
Article:
Konečný, Lukáš, Jedličková, Lucie, Ibnahaten, Zaina et al. (3 more authors) (2025) Eggs-posed:revision of Schistosoma mansoni venom allergen-like proteins unveils new genes and offers new insights into egg-host interactions. BMC Genomics. 189. ISSN 1471-2164

https://doi.org/10.1186/s12864-025-11369-4

BMC Genomics

RESEARCH

Open Access

Eggs-posed: revision of *Schistosoma mansoni* venom allergen-like proteins unveils new genes and offers new insights into egg-host interactions

Lukáš Konečný[1,2*], Lucie Jedličková[1,3], Zaina Ibnahaten[4], Adam Roberts[4,6,7], Cecile Crosnier[4,7] and Jan Dvořák[2,5]

Abstract

Background Venom allergen-like proteins (VALs) are abundant in the excretory-secretory products (ESPs) of numerous parasitic helminths and have been extensively studied for over 30 years because of their potential to interact with host systems. Despite substantial research, however, the precise functions of these proteins remain largely unresolved. Schistosomes, parasites of the circulatory system, are no exception, with 29 SmVAL genes identified in the genome of *Schistosoma mansoni* to date. The eggs of these parasites, as primary pathogenic agents, interact directly with host tissues and release excretory-secretory products that aid their egress from the host. Although SmVALs have been detected in the egg secretome in the past, direct evidence of their secretion and functional interaction with host molecules has never been demonstrated. These findings fuel the ongoing debate as to whether egg-expressed SmVALs interact with the mammalian host or are rather miracidial proteins synthesized within the egg during larval development.

Results Based on complete revision of the SmVAL family and an associated robust transcriptomic meta-analysis of gene expression across the life cycle, we show that many of SmVAL genes, including 6 newly identified genes, are expressed in the infective larvae-producing stages (eggs and sporocysts). Following localization of two "egg-specific" SmVAL9 and SmVAL29 did not prove active secretion of these molecules into surrounding tissues but were aligned with miracidial structures interfacing with the molluscan host, specifically the larval surface and penetration glands. Finally, we show the complete lack of interactions between candidate SmVAL proteins and an array of 755 human cell receptors via a state-of-the-art SAVEXIS screen.

Conclusions Overall, we conclude that these "egg" SmVALs are not involved in direct host–parasite interactions in the mammalian host and are rather proteins employed during intermediate host invasion. Our study revisits and updates the SmVAL gene family, highlighting the limitations of in silico protein function predictions while emphasizing the need for up-to-date datasets and tools together with experimental validation in host-parasite interactions. By uncovering the diversity, expression patterns, and interaction dynamics of SmVALs, we open new avenues for understanding host manipulation and reevaluating orthologous proteins in other helminths.

Keywords Schistosomiasis, Venom allergen-like, Egg, Miracidium, Transcriptome, SCP, CAP, SAVEXIS

*Correspondence:
Lukáš Konečný
lukkas.konecny@gmail.com
Full list of author information is available at the end of the article

Introduction

Schistosoma mansoni is a major causative agent of schistosomiasis, for which more than 250 million individuals required preventive treatment in 2021 [1]. These parasitic flatworms are distinctive among trematodes in terms of their gonochoristic lifestyle, where adults form pairs within the host's vascular system and constantly lay eggs directly into the bloodstream. Successful egg excretion relies on parasite-induced granuloma formation, which aids in transporting eggs across the vascular and intestinal walls into the external environment [2]. The interaction between *S. mansoni* eggs and host tissues is believed to be facilitated mainly by egg excretory-secretory products (ESPs), which reportedly originate from the syncytial outer subshell envelope of the egg [2, 3]. To date, nearly 200 proteins have been identified in the *S. mansoni* egg secretome [4–6]. Of these however, only IPSE-1/alpha-1 and omega-1 have been intensively studied, and their immunomodulatory functions have been revealed [7–10] whereas other identified proteins remain enigmatic.

A significant group of alleged egg-secreted molecules are venom allergen-like proteins (VALs) [5, 11, 12]. The VAL protein family is part of the sperm coating protein/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS) superfamily (SMART: SM00198, Pfam: PF00188). All VALs are defined by a CAP domain in their core, which stands for cysteine-rich secretory proteins, antigen 5 proteins, and pathogenesis-related 1 proteins. The CAP domain is a structurally conserved, cysteine-rich, approximately 15 kDa domain with a typical α-β-α sandwich fold and encompasses the SCP/TAPS domain within its structure. This strong conservation of the tertiary structure could indicate shared common biochemical activity for all SCP/TAPS proteins [12–14]. Despite the general diversity in ESPs of various stages of helminths, VAL proteins are ubiquitous in these products and are likely specifically associated with the parasitic lifestyle [15].

To date, 29 VALs with complete SCP/TAPS domains have been identified in *S. mansoni* (SmVAL). The SmVAL superfamily is divided into two groups. Group 1 SmVALs were described as having 3 conserved disulfide bonds, an extended first loop region, and, importantly, predicted signal peptides indicating their active secretion. Group 2 representatives (SmVAL6, 11, 13, 16, and 17) lack these features but have unique conserved histidine and tyrosine residues [14]. The gene expression of SmVALs seems to be stage- and tissue-specific [14, 16–19], with several of these genes from group 1 being exclusive to the egg and miracidia stages [14, 16, 20]. Moreover, group 1 SmVALs were identified among the most highly expressed genes in mature eggs from both liver and intestinal tissues [21]. Owing to the assumption that these proteins are actively secreted from the egg, many functions have been attributed to them in the past, including possible anticoagulant, immunomodulatory and tissue remodelling functions [12, 22]. However, the lack of direct evidence fuels an ongoing debate regarding whether SmVALs are truly secreted from the egg to the mammalian tissues or are rather miracidial proteins expressed by the egg and the developing miracidium stage *in ovo* [13, 23, 24].

In this work, we sought to resolve this debate by localizing the putative secretion of these proteins from the egg into the surrounding tissues of the infected host. This is preceded by a complete revision and de novo identification of all SmVAL genes according to the recent genome version and a reanalysis of their sequences, annotations and phylogenetic relationships, which in the past were only adopted from seminal studies. Furthermore, we employed a high-throughput state-of-the-art protein–protein interaction screen to investigate whether egg-specific SmVALs can directly interact extracellularly with host cell receptors. This approach aimed to further confirm or challenge their potential functions within the mammalian host.

Materials and methods

Ethics statement

Research involving experimental animals was performed in accordance with the animal welfare laws of Czechia and under the European regulations for transport, housing and care of laboratory animals (CZ Act No. 246/1992, EU Directive 2010/63/EU amended by Regulation 2019/1010/EU). Experiments were approved by the Animal Welfare Committees of University of Life Sciences and the Ministry of Education, Youth and Sports of Czechia (MSMT-29299/2020–3). All animals used in the study were maintained by a certified person (Certificate Number CZ 02847) in an accredited facility. Ketamine-xylazine anaesthesia was employed to sacrifice the mice as indicated in "Parasite material" section.

Parasite material

The life cycle of the *S. mansoni* Puerto Rican strain is maintained in our laboratories using *Biomphalaria glabrata* as an intermediate host snail and C57BL/6JRj (Janvier Labs) female mice as the definitive host. The mice were infected in a water bath with 150 cercariae per mouse for one hour. The mice were sacrificed 7 weeks post infection by an intraperitoneal injection of ketamine (Narkamon 5%—1.2 ml/kg body weight) and xylazine (Rometar 2%—0.6 ml/kg body weight), with heparin to prevent blood clotting. The small intestines and livers were harvested and processed as described previously [21]. The infected hepatopancreases were dissected from the snails 6 weeks after infection with miracidia.

SmVALgenes de novo identification and analysis

To identify all *S. mansoni* genes coding for SmVAL proteins, the SCP/TAPS Hidden-Markov Model (HMM) consensus domain profile was built with hmmbuild on the basis of seed alignment acquired from Pfam (PF00188). The protein sequences for all *S. mansoni* protein-coding genes were acquired from the WormBase ParaSite database (*S. mansoni* genome PRJEA36577 v9; database version WBPS19) [25, 26]. Sequences containing the SCP/TAPS domain were identified via hmmscan and were manually evaluated to determine whether they contained the complete domain. Only those sequences with identified domains > 100 amino acids and domain-specific features such as conserved cysteines and residues at specific positions identified previously were included in the subsequent analyses [14, 27]. The whole peptide sequences of proteins identified as SmVALs were retrieved via BioMart at the WormBase ParaSite server [25]. The prediction of signal peptides and transmembrane elements was performed via Phobius [28]. Additionally, SignalP 6.0 [29] was used for comparison and presence of signal peptides was assumed when the probability score was ≥ 0.5. The amino acid positions bordering the peptide sequences identified as forming the CAP or SCP/TAPS domain (for phylogenetic analysis and domain alignment, respectively) served as input for esl-fetch, which was used to extract the respective domain sequences. The tools hmmbuild, hmmscan and esl-sfetch used for the genes and domain recognition are part of HMMER v3.4 software [30]. All relevant intermediate files are available at https://github.com/lukykony/SmVAL_identification).

Alignment and phylogenetic analysis

To increase the reliability of the phylogenetic analysis, the CAP domain sequences were used, as they are the most conserved part of the SCP/TAPS proteins, providing a stable basis for evolutionary comparisons. The amino acid sequences of 36 complete CAP domains encoded by 35 identified SmVAL genes were aligned with MAFFT v7.520 using the L-INS-I strategy [31]. The resulting alignment in fasta format was trimmed with trimAl using default settings to remove poorly aligned regions that would later affect the accuracy of the phylogenetic analysis [32]. The best-fitting model WAG + R3 was chosen on the basis of ModelFinder analysis as a part of IQtree analysis [33]. The maximum likelihood analysis was performed via IQtree v2.3.5 with bootstrap values based on 1000 replicates. The resulting phylogenetic tree was visualized with FigTree v1.4.4 software [34]. For sequence analysis, the amino acid sequences predicted to form SCP/TAPS domains were extracted. 36 SCP/TAPS-forming protein sequences were aligned with MAFFT v7.520 using the L-INS-I strategy, and the resulting multiple sequence alignment was manually inspected with Jalview 2 [35].

Gene expression of SmVALs throughout the parasite life cycle

The RNA-Seq datasets of selected parasite stages, which are freely available in the NCBI Sequence Read Archive (SRA) or European Nucleotide Archive (ENA), were processed so that the expression of each gene was directly comparable between samples and within the sample. This was achieved by selecting public RNA-seq datasets prepared using the same methodology, specifically library preparation via PolyA selection and paired-end sequencing layout. Three datasets for each stage (mammalian liver eggs, miracidia, sporocysts, cercariae, schistosomula and intramammalian parasitic stages collected at 6, 13, 17, 21, 28 and 35 days post infection) from previously published studies were selected for comparative gene expression analysis [16, 36] (see Additional file 1: Table S1 for details on used datasets). The counts of aligned reads per run were acquired from the WormBase ParaSite database [25, 36]. The read counts in the database were produced as described by Wangwiwatsin et al. (2020) [36]. Briefly, the acquired reads were mapped to the *S. mansoni* genome via TopHat version 2.0.8 and counted with HTSeq-count version 0.7.1 (for HTSeq counts see Additional file 2: Table S2). Gene length corrected trimmed mean of M-values (GeTMM) normalization, which allows for accurate inter- and intrasample comparisons, was performed in EdgeR according to a published methodology [37]. The resulting normalized GeTMM gene counts for each stage were averaged, and expression values for individual genes encoding SmVALs were retrieved via respective gene identifiers (for GeTMM normalised counts see Additional file 3: Table S3). The gene expression across the parasite life cycle was visualized as a heatmap created via the pheatmap R package. For better resolution of SmVAL expression in *S. mansoni* eggs, the data from immature and mature eggs from liver and intestinal tissues from our recent work were also analysed (data from Supplementary Table 2 from Peterková et al. (2024) [21]. Because cDNA library preparation for these datasets differed from that for the other publicly available datasets mentioned above, these data were assessed and visualized separately.

SmVAL genes amplification

To verify the sequences and in silico expression of the predicted novel SmVALs, genes from life stages where they are expected to be most highly expressed according to the gene expression meta-analysis were amplified via

conventional PCR. Total RNA was isolated from parasite eggs and sporocysts with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. The concentration of isolated RNA was measured on a NanoDrop (Thermo Fisher), and 1 µg of RNA per sample was treated with Turbo DNase (Ambion) to remove contaminating genomic DNA. cDNA was generated with a SuperScript IV RT kit (Invitrogen). The gene sequences of the SmVALs were retrieved from the WormBase ParaSite database, and specific primers were designed on the basis of CDSs (see Additional file 4: Document S1). PCR was performed with DreamTaq PCR Master Mix 2x (Thermo Scientific) according to the manufacturer's instructions. The resulting PCR amplicons were purified via the PCR Cleanup Kit (Geneaid) and cloned with the pGEM T Easy Vector system (Promega Corporation), and the correct gene sequences were confirmed via Sanger sequencing. SmVAL9 and SmVAL29, the two most highly expressed SmVALs throughout the life cyclxe, were chosen for further studies.

Recombinant production of SmVAL9 and SmVAL29 in E. coli

For polyclonal antibody production, SmVAL9 and SmVAL29 were heterologously expressed in an *E. coli* expression system. The coding sequences without signal peptides and stop codons were amplified with PCR using PrimeSTAR Max DNA Polymerase (Takara Bio) and cloned into the pET22b(+) vector in frame with a 6×His tag via an In-Fusion cloning kit (Takara Bio) (for details on constructs see Additional file 5: Document S2). The recombinant proteins were expressed in *E. coli* BL21 (DE3) cells in LB medium at 37 °C for 6 h upon 0.5 mM IPTG induction and further purified from inclusion bodies. Briefly, the cells were resuspended in 20 ml (per 200 ml culture) of lysis buffer (50 mM NaH2PO4, 0.3 M NaCl, and 10 mM imidazol, pH 8.0) and left at -80 °C overnight. The samples were subjected to three freeze–thaw cycles in an ice bath (methanol and dry ice) with 4×15 s of sonication after every cycle. The cell lysates were centrifuged at 12,000×g for 30 min at 4 °C, and the supernatant was discarded. The inclusion bodies were further purified via consecutive sonication-assisted solubilization via incubation in 50 mM Tris base, 2 M urea, 150 mM NaCl, and 1 mM EDTA, pH 8.0, followed by incubation in 50 mM Tris base, 150 mM NaCl, 1 mM EDTA, and 0.1% Triton-X 100, pH 8.0 for 10 min, after which both solubilization steps were followed by 4×15 s sonication and 30 min of centrifugation at 12,000×g. Then, the pellets were repeatedly washed with two volumes of dH2O, followed by centrifugation until the supernatant became clear. The resulting inclusion bodies were dissolved in equilibration buffer (50 mM phosphate buffer with 300 mM NaCl and 10 mM imidazole, pH 8.0) and centrifuged at 12,000×g for 30 min. The supernatant was loaded onto a 5 ml Ni–NTA column. The bound rSmVAL proteins were washed with 10 volumes of equilibration buffer with 20 mM imidazole and eluted with equilibration buffer with 250 mM imidazole. Fractions containing the purified proteins were pooled and used to prepare polyclonal mouse antisera.

Polyclonal antibody production

Polyclonal antibodies against rSmVAL9 and rSmVAL29 were produced in NMRI mice by injecting 30 µg of purified proteins in TiterMax Gold adjuvant (Sigma–Aldrich) according to the manufacturer's instructions intraperitoneally three times at intervals of 14 days. The immunized mice were bled under deep ketamine/xylazine anaesthesia 5 days after the last injection, and the resulting antisera were collected via centrifugation. Control serum samples were collected from the mice prior to immunization. The specificity of the produced antibodies was confirmed by immunoblotting following established protocol [38].

Immunolocalization of SmVALs in eggs and surrounding host tissues

For immunolocalization of the SmVALs in the parasite eggs and surrounding tissues, liver and small intestine samples were collected from female C57BL/6JRj mice (Janvier) at 7 weeks post infection. The organs were rinsed in PBS and then fixed in 4% neutral buffered formalin for 3 days. The fixed tissues were dehydrated through a graded ethanol series (50%, 70%, 90%, 96%, and 100% v/v ethanol), with each step lasting 20 min. Following dehydration, the tissues were cleared in xylene (VWR) twice for 15 min each. The cleared tissues were then embedded in Paraplast (Sigma–Aldrich). Sections (5 µm thick) were prepared via a Shandon Finesse® ME+ microtome (Thermo Fisher Scientific) and mounted onto X-tra adhesive slides (Leica). The paraffin-embedded sections were deparaffinized and rehydrated. Antigen retrieval was performed by boiling the sections in 0.05 M citrate buffer (pH 6.0) containing 0.05% Tween 20 in a microwave oven (500 W) for three cycles of 3 min each, followed by a 20-min cooling period. The sections were then blocked with buffer containing PBS, 2% BSA, 0.25% Triton X-100, and a goat anti-mouse polyclonal antibody (Abcam, ab6668) diluted 1:50 for 1 h. After blocking, the sections were washed three times in PBS-T (PBS with 0.25% Triton X-100) for 5 min each. The sections were then incubated overnight at 4 °C in a wet chamber with immunized mouse serum against SmVAL9 or SmVAL29 diluted 1:50 in an antibody mixture (PBS, 1% BSA, 0.25% Triton X-100). Preimmunized

mouse serum was used as the negative control. Following primary antibody incubation, the sections were washed three times in PBS-T for 5 min each. The slides were then incubated with goat anti-mouse IgG (H + L) conjugated with Alexa Fluor® Plus 647 (Invitrogen, A48289) diluted 1:500 in PBS-T with 1% BSA for 1 h in the dark. The sections were subsequently washed three times in PBS-T for 5 min each and mounted with Fluoroshield containing DAPI (Sigma–Aldrich). The fluorescence signals were visualized via a Leica TCS SP8 confocal microscope. The background signal threshold was determined via slides probed with control serum.

SAVEXIS protein–protein interaction screen

To identify whether egg/miracidium-specific SmVALs could directly interact with mammalian host tissues, a resource of 755 recombinant human surface receptors was used for protein–protein interaction screening. Recombinant monobiotinylated *S. mansoni* SmVALs and human receptor proteins were produced via transient transfection of HEK293-6E cells grown in Freestyle 293 media supplemented with 0.1% Kolliphor, 0.5 µg/mL G418 and 48 µg/mL D-biotin and cotransfected with a plasmid encoding the BirA biotin ligase as previously described [39]. Proteins were purified via His MultiTrap plates with a previously described purification platform and pneumatic press [40], followed by confirmation of expression via protein gel electrophoresis and western blotting with HRP-streptavidin (Jackson ImmunoResearch, 016–030-084, 0.2 µg/mL). Protein concentrations were measured via the Bradford assay. To create "preys" for SAVEXIS, SmVAL proteins were normalized to 17.5 nM and tetramerised with PierceHigh Sensitivity Streptavidin-HRP (Thermo Scientific, 21,130) as previously described. The human protein library was assembled from 755 proteins primarily expressed on the surface of leukocytes [41] with additional proteins from erythrocytes [42] and megakaryocytes [40] as proteins accessible to extracellular parasite antigens and arrayed on 384-well streptavidin-coated microtiter plates [41]. The preys were incubated with the baits for one hour and then washed in HBS-T with desthiobiotin; the host–parasite interactions were then revealed with a TMB/E chromogenic substrate. The plates were measured with a Tecan Spark plate reader at an absorbance of 650 nm as previously described [41].

Results

Revision of the SmVAL family revealed 6 novel SmVAL genes

The first objective of this study was to uncover a complete repertoire of genes encoding *S. mansoni* VAL proteins to characterize their sequence features on the basis of the latest databases and bioinformatics tools. The search using the SCP/TAPS HMM profile revealed 40 transcripts, 38 of which contained the complete CAP domain. The incomplete SCP/TAPS domains of the identified transcripts Smp_345050.1 and Smp_176170.1 are only 26 and 73 amino acids long, respectively, therefore they lack critical conserved sites and are not expressed at any stage of the parasite according to our analyses. In addition to the SmVAL1-28 genes already described by Chalmers et al. (2008) [14] and SmVAL29 identified by Wu et al. (2009) [24], our analysis revealed 6 previously undescribed members of this family. In addition to these newly identified genes, one previously undiscovered splice variant, SmVAL11.2, was detected, where the sixth exon not transcribed, thus the resulting protein product is 23 amino acids shorter than that in SmVAL11.1. The new genes were named on the basis of the numerical order of their Smp identifiers (Smp_131370—SmVAL30, Smp_200460—SmVAL31, Smp_300080—SmVAL32, Smp_313710—SmVAL33, Smp_317520—SmVAL34 and Smp_347320—SmVAL35). These newly discovered genes, except for SmVAL32, were successfully amplified from the cDNA stages where these genes are the most expressed based on our gene expression analysis across the life cycle (SmVAL30, SmVAL34 and SmVAL35 from sporocysts and SmVAL31 and SmVAL33 from eggs). SmVAL32 is identical to SmVAL7; therefore, it was not possible to distinguish these genes via conventional PCR. However, for example, the successful amplification and resolution of the SmVAL34 and SmVAL35 genes, which share 93% identity, convinced us that SmVAL32, like the other newly discovered genes, is real and not just an artifact of the genome assembly. The updated SmVAL gene family and gene characteristics, including the new nomenclature, can be found in Table 1. A comparison of the current state of knowledge with previously published data is shown in Additional file 6: Table S4.

Updated analyses of SmVAL proteins confirm two distinct groups with unique features, challenging historical views on signal peptides

Updated phylogenetic analysis of all CAP domains of the previously known and newly identified SmVALs via the maximum likelihood method strongly supported the previously published division of this family into two distinct phylogenetic branches (100% bootstrap value) and proposed division of these proteins into group 1 and group 2 [13, 14, 23] (Fig. 1). Accordingly, 5 newly identified SmVAL genes belong to group 1 (SmVAL30, SmVAL32, SmVAL33, SmVAL34 and SmVAL35), and one belongs to group 2 (SmVAL31). In line with previously published phylogenetic analyses [13, 14, 23], the second domain of SmVAL11 belongs to group 2 according to our phylogenetic analysis. Moreover, the division of two groups is

Table 1 Overview and characteristics of an updated SmVAL family

Name	Chromosome	Gene ID	Transcript ID	Gene start (bp)	Gene end (bp)	Gene length	Protein length	SP/TM (Phobius)	Group
Previously identified SmVALs									
SmVAL1	6	Smp_193680	Smp_193680.1	12,437,483	12,439,150	1667	234	YES	1
SmVAL2	6	Smp_002630	Smp_002630.1	12,468,974	12,470,914	1940	229	YES	1
SmVAL3	6	Smp_316760	Smp_316760.1	17,278,025	17,279,298	1273	178	N/O	1
SmVAL4	6	Smp_002070	Smp_002070.1	4,187,554	4,193,626	6072	181	YES	1
SmVAL5	6	Smp_336880	Smp_336880.1	17,331,374	17,335,886	4512	184	YES	1
SmVAL6.1	1	Smp_124050	Smp_124050.1	11,722,354	11,790,303	67,949	429	N/O	2
SmVAL6.2	1	Smp_124050	Smp_124050.2	11,722,354	11,790,303	67,949	423	N/O	2
SmVAL6.3	1	Smp_124050	Smp_124050.3	11,722,354	11,790,303	67,949	414	N/O	2
SmVAL7	6	Smp_300070	Smp_300070.1	17,323,751	17,327,231	3480	193	TM	1
SmVAL8	6	Smp_123550	Smp_123550.1	12,477,072	12,480,198	3126	249	TM	1
SmVAL9	6	Smp_176180	Smp_176180.1	16,367,696	16,373,800	6104	182	YES	1
SmVAL10	6	Smp_002060	Smp_002060.1	4,171,241	4,183,900	12,659	170	YES	1
SmVAL11.1	2	Smp_012350	Smp_012350.1	25,764,239	25,784,550	20,311	400	N/O	2
SmVAL11.2	2	Smp_012350	Smp_012350.1	25,764,239	25,784,550	20,311	423	N/O	2
SmVAL12	6	Smp_123540	Smp_123540.1	12,487,474	12,491,566	4092	204	TM	1
SmVAL13	1	Smp_124060	Smp_124060.1	11,676,364	11,702,352	25,988	236	N/O	2
SmVAL14	1	Smp_078490	Smp_078490.1	2,040,271	2,042,090	1819	219	YES	1
SmVAL15	6	Smp_070250	Smp_070250.1	17,308,316	17,312,821	4505	248	YES	1
SmVAL16	1	Smp_124070	Smp_124070.1	11,644,064	11,669,918	25,854	169	N/O	2
SmVAL17	1	Smp_331830	Smp_331830.1	19,494,075	19,508,994	14,919	157	N/O	2
SmVAL18	6	Smp_001890	Smp_001890.1	3,518,374	3,526,057	7683	194	TM	1
SmVAL19	6	Smp_123090	Smp_123090.1	3,503,241	3,510,772	7531	186	YES	1
SmVAL20	6	Smp_127130	Smp_127130.1	22,505,531	22,514,880	9349	225	N/O	1
SmVAL21	6	Smp_159290	Smp_159290.1	12,417,430	12,419,089	1659	234	YES	1
SmVAL22	4	Smp_139450	Smp_139450.1	2,688,513	2,690,855	2342	219	YES	1
SmVAL23	6	Smp_160250	Smp_160250.1	17,281,855	17,283,130	1275	182	YES	1
SmVAL24	6	Smp_141550	Smp_141550.1	17,213,661	17,215,006	1345	195	YES	1
SmVAL25	6	Smp_141560	Smp_141560.1	17,239,323	17,240,623	1300	195	YES	1
SmVAL26	6	Smp_154260	Smp_154260.1	16,413,349	16,416,075	2726	182	YES	1
SmVAL27	6	Smp_154290	Smp_154290.1	16,593,216	16,595,909	2693	182	YES	1
SmVAL28	6	Smp_176160	Smp_176160.1	16,378,302	16,381,034	2732	182	YES	1
SmVAL29	6	Smp_120670	Smp_120670.1	16,391,667	16,396,592	4925	153	N/O	1
Newly identified SmVALs									
SmVAL30	6	Smp_131370	Smp_131370.1	21,769,518	21,776,061	6543	208	N/O	1
SmVAL31	1	Smp_200460	Smp_200460.1	4,030,097	4,049,167	19,070	197	N/O	2
SmVAL32	6	Smp_300080	Smp_300080.1	17,338,761	17,341,822	3061	193	TM	1
SmVAL33	6	Smp_313710	Smp_313710.1	17,299,341	17,300,791	1450	178	N/O	1
SmVAL34	6	Smp_317520	Smp_317520.1	17,237,013	17,238,811	1798	185	YES	1
SmVAL35	6	Smp_347320	Smp_347320.1	17,210,799	17,213,179	2380	185	YES	1

SP/TM – signal peptide/transmemembrane domain – YES marks signal peptide, N/O marks the lack of signal peptide, TM marks the presence of transmembrane domain

also supported by the pairwise identity matrix (see Additional file 7: Figure S1).

The division of SmVALs into two groups has historically been associated with the presence of signal peptides in group 1 and their absence in group 2. The prediction of signal peptides via the latest version of SignalP 6.0 indicated that the previously described Group 1 SmVAL3, SmVAL5, and SmVAL20 lack this feature together with the newly analysed Group 1 SmVAL29, SmVAL30 and SmVAL33. Further analysis with Phobius revealed that the

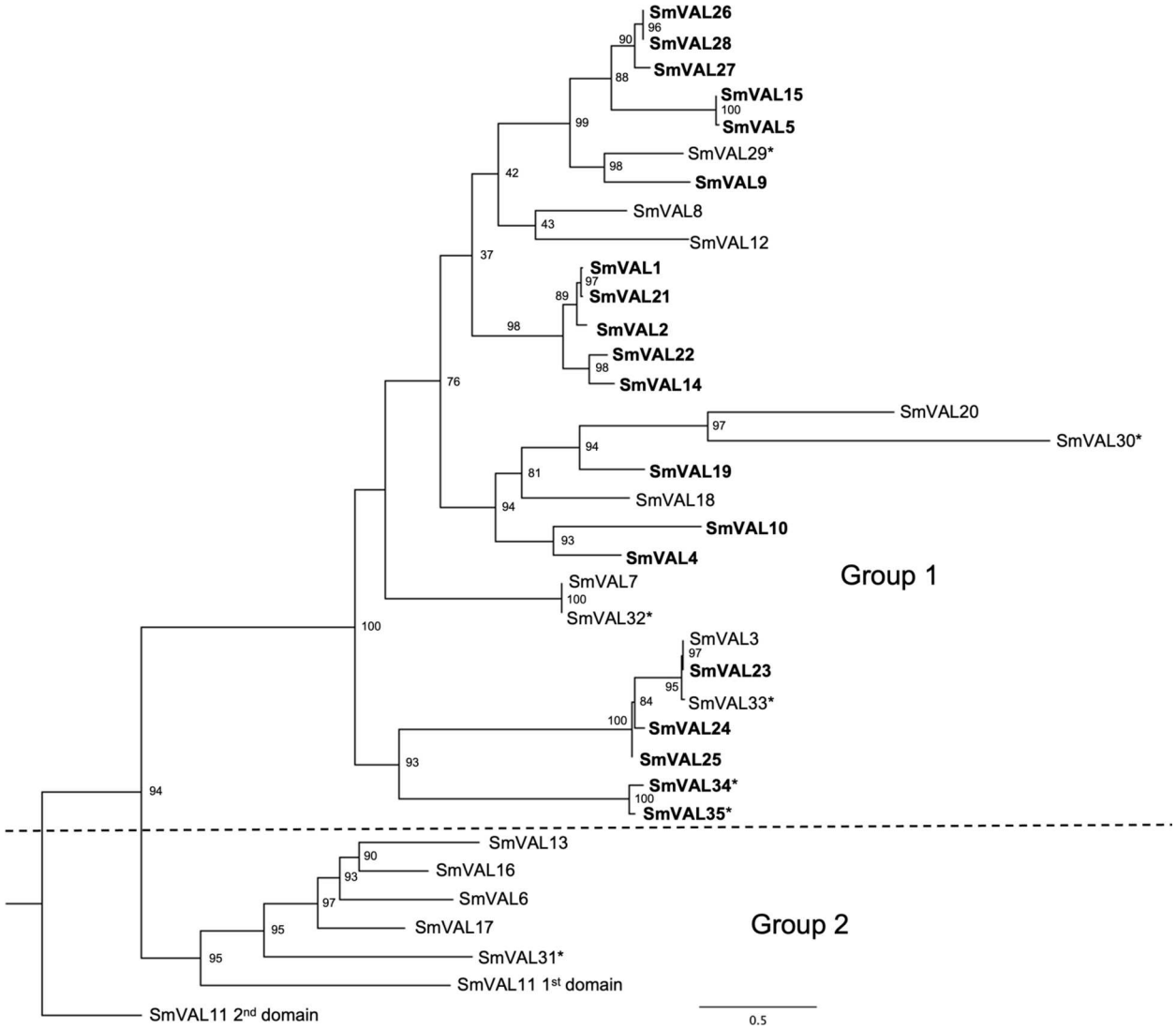

Fig. 1 Phylogenetic analysis of *S. mansoni* venom allergen-like protein CAP domains. The updated phylogenetic tree supports an evolutionary split between two distinct branches of previously proposed group 1 and group 2 proteins. The relative branch lengths are indicated by the scale bar. Newly identified SmVALs within this study and SmVALs put in the phylogenetic analysis are marked with an asterisk. SmVALs containing signal peptides within the sequence according to Phobius analysis are in bold

signal peptides of Group 1 SmVAL7, SmVAL8, SmVAL12, SmVAL18 and SmVAL32 are in fact transmembrane domains. This finding contradicts previously published data; however, our updated multiple sequence alignment further supported the division of the SmVAL family into two distinct groups. All five newly identified group 1 SmVALs contain 6 disulfide bond-forming cysteines characteristic of this group. Additionally, SmVAL34 and SmVAL35, which are closely phylogenetically related, contain two extra cysteine residues at positions 21 and 66 within the SCP/TAPS domain, which might indicate specific folds of these two proteins and possibly distinct shared functions (Fig. 2). In addition to the previously described common deletion in the SCP/TAPS domain for all old and new members of group 1, the typical deletion for all the members of group 2 is also present.

Detailed analysis of individual identified gene features also revealed differences in the length of genomic sequences for most of the previously described SmVALs (Table 1). In addition, three of the 13 coding sequences verified by PCR and deposited in GenBank (SmVAL1-13) by Chalmers et al. (2008) [14] have different coding sequences according to WormBase ParaSite prediction. While the predicted CDS of SmVAL1 according to the

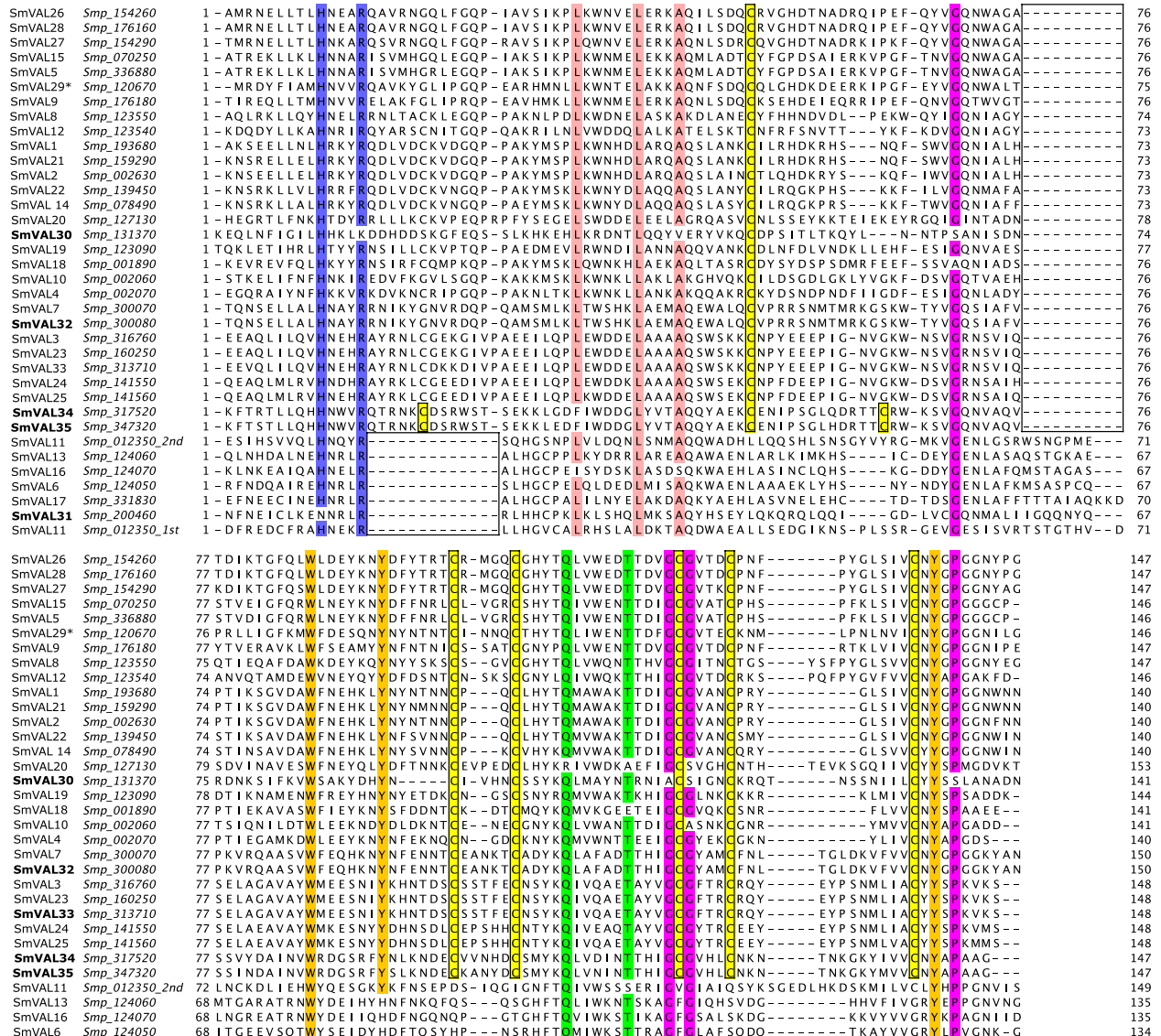

Fig. 2 Multiple sequence alignment of 36 *S. mansoni* SCP/TAPS domains. Newly identified SmVALs within this study are in bold. The asterisk at SmVAL29 signifies that although identified before this study, it has been first time analysed. The black bar below the sequence alignment indicates CAP domain location. Highly conserved residues (≥ 90% identity threshold) are highlighted based on physicochemical properties (aliphatic/hydrophobic—pink, aromatic—orange, positive—blue, negative—red, hydrophilic—green, conformationally special—magenta). Conserved cysteine residues are highlighted in yellow. Distinct deletion regions typical for previously proposed group 1 and group 2 SmVALs are boxed. The alignment was processed in Jalview [35]

latest genome annotation differs by only 9 nucleotides from the original characterization, the SmVAL3 and SmVAL5 CDSs are now shorter at the 5' end by 105 bp and 66 bp, respectively. The complete comparison of previously published characterizations of individual genes with the current analysis we present in this work, including the changes in nomenclature, can be found in Additional file 6: Table S4.

Life cycle expression patterns of SmVAL genes highlight dominance in early larval stages and stage-specific profiles for SmVAL9 and SmVAL29

After SmVAL family members were identified according to the recent genome version, it was important to determine gene expression across the life cycle to identify which of these genes are specific for the egg stage. This was accomplished via comparative bioinformatic

analysis of publicly available gene expression datasets for each stage. In selecting the datasets, emphasis was placed on covering the widest possible range of life stages while maintaining the comparability of these datasets with respect to the methodology of their preparation and appropriate data standardization. For this purpose, GeTMM normalization was used, as it allows not only a direct comparison of relative gene expression between stages but also of individual genes within a stage. According to our analysis, SmVALs are relatively highly expressed at the miracidia stage, accounting for more than 41% of all SmVAL transcripts cumulatively across the life cycle. The second most active stage in terms of expression was the egg stage, with just under 21% of the transcripts being expressed, and the third stage was the sporocyst stage, with just over 12% of the total relative SmVAL expression. For the remaining life stages analysed, the relative expression of SmVAL genes was very low, ranging from as low as 1.31% in 21 dpi juveniles to 5.84% in cercariae (see Additional file 8: Table S5).

As Chalmers et al. (2008) [14] and Farias et al. (2019) [23] revealed through qPCR analysis in their seminal work, our meta-analysis of publicly available RNA-Seq data pointed to both universal and development-specific expression patterns of various SmVALs across different life cycle stages (Fig. 3A). First, of a total of 35 genes, SmVAL20, SmVAL26, SmVAL27, SmVAL28 and SmVAL30 do not seem to be transcribed at any life stage (< 10 GeTMM normalized counts). In contrast, SmVAL11, the only member of the SmVAL family that contains two SCP/TAPS domains, appears to be consistently expressed regardless of developmental stage. A similar trend was observed for SmVAL6 and SmVAL7. In addition, the newly identified SmVAL32 gene was also relatively universally expressed, except at the egg and miracidium stages. In terms of developmentally specific expression, the sporocyst is a stage where the most stage-specific SmVALs are expressed. While SmVAL14, SmVAL19, SmVAL21, SmVAL22 and SmVAL35 are expressed relatively weakly, SmVAL1 and SmVAL34 expression is notable (between 100 and 200 GeTMM normalized counts); SmVAL4, SmVAL10, and SmVaL18 are among the most expressed SmVAL genes across the life cycle (> 250 GeTMM normalized counts). Another developmentally specific expression pattern was observed for SmVAL8 and SmVAL12, which formed

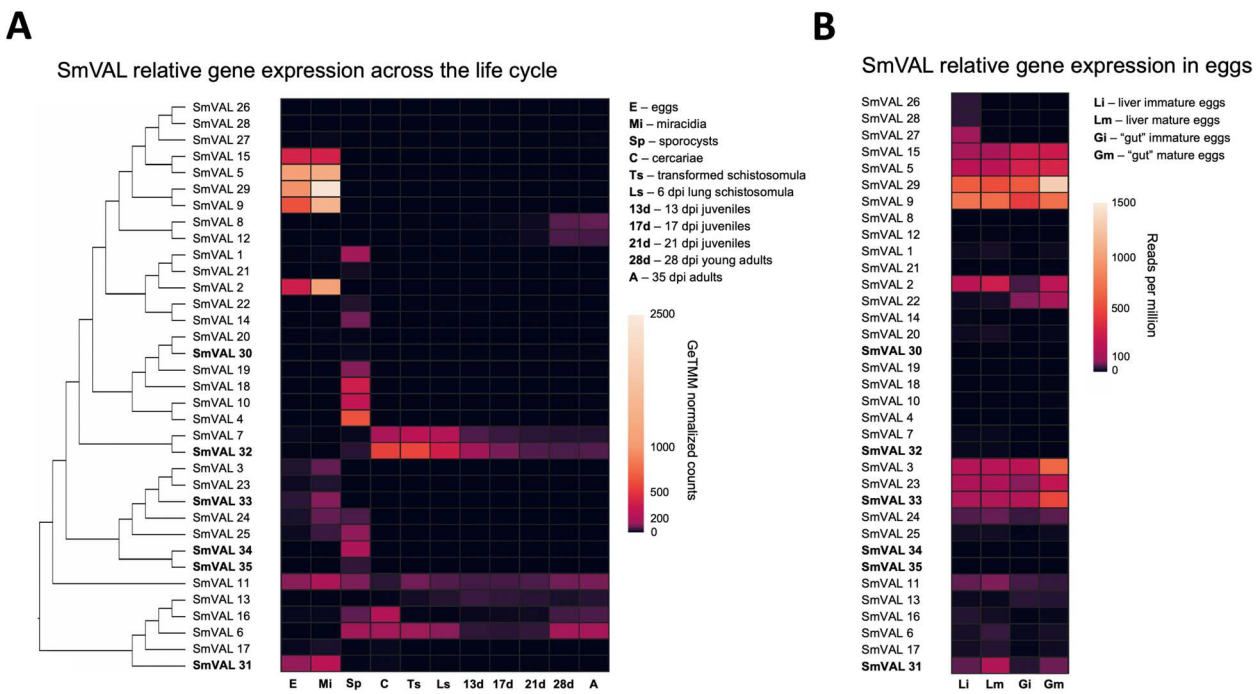

Fig. 3 **A** Relative gene expression of SmVALs across the *S. mansoni* life cycle. The three datasets of expression counts for each stage were obtained as described in the Methods section. The heatmap shows the means of GeTMM-normalized counts, allowing for intra- and intersample comparisons. Newly identified SmVALs within this study are in bold. To add phylogenetic context to expression patterns across the life cycle, a proportionally transformed tree from Fig. 1 was added. **B** Gene expression of the SmVAL family compared between immature and mature *S. mansoni* eggs from the intestinal and liver tissues of experimentally infected mice. The heatmap shows the mean gene expression levels normalized to reads per million (RPM), which highlights the proportional contributions of each subgroup to the overall gene expression profile in a given sample. The expression data were obtained from Supplementary Table S2 of Peterková et al. 2024 [21]

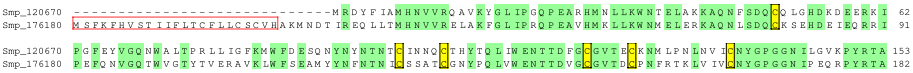

Fig. 4 Pairwise sequence alignment of SmVAL29 and SmVAL9. Identical residues are highlighted in green, and the predicted signal peptide between positions 24 and 25 according to Phobius analysis is indicated by a red box. Conserved cysteine residues are highlighted in yellow

a common branch according to our updated phylogenetic analysis. These two genes are the only SmVALs exclusively expressed in juvenile and adult worms in the definitive host. The expression of SmVAL8 and SmVAL12 is first detectable in the third week after host infection, and their levels increase markedly in the fourth week in freshly paired adults, reaching a maximum at the fifth week after infection in egg-laying adults. The most striking trend in our analysis was the observed combination of the abundance and intensity of SmVAL expression in the eggs and miracidia. A total of 9 different SmVALs were specifically expressed at these stages. While this stage-specific expression has been previously described for SmVAL2, SmVAL3, SmVAL5 and SmVAL9, our complete comparative analysis of all 35 SmVAL genes revealed that SmVAL15, SmVAL23, SmVAL29, SmVAL31 and SmVAL33 also follow this pattern. Furthermore, SmVAL2, SmVAL5, SmVAL9 and SmVAL29 presented the highest expression compared with the other genes in the whole family. The most highly transcribed gene from this family is SmVAL29, with 2344 GeTMM normalized counts in miracidia, almost twice as high as the second most highly expressed SmVAL of all, SmVAL9, and many fold greater than most of these genes throughout the life cycle. Together with SmVAL5, these two genes were also the most transcriptionally active at the intramammalian egg stage (Fig. 3B).

Since our recent study revealed that *S. mansoni* egg gene expression is critically dependent on the developmental stage of the egg and, more importantly, the tissue of origin [21], we also analysed SmVAL gene expression within these datasets (further referred to as "organ datasets"). These data could not be directly compared with the rest of the datasets used because of different library preparations and were therefore analysed separately. When examining the overall pattern of expression in different groups of eggs, it is clear that the expression patterns of the most highly expressed genes correspond to the meta-analysis where mixed liver eggs were used. These genes are primarily SmVAL2, SmVAL5, SmVAL9, SmVAL15 and SmVAL29.

However, some genes have remained undetected due to the limited resolution of life cycle expression analyses. While no expression of SmVAL26 or SmVAL28 was detected in the GeTMM-normalized datasets, low expression in immature liver eggs was detected within the organ egg datasets. SmVAL27 was also clearly detected in liver immature eggs, although it was present in very low numbers in the life cycle expression meta-analysis (see Additional file: 2 Table S2) and this signal was lost due to data normalization. Another example is SmVAL22, which is expressed only in intestinal eggs; therefore, its expression was not detected in the life-cycle analysis because the source study for the egg dataset used only liver eggs [16]. This is likely the reason why the expression levels of SmVAL3 and SmVAL33, which are prominent in intestinal eggs within organ datasets, are only weakly evident in the life cycle analysis. In the organ datasets, SmVAL9 and SmVAL29 were also the most highly expressed SmVALs, which were steadily expressed from immature to mature eggs in both liver and intestine of the host. Compared with that of SmVAL9, the gene expression of SmVAL29 in intestinal eggs increased even more markedly at maturity (Fig. 3B).

The first secretomic study identified SmVAL9, which has a signal peptide, in egg secretions and this protein has also previously been hypothesized to assist in extracellular matrix (ECM) remodelling during the migration of eggs through tissues [5, 22]. On the other hand, the closely related SmVAL29 protein lacks a signal peptide (Fig. 4), has not been detected in egg secretions, and its function is unknown. Both SmVALs are the most dominantly expressed in eggs and miracidia, and for these reasons, these two proteins were selected for further studies.

The most transcribed SmVAL9 and SmVAL29 are localized to miracidial glands and surface with no detected secretion from eggs

For the immunolocalization of the most highly expressed SmVAL9 and SmVAL29 in eggs and miracidia, antisera were raised against proteins produced recombinantly in the bacterial system. Primarily mature eggs were selected for immunolocalization, as these are the eggs that should be the most active in terms of secretion [3, 43]. Our experiments revealed that the tissue localization of both proteins in eggs and miracidia is identical. Neither SmVAL9 nor SmVAL29 were observed to be secreted into the surrounding liver or intestinal tissues, nor did the signal correspond to the egg subshell envelope, where the egg-secreted proteins are believed to be synthesized [3]. In contrast, on the basis of the atlas of *S. mansoni* organs [44], the observed localizations corresponded

with the lateral (penetration) glands of the miracidium, in both intestinal and liver sections. These effects were noticeably observed in both transverse and longitudinal sections through the egg/miracidium. Moreover, anti-SmVAL9 and anti-SmVAL29 immunoreactivity was also clearly associated with the surface of developed miracidia larvae (Fig. 5), again in sections from both intestine and liver tissues. No significant signal was observed in the samples treated with the control sera (Fig. 5).

SmVAL9 and SmVAL29 do not interact with the array of human cell receptors

Because our immunolocalization experiments did not indicate that SmVAL9 and SmVAL29 are secreted from the egg, we decided to support these findings by screening for potential protein interactions between the parasite and its mammalian host. The *S. mansoni* SmVAL proteins SmVAL9 and SmVAL29 were produced recombinantly in HEK293-6E cells and tested with SAVEXIS against a library of 755 human cell surface receptors. In these assays, no significant interactions were observed between the parasite proteins and human receptors, as shown by the lack of blue signals in any of the wells aside from the positive controls and a mutated LAIR1 receptor (containing T67L, N69S, and A77T mutations; well M23 of plate B), which was proven to bind indiscriminately to many different proteins in the SAVEXIS screens and was therefore excluded from the final results (Fig. 6). These observations support our findings that these proteins are not secreted into the vicinity of the egg and are more likely function in the miracidium stage. For readings for each well see Additional file: 9 Table S6.

Discussion

Our comprehensive study of VAL in *S. mansoni* proteins challenges existing assumptions about the roles of SmVALs previously attributed to the egg stage, particularly in relation to mammalian host interactions. Initial hypotheses about numerous egg SmVALs were largely based on their detection in the early secretome [5] leading to assumptions that these proteins might contribute to mammalian host immunomodulation and tissue remodelling and aid in egg migration [12–14, 21–23]. However, our findings challenge this view, highlighting the complexity of the SmVAL family and raising new questions about their precise roles in host invasion and tissue interactions.

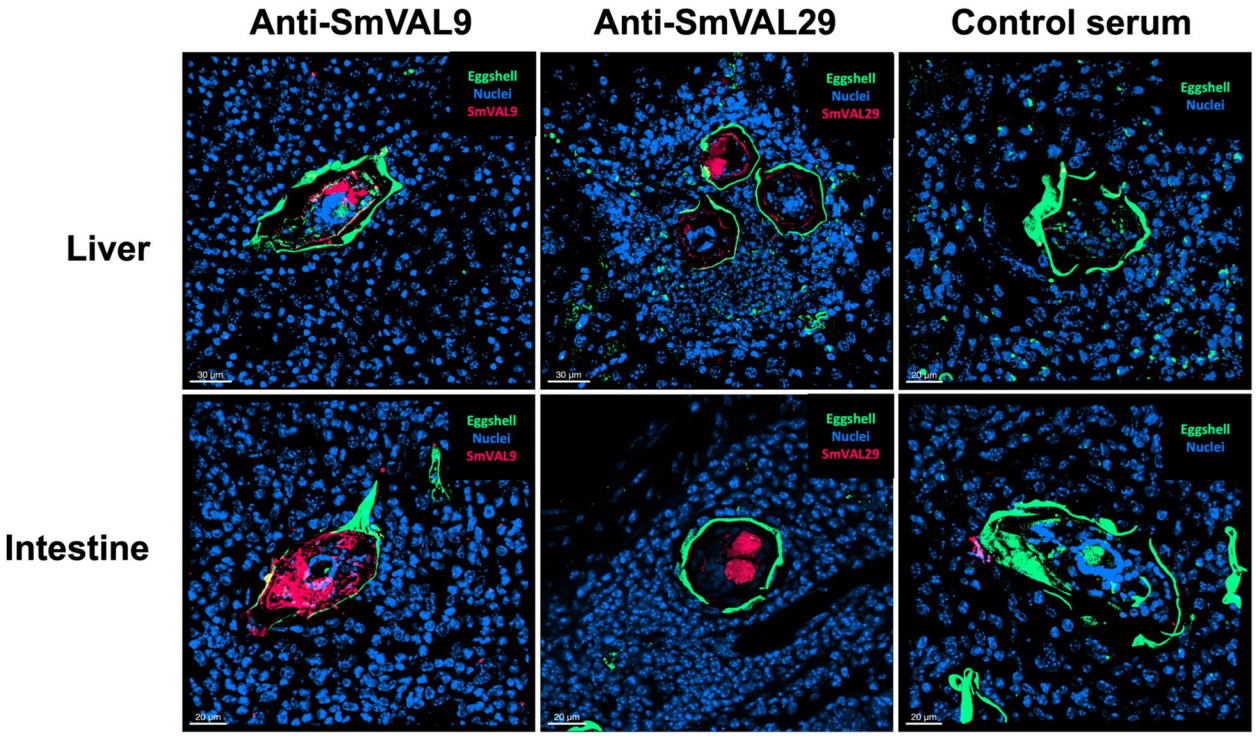

Fig. 5 Immunolocalization SmVAL9 and SmVAL29 in the sections of *S. mansoni* infected mouse liver and intestine. Confocal close-ups illustrate mature eggs with developed miracidia. The red signal represents anti-SmVAL serum-specific sites, the green signal indicates eggshell autofluorescence, and the blue signal indicates the DAPI-labelled cell nuclei. Both SmVAL proteins are localized in the miracidia penetration glands and the surface of the developed larvae. There was no specific protein signal in the control serum obtained from an non-immunized mouse

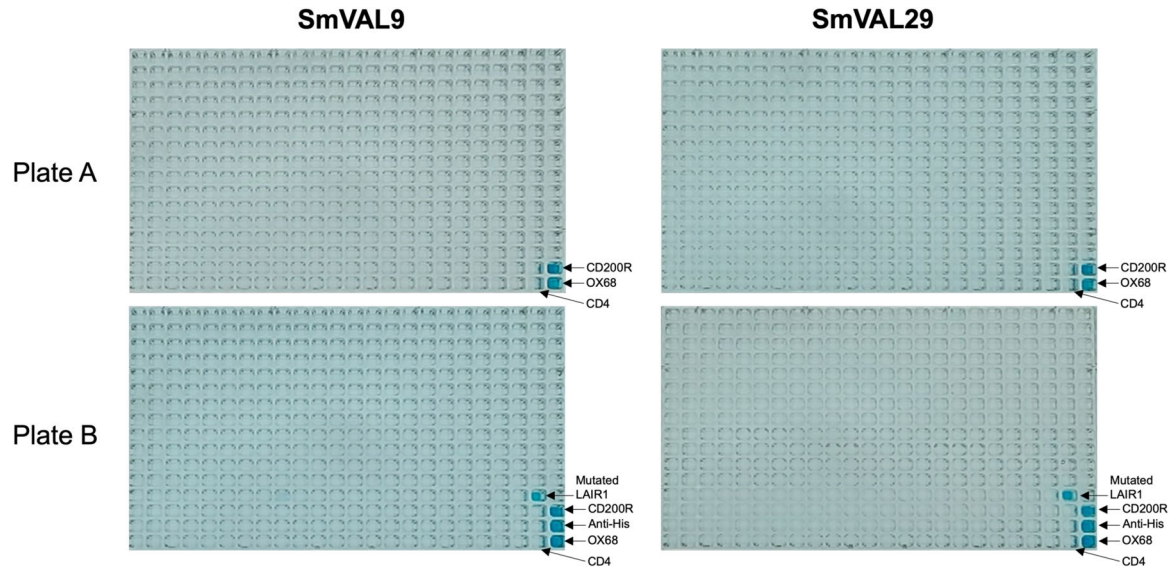

Fig. 6 SAVEXIS plates for SmVAL9 and SmVAL29 preys tested against a library of 755 human receptors. A library of 755 human receptors was arrayed across two 384-well plates (plates A and B). Controls are located in the bottom right corner of each plate. OX68 (well P24 on plates A and B) and anti-His (well O24 on plate B) are two biotinylated antibodies used as baits and target the Cd4 tag and 6 × His tag of the recombinant proteins, respectively. They act as positive controls by capturing prey through the relevant tags. CD4 (well P23 on plates A and B) corresponds to the biotinylated rat Cd4 (d3 + 4) tag used as bait. It acts as a negative control, and any signal observed in this well is considered background noise from the prey. CD200R (well O24 in plate A and N24 in plate B) is a positive control interaction between rat Cd200, which is used as bait, and rat Cd200R, which is used as prey. A mutated human LAIR1 ectodomain (well M23 in plate B) was included as bait in the SAVEXIS assay and proved to be very promiscuous, binding to > 75% of all preys tested in the assay

Since the last SmVAL family revision, the *S. mansoni* genome assembly has advanced significantly [26], as has the capacity and accuracy of bioinformatic tools [25]. Using an updated genomic dataset, we identified six previously undescribed SmVAL genes, notably expanding the known family diversity. Consequent sequence analysis of the revised family challenges the classification of SmVALs on the basis of signal peptide presence, which was previously used to divide SmVALs into secretory (Group 1) and nonsecretory (Group 2) categories, reflecting their phylogenetic relationships [12–14, 23]. Using the most advanced version of SignalP (v6.0), we found that previous assumptions about group features based on SignalP 3.0 may have been flawed by an early version of the tool. Consequent analysis with the Phobius tool, which enables the differentiation of signal peptides and transmembrane elements, revealed that in several SmVAL group 1 proteins these elements were likely confused, leading to incorrect conclusions. Even though predictive tools are not always completely accurate and extra care must be taken when interpreting in silico data (reviewed in [21]), we believe that our combined approach yields more accurate insights. Together, our revision suggests that earlier assumptions about these proteins and their potential secretion and functional roles of these proteins may require substantial reconsideration while highlighting the importance of regular re-evaluation as genomic datasets and prediction tools improve.

Following SmVAL family revision and sourcing high-quality data for eggs and miracidia from existing databases, along with the application of an innovative normalization method [37], we created a new gene expression atlas for both the updated SmVAL family and all *S. mansoni* genes. Although the expression profiles of SmVALs throughout the life cycle have been investigated relatively recently via RT–PCR [23], the expression in the egg stage was omitted, as were expression profiles of newly identified genes presented in this study. Additionally, RT–PCR may not be reliable for comparing gene expression across schistosome stages because of the lack of stable reference genes [45]. In contrast, RNA-Seq-based approaches have proven invaluable, but high-quality egg and miracidia RNA-Seq data have long been lacking in current gene expression atlases [16, 46]. In addition to supporting previous hypotheses that SmVALs likely play distinct roles throughout schistosome development, the pronounced expression of distinct SmVALs in eggs and sporocysts (stages that produce invasive

larvae) indicates that these proteins likely play a role in host invasion.

That SmVALs are likely to be employed during host invasion rather than host exit is also supported by our localization experiments. SmVAL9 and SmVAL29, the most highly expressed SmVALs in eggs, are concentrated within miracidial penetration glands and the larval surface which are critical for the establishment of infection in the molluscan intermediate host. Preceding studies have suggested that SmVAL proteins expressed in eggs are actively secreted into mammalian host tissues on the basis of their detection in excretory–secretory products (ESPs). However, while SmVAL2, SmVAL3, SmVAL5 and SmVAL9 were identified in the first proteomic study of egg ESPs [5], they were not identified in later egg secretomic studies [4, 6]. In contrast, SmVAL2, SmVAL3 and SmVAL9 were detected during miracidium-to-sporocyst transformation, indicating their role in molluscan invasion [24]. Additionally, SmVAL9 was previously localized within the miracidium, and its proposed role in snail invasion this protein has been attributed a possible role in mammalian tissue remodelling and egg translocation [22]. However, this protein was only localized in the parenchyma of the miracidium, and no evidence was provided regarding its localization in the egg or its secretions [22]. Although SmVAL9-specific igGs were detected in the mice at 14 weeks post infection, thus indicating protein release from the eggs, it is likely a consequence of the late, chronic phase of infection, in which the death and disintegration of eggs trapped in the host tissues take place [2, 22, 47]. The only SmVAL that has been detected in egg secretions without the use of high-throughput omics methods is SmVAL26, but this protein appears to be a component of hatch fluid and is not truly secreted from intact eggs [4, 6, 11]. In line with the provided background, we found no evidence of SmVAL9 or SmVAL29 secretion from eggs into mammalian tissues, further supporting its miracidial origin.

Another support for our conclusions is the conspicuous similarity to the analogous situation in cercariae, which are invasive larvae that infect mammals. We found that SmVAL4, SmVAL10 and SmVAL18 were the most highly expressed SmVALs in sporocysts. These proteins have also been repeatedly identified in cercarial secretomes and localize to penetration glands of cercariae, likely supporting tissue invasion when schistosomes encounter the mammalian host [48–50]. Egg expression may therefore reflect the protein weaponry of miracidia in the same way that sporocysts reflect the protein weaponry of cercariae, as has been shown several times in the past [51–53]. These similarities further suggest that SmVALs may function as multipurpose proteins essential for the invasion of both molluscan and mammalian hosts rather than facilitating tissue remodelling for egg migration in mammalian tissues.

Finally, the lack of binding to the 755 tested human cell receptors in our SAVEXIS screen further suggests that egg SmVALs may not engage directly in protein-mediated interactions with mammalian cells. Despite the previously reported ability of SmVAL9 to influence the expression of ECM remodelling-related molecules in murine macrophages [22], our screen did not identify any human receptors able to interact with the protein. This finding suggests that the previously observed interaction is either mediated by a cellular receptor that was not present in our high-throughput assay or that the ECM remodelling effects observed in earlier studies may be due to a nonspecific response to artificial protein treatment rather than a true biological interaction. Alternatively, interactions of different natures may take place, such as signalling lipid binding at the cell surface, a likely scenario given the lipid-binding properties of the CAP domain [54–57]. Nevertheless, the absence of specific interactions with mammalian host receptors aligns with the localization experiments, further supporting the conclusion that these proteins function in the miracidia rather than the egg.

Conclusions

Our findings suggest that egg-expressed SmVALs are unlikely to directly mediate interactions with mammalian host tissues, as previously proposed. Rather, these molecules appear to be adapted primarily for the invasion of molluscan hosts. Notably, the identification of new SmVAL genes and their distinct expression profiles throughout the *S. mansoni* life cycle further supports the idea that SmVALs function at critical stages of invasion by miracidia and cercariae. While we cannot entirely dismiss a role for egg-expressed SmVALs in host–parasite interactions owing to their release upon egg damage, we believe that SmVAL molecules involved in host manipulation are more effectively studied at the life stages where the secretion of these proteins is undeniable. In future work, we therefore plan to apply a high-throughput SAVEXIS screen to cercarial SmVALs to definitively prove or disprove the ability of these proteins to bind to mammalian host cell receptors. In addition, techniques such as laser microdissection and improved proteomics could be implemented to further elucidate the localization and dynamics of SmVAL secretion across life stages, allowing a more precise understanding of these enigmatic proteins.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11369-4.

Additional file 1: Table S1. Datasets for gene expression meta-analysis. Accession numbers of publicly available RNA-Seq datasets used for the meta-analysis of gene expression of SmVALs across the *S. mansoni* lifecycle.

Additional file 2: Table S2. HTSeq counts for gene expression meta-analysis. HTSeq counts for all the compared life stages and replicates retrieved from Wormbase ParaSite to which the GeTMM normalisation was applied.

Additional file 3: Table S3. GeTMM counts for gene expression meta-analysis. GeTMM normalised counts for all *S. mansoni* genes used for direct intra- and intersamples comparisons of gene expression throughout the life-cycle.

Additional file 4: Document S1. Novel SmVAL primers and sequencing data. The sequences of primers used to amplify novel SmVAL30-35 genes and sequencing data of amplicons.

Additional file 5. Document S2. Primers and cloning for *E. coli* expression. The primers for SmVAL9 and SmVAL29 gene constructs for In-fusion cloning and subsequent *E. coli* protein expression.

Additional file 6: Table S4. Revision of the SmVAL family. The complete revision of SmVAL family including newly identified SmVALs and comparison of current analysis with previously published data on the family regarding gene details, signal peptide predictions and nomenclature.

Additional file 7: Figure S1. *S. mansoni* SCP/TAPS domains distance matrix. The similarity distance matrix of identified *S. mansoni* SCP/TAPS domains.

Additional file 8: Table S5. Relative SmVAL expression across the life cycle. Relative expression of all the members of SmVAL family throughout the lifecycle.

Additional file 9: Table S6. SmVAL SAVEXIS results. Complete readout of the SmVAL9 and SmVAL29 plates used for SAVEXIS protein interaction screen.

Acknowledgements
Microscopy was performed in the Vinicna Microscopy Core Facility co-financed by the Czech-BioImaging large RI project LM2023050. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program "Projects of Large Research, Development and Innovations Infrastructures". We would like to acknowledge Gavin Wright for providing access to the library of human receptors, Kristýna Peterková for her help with the study design and brainstorming and Tomáš Macháček for reading and commenting on the manuscript.

Authors' contributions
LK has performed the bioinformatic analyses, cloning and PCR verification, E. coli protein expression and animal immunization, data interpretation and visualisation and composed the manuscript. LJ performed the histological processing and immunolocalization experiments. ZI, AR and CC performed the HEK protein expression followed by SAVEXIS protein–protein interaction screen. JD supervised the project and reviewed the manuscript. All authors read and approved the final manuscript.

Funding
This study was supported by the Czech Science Foundation 23-06638S (https://gacr.cz/en) (to JD), Schistosomiasis Research Grant FR4003935 by ARES Trading S.A., A., an affiliate of Merck KGaA, Darmstadt, Germany (https://www.merckgroup.com/en) (to JD), the Grant Agency of Charles University project 198722 (https://cuni.cz/UKEN-753.html) (to LK), and Charles University institutional funding (https://cuni.cz/UKEN-1889.html): Cooperatio Biology, UNCE24/SCI/011 (LJ), SVV 260678/2023 (to LK) and an MRC Career Development Award MR/W016397/1 (to CC)—this UK funded award is carried out in the frame of the Global Health EDCTP3 Joint Undertaking". Microscopy was performed in the Vinicna Microscopy Core Facility co-financed by the Czech-BioImaging large RI project LM2023050 funded by MEYS CR. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140 funded by MEYS CR) provided within the program "Projects of Large Research, Development and Innovations Infrastructures". The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability
Nucleotide sequence data reported in this paper are available in GenBank™ under the accession numbers PQ565842 for SmVAL29, PQ565715 for SmVAL30, PQ565716 for SmVAL31, PQ565730 for SmVAL33, PQ565731 for SmVAL34, and PQ565841 for SmVAL35. All relevant intermediate files regarding SmVAL gene identification are available at https://github.com/lukykony/SmVAL_identification. All data generated or analysed during this study are included in this published article and its supplementary information files. Any intermediate datasets in the analyses performed and relevant data not directly present in the manuscript can be obtained from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Research involving experimental animals was performed in accordance with the animal welfare laws of Czechia and under the European regulations for transport, housing and care of laboratory animals (CZ Act No. 246/1992, EU Directive 2010/63/EU amended by Regulation 2019/1010/EU). Experiments were approved by the Animal Welfare Committees of University of Life Sciences and the Ministry of Education, Youth and Sports of Czechia (MSMT-29299/2020–3). All animals used in the study were maintained by a certified person (Certificate Number CZ 02847) in an accredited facility. Ketamine-xylazine anaesthesia was employed to sacrifice the mice as indicated in "Parasite material" section.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Parasitology, Faculty of Science, Charles University, Prague, Czechia. [2]Department of Ecology, Center of Infectious Animal Diseases, Faculty of Environmental Sciences, Czech University of Life Sciences, Prague, Czechia. [3]Department of Zoology and Fisheries, Center of Infectious Animal Diseases, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences, Prague, Czechia. [4]Department of Biology, University of York, York, UK. [5]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czechia. [6]Hull York Medical School, University of York, York, UK. [7]York Biomedical Research Institute, University of York, York, UK.

Received: 4 December 2024 Accepted: 14 February 2025
Published online: 24 February 2025

References

1. World Health Organisation. Schistosomiasis. 2023. In: https://www.who.int/news-room/fact-sheets/detail/schistosomiasis.
2. Schwartz C, Fallon PG. Schistosoma "Eggs-iting" the host: Granuloma formation and egg excretion. Front Immunol. 2018. https://doi.org/10.3389/fimmu.2018.02492.
3. Ashton PD, Harrop R, Shah B, Wilson R, a,. The schistosome egg: development and secretions. Parasitology. 2001;122:329–38.
4. Carson JP, Robinson MW, Hsieh MH, Cody J, Le L, You H, McManus DP, Gobert GN. A comparative proteomics analysis of the egg secretions of three major schistosome species. Mol Biochem Parasitol. 2020. https://doi.org/10.1016/j.molbiopara.2020.111322.
5. Cass CCL, Johnson JR, Califf LL, Xu T, Hernandez HJ, Stadecker MJ, Yates JR, Williams DL. Proteomic analysis of Schistosoma mansoni egg secretions. Mol Biochem Parasitol. 2007;155:84–93.

6. Mathieson W, Wilson RA. A comparative proteomic study of the undeveloped and developed Schistosoma mansoni egg and its contents: The miracidium, hatch fluid and secretions. Int J Parasitol. 2010;40:617–28.
7. Colley DG, Secor WE. Immunology of human schistosomiasis. Parasite Immunol. 2014;36:347–57.
8. Costain AH, MacDonald AS, Smits HH. Schistosome Egg Migration: Mechanisms, Pathogenesis and Host Immune Responses. Front Immunol. 2018. https://doi.org/10.3389/fimmu.2018.03042.
9. Everts B, Hussaarts L, Driessen NN, et al. Schistosome-derived omega-1 drives Th2 polarization by suppressing protein synthesis following internalization by the mannose receptor. J Exp Med. 2012;209:1753–67.
10. Knuhr K, Langhans K, Nyenhuis S, Viertmann K, Overgaard Kildemoes AM, Doenhoff MJ, Haas H, Schramm G. Schistosoma mansoni egg-released IPSE/alpha-1 dampens inflammatory cytokine responses viabasophil interleukin (IL)-4 and IL-13. Front Immunol. 2018. https://doi.org/10.3389/fimmu.2018.02293.
11. Farias LP, Rodrigues D, Cunna V, Rofatto HK, Faquim-Mauro EL, Leite LCC. Schistosoma mansoni venom allergen like proteins present differential allergic responses in a murine model of airway inflammation. PLoS Negl Trop Dis. 2012. https://doi.org/10.1371/journal.pntd.0001510.
12. Wilbers RHP, Schneiter R, Holterman MHM, Drurey C, Smant G, Asojo OA, Maizels RM, Lozano-Torres JL. Secreted venom allergen-like proteins of helminths: Conserved modulators of host responses in animals and plants. PLoS Pathog. 2018. https://doi.org/10.1371/journal.ppat.1007300.
13. Chalmers IW, Hoffmann KF. Platyhelminth Venom Allergen-Like (VAL) proteins: Revealing structural diversity, class-specific features and biological associations across the phylum. Parasitology. 2012;139:1231–45.
14. Chalmers IW, McArdle AJ, Coulson RMR, Wagner MA, Schmid R, Hirai H, Hoffmann KF. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the Schistosoma mansoni venom allergen-like (SmVAL) gene family. BMC Genomics. 2008. https://doi.org/10.1186/1471-2164-9-89.
15. Hunt VL, Tsai IJ, Coghlan A, et al. The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat Genet. 2016;48:299–307.
16. Buddenborg SK, Lu Z, Sankaranarayan G, Doyle SR, Berriman M. The stage- and sex-specific transcriptome of the human parasite Schistosoma mansoni. Sci Data. 2023. https://doi.org/10.1038/s41597-023-02674-2.
17. Fernandes RS, Barbosa TC, Barbosa MMF, Miyasato PA, Nakano E, Leite LCC, Farias LP. Stage and tissue expression patterns of Schistosoma mansoni venom allergen-like proteins SmVAL 4, 13, 16 and 24. Parasit Vectors. 2017. https://doi.org/10.1186/s13071-017-2144-2.
18. Perally S, Geyer KK, Farani PSG, et al. Schistosoma mansoni venom allergen-like protein 6 (SmVAL6) maintains tegumental barrier function. Int J Parasitol. 2021;51:251–61.
19. Rofatto HK, Parker-Manuel SJ, Barbosa TC, Tararam CA, Alan Wilson R, Leite LCC, Farias LP. Tissue expression patterns of Schistosoma mansoni Venom Allergen-Like proteins 6 and 7. Int J Parasitol. 2012;42:613–20.
20. Attenborough T, Rawlinson KA, Soria CLD, Ambridge K, Sankaranarayanan G, Graham J, Cotton JA, Doyle SR, Rinaldi G, Berriman M. A single-cell atlas of the miracidium larva of the human blood fluke Schistosoma mansoni: cell types, developmental pathways and tissue architecture. 2024. https://doi.org/10.7554/eLife.95628.2.
21. Peterková K, Konečný L, Macháček T, Jedličková L, Winkelmann F, Sombetzki M, Dvořák J. Winners vs. losers: Schistosoma mansoni intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity. PLoS Pathog. 2024. https://doi.org/10.1371/journal.ppat.1012268.
22. Yoshino TP, Brown M, Wu XJ, Jackson CJ, Ocadiz-Ruiz R, Chalmers IW, Kolb M, Hokke CH, Hoffmann KF. Excreted/secreted Schistosoma mansoni venom allergen-like 9 (SmVAL9) modulates host extracellular matrix remodelling gene expression. Int J Parasitol. 2014;44:551–63.
23. Farias LP, Chalmers IW, Perally S, et al. Schistosoma mansoni venom allergen-like proteins: phylogenetic relationships, stage-specific transcription and tissue localization as predictors of immunological cross-reactivity. Int J Parasitol. 2019;49:593–9.
24. Wu XJ, Sabat G, Brown JF, Zhang M, Taft A, Peterson N, Harms A, Yoshino TP. Proteomic analysis of Schistosoma mansoni proteins released during in vitro miracidium-to-sporocyst transformation. Mol Biochem Parasitol. 2009;164:32–44.
25. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2–10.
26. Buddenborg SK, Tracey A, Berger DJ, et al. Assembled chromosomes of the blood fluke Schistosoma mansoni provide insight into the evolution of its ZW sex-determination system. 2021. https://doi.org/10.1101/2021.08.13.456314.
27. Prawer YDJ, Stroehlein AJ, Young ND, Kapoor S, Hall RS, Ghazali R, Batterham P, Gasser RB, Perry T, Anstead CA. Major SCP/TAPS protein expansion in Lucilia cuprina is associated with novel tandem array organisation and domain architecture. Parasit Vectors. 2020. https://doi.org/10.1186/s13071-020-04476-6.
28. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res. 2007. https://doi.org/10.1093/nar/gkm256.
29. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40:1023–5.
30. HMMER: biosequence analysis using profile hidden Markov models. http://hmmer.org.
31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
32. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
33. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, Teeling E. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4.
34. FigTree. http://tree.bio.ed.ac.uk/software/figtree/.
35. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91.
36. Wangwiwatsin A, Protasio AV, Wilson S, Owusu C, Holroyd NE, Sanders MJ, Keane J, Doenhoff MJ, Rinaldi G, Berriman M. Transcriptome of the parasitic flatworm schistosoma mansoni during intra-mammalian development. PLoS Negl Trop Dis. 2020;14:1–25.
37. Smid M, Coebergh van den Braak RRJ, van de Werken HJG, et al. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. BMC Bioinformatics. 2018. https://doi.org/10.1186/s12859-018-2246-7.
38. Jedličková L, Dvořák J, Hrachovinová I, Ulrychová L, Kašný M, Mikeš L. A novel Kunitz protein with proposed dual function from Eudiplozoon nipponicum (Monogenea) impairs haemostasis and action of complement in vitro. Int J Parasitol. 2019;49:337–46.
39. Crosnier C, Hokke CH, Protasio AV, et al. Screening of a Library of Recombinant Schistosoma mansoni Proteins With Sera From Murine and Human Controlled Infections Identifies Early Serological Markers. In: Journal of Infectious Diseases. Oxford University Press, 2022. pp 1435–1446.
40. Sun Y, Gallagher-Jones M, Barker C, Wright GJ. A benchmarked protein microarray-based platform for the identification of novel low-affinity extracellular protein interactions. Anal Biochem. 2012;424:45–53.
41. Shilts J, Severin Y, Galaway F, Müller-Sienerth N, Chong ZS, Pritchard S, Teichmann S, Vento-Tormo R, Snijder B, Wright GJ. A physical wiring diagram for the human immune system. Nature. 2022;608:397–404.
42. Crosnier C, Bustamante LY, Bartholdson SJ, et al. Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. Nature. 2011;480:534–7.
43. Takaki KK, Rinaldi G, Berriman M, Pagán AJ, Ramakrishnan L. Schistosoma mansoni Eggs Modulate the Timing of Granuloma Formation to Promote Transmission. Cell Host Microbe. 2021;29:58-67.e5.
44. Collins JJ, King RS, Cogswell A, Williams DL, Newmark PA. An atlas for schistosoma mansoni organs and life-cycle stages using cell type-specific markers and confocal microscopy. PLoS Negl Trop Dis. 2011. https://doi.org/10.1371/journal.pntd.0001009.
45. Haeberlein S, Angrisano A, Quack T, Lu Z, Kellershohn J, Blohm A, Grevelding CG, Hahnel SR. Identification of a new panel of reference genes to study pairing-dependent gene expression in Schistosoma mansoni. Int J Parasitol. 2019;49:615–24.
46. Lu Z, Sessler F, Holroyd N, Hahnel S, Quack T, Berriman M, Grevelding CG. Data Descriptor: A gene expression atlas of adult Schistosoma mansoni and their gonads. Sci Data. 2017. https://doi.org/10.1038/sdata.2017.118.

47. Linder E. The Schistosome Egg in Transit. Ann Clin Pathol. 2017;5:1110.
48. Curwen RS, Ashton PD, Sundaralingam S, Wilson RA. Identification of Novel Proteases and Immunomodulators in the Secretions of Schistosome Cercariae That Facilitate Host Entry. Mol Cell Proteomics. 2006;5:835–44.
49. Fernandes RS, Fernandes LGV, de Godoy AS, Miyasato PA, Nakano E, Farias LP, Nascimento ALTO, Leite LCC. Schistosoma mansoni venom allergen-like protein 18 (SmVAL18) is a plasminogen-binding protein secreted during the early stages of mammalian-host infection. Mol Biochem Parasitol. 2018;221:23–31.
50. Hansell E, Braschi S, Medzihradszky KF, Sajid M, Debnath M, Ingram J, Lim KC, McKerrow JH. Proteomic analysis of skin invasion by blood fluke larvae. PLoS Negl Trop Dis. 2008. https://doi.org/10.1371/journal.pntd.0000262.
51. Dolečková K, Albrecht T, Mikeš L, Horák P. Cathepsins B1 and B2 in the neuropathogenic schistosome Trichobilharzia regenti: Distinct gene expression profiles and presumptive roles throughout the life cycle. Parasitol Res. 2010;107:751–5.
52. Parker-Manuel SJ, Ivens AC, Dillon GP, Wilson RA. Gene expression patterns in larval Schistosoma mansoni associated with infection of the mammalian host. PLoS Negl Trop Dis. 2011. https://doi.org/10.1371/journal.pntd.0001274.
53. Ruth Lawson J, Wilson RA. The survival of the cercariae of Schistosoma mansoni in relation to water temperature and glycogen utilization. Parasitology. 1980;81:337–48.
54. Darwiche R, Mène-Saffrané L, Gfeller D, Asojo OA, Schneiter R. The pathogen-related yeast protein Pry1, a member of the CAP protein superfamily, is a fatty acid-binding protein. J Biol Chem. 2017;292:8304–14.
55. Kelleher A, Darwiche R, Rezende WC, Farias LP, Leite LCC, Schneiter R, Asojo OA. Erratum : Schistosoma mansoni venom allergen-like protein 4 (SmVAL4) is a novel lipid-binding SCP, TAPS protein that lacks the prototypical CAP motifs. Corrigendum (Acta Crystallographica Section D: Biological Crystallography D70, 2014 2186–2196). Acta Crystallogr D Biol Crystallogr. 2015;71:1022.
56. Ma D, Xu X, An S, et al. A novel family of RGD-containing disintegrins (Tablysin-15) from the salivary gland of the horsefly Tabanus yao targets αIIbβ3 or αVβ3 and inhibits platelet aggregation and angiogenesis. Thromb Haemost. 2011;105:1032–45.
57. Xu E, Francischetti IMB, Lai R, Ribeiro JMC, Andersen JF. Structure of protein having inhibitory disintegrin and leukotriene scavenging functions contained in single domain. J Biol Chem. 2012;287:10967–76.

Publisher's Note