation of schistosome eggs and miracidia free of contaminating host tissues. *Parasitology.* 1997;115(Pt 1):29–32.
30. Ramalho-Pinto FJ, Gazzinelli G, Howells RE, et al. *Schistosoma mansoni*: defined system for stepwise transformation of cercaria to schistosomule in vitro. *Exp Parasitol.* 1974;36:360–72.
31. Basch PF. Cultivation of *Schistosoma mansoni* in vitro. I. Establishment of cultures from cercariae and development until pairing. *J Parasitol.* 1981;67:179–85.
32. Taft AS, Vermeire JJ, Bernier J, et al. Transcriptome analysis of *Schistosoma mansoni* larval development using serial analysis of gene expression (SAGE). *Parasitology.* 2009;136:469–85.
33. Boguski MS, Lowe TM, Tolstoshev CM. dbEST—database for “expressed sequence tags”. *Nat Genet.* 1993;4:332–3.
34. Fitzpatrick JM, Peak E, Perally S, et al. Anti-schistosomal intervention targets identified by lifecycle transcriptomic analyses. *PLoS Negl Trop Dis.* 2009;3:e543.
35. Hertz-Fowler C, Peacock CS, Wood V, et al. GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.* 2004;32:D339–43.
36. Pagni M, Ioannidis V, Cerutti L, et al. MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.* 2007;35:W433–7.
37. Hanington PC, Lun CM, Adema CM, Loker ES. Time series analysis of the transcriptional responses of *Biomphalaria glabrata* throughout the course of intramolluscan development of *Schistosoma mansoni* and *Echinostoma paraensei*. *Int J Parasitol.* 2010;40:819–31.
38. Lorenzi HA, Puiu D, Miller JR, et al. New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information. *PLoS Negl Trop Dis.* 2010;4:e716.
39. Heermann R, Lippert ML, Jung K. Domain swapping reveals that the N-terminal domain of the sensor kinase KdpD in *Escherichia coli* is important for signaling. *BMC Microbiol.* 2009;9:133.
40. Drumm JE, Mi K, Bilder P, et al. *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. *PLoS Pathog.* 2009;5:e1000460.
41. Zelck UE, Von JB. Antioxidant enzymes in intramolluscan *Schistosoma mansoni* and ROS-induced changes in expression. *Parasitology.* 2004;128:493–501.
42. Hahn UK, Bender RC, Bayne CJ. Killing of *Schistosoma mansoni* sporocysts by hemocytes from resistant *Biomphalaria glabrata*: role of reactive oxygen species. *J Parasitol.* 2001;87:292–9.
43. Schembri MA, Hjerrild L, Gjermansen M, Klemm P. Differential expression of the *Escherichia coli* autoaggregation factor antigen 43. *J Bacteriol.* 2003;185:2236–42.
44. Jiang LH, Rassendren F, Surprenant A, North RA. Identification of amino acid residues contributing to the ATP-binding site of a purinergic P2X receptor. *J Biol Chem.* 2000;275:34190–6.
45. Milburn MV, Tong L, deVos AM, et al. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science.* 1990;247:939–45.
46. Swierczewski BE, Davies SJ. Developmental regulation of protein kinase A expression and activity in *Schistosoma mansoni*. *Int J Parasitol.* 2010;40:929–35.
47. Kim YC, Jung YC, Xuan Z, et al. Pan-genome isolation of low abundance transcripts using SAGE tag. *FEBS Lett.* 2006;580:6721–9.

Supplementary Data

Supplementary File 1

Nucleotide sequences of *Schistosoma mansoni* universal stress protein genes

Supplementary File 2

Amino acid sequences of 8 *Schistosoma mansoni* universal stress proteins

Supplementary File 3

Gene Expression Datasets Integration and Analysis for *Schistosoma mansoni* Genes Encoding Universal Stress Proteins

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>