

RESEARCH ARTICLE

# MiR-277/4989 regulate transcriptional landscape during juvenile to adult transition in the parasitic helminth *Schistosoma mansoni*

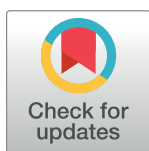
Anna V. Protasio<sup>1aa\*</sup>, Stijn van Dongen<sup>2</sup>, Julie Collins<sup>3</sup>, Leonor Quintais<sup>2</sup>, Diogo M. Ribeiro<sup>1ab</sup>, Florian Sessler<sup>1</sup>, Martin Hunt<sup>1</sup>, Gabriel Rinaldi<sup>1</sup>, James J. Collins<sup>3</sup>, Anton J. Enright<sup>2</sup>, Matthew Berriman<sup>1</sup>

**1** Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, **2** European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **3** Department of Pharmacology, UT Southwestern Medical Center, Dallas, TX, United States of America

<sup>aa</sup> Current address: Institute for Stem Cell Biology and Regenerative Medicine, Bangalore, India.

<sup>ab</sup> Current address: Aix-Marseille University, TAGC Inserm U1090, Marseille, France.

\* [ap6@sanger.ac.uk](mailto:ap6@sanger.ac.uk)



 OPEN ACCESS

**Citation:** Protasio AV, van Dongen S, Collins J, Quintais L, Ribeiro DM, Sessler F, et al. (2017) MiR-277/4989 regulate transcriptional landscape during juvenile to adult transition in the parasitic helminth *Schistosoma mansoni*. *PLoS Negl Trop Dis* 11(5): e0005559. <https://doi.org/10.1371/journal.pntd.0005559>

**Editor:** Paul J. Brindley, George Washington University School of Medicine and Health Sciences, UNITED STATES

**Received:** April 20, 2016

**Accepted:** April 5, 2017

**Published:** May 23, 2017

**Copyright:** © 2017 Protasio et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Schistosomula poly-A enriched and small RNAseq data: European Nucleotide Archive study accession number PRJEB3190. Adult poly-A enriched RNAseq data: ArrayExpress accession number E-ERAD-478.

**Funding:** The work was funded by the Wellcome Trust (grant WT098051) to MB and NIH (R01AI121037\_ to JJC). The funders had no role in

## Abstract

Schistosomes are parasitic helminths that cause schistosomiasis, a disease affecting circa 200 million people, primarily in underprivileged regions of the world. *Schistosoma mansoni* is the most experimentally tractable schistosome species due to its ease of propagation in the laboratory and the high quality of its genome assembly and annotation. Although there is growing interest in microRNAs (miRNAs) in trematodes, little is known about the role these molecules play in the context of developmental processes. We use the completely unaware “miRNA-blind” bioinformatics tool Sylamer to analyse the 3'-UTRs of transcripts differentially expressed between the juvenile and adult stages. We show that the miR-277/4989 family target sequence is the only one significantly enriched in the transition from juvenile to adult worms. Further, we describe a novel miRNA, sma-miR-4989 showing that its proximal genomic location to sma-miR-277 suggests that they form a miRNA cluster, and we propose hairpin folds for both miRNAs compatible with the miRNA pathway. In addition, we found that expression of sma-miR-277/4989 miRNAs are up-regulated in adults while their predicted targets are characterised by significant down-regulation in paired adult worms but remain largely undisturbed in immature “virgin” females. Finally, we show that sma-miR-4989 is expressed in tegumental cells located proximal to the oesophagus gland and also distributed throughout the male worms' body. Our results indicate that sma-miR-277/4989 might play a dominant role in post-transcriptional regulation during development of juvenile worms and suggest an important role in the sexual development of female schistosomes.

## Author summary

Schistosomes are parasitic helminths that infect a range of different animals causing great morbidity and some level of mortality among endemic populations in sub-Saharan Africa, southeast Asia and regions of south America. RNA has long been known as a translator of

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

the message between DNA and protein. However, it is nowadays well accepted that RNA roles go beyond that of a translator. Such RNA molecules do not translate DNA into protein and are therefore referred to as non-coding RNAs. Here we study a particular type of RNA molecules called microRNA (miRNA). These small RNA molecules are ~ 19–22 nucleotides (nt) in length and their most characterised function so far is that of regulating the bioavailability of messenger RNA for the production of protein. We call this effect post-transcriptional regulation. Previously, it has been shown that different stages of the *Schistosoma* parasites express different types of miRNAs. Our work focuses in utilising differences in gene expression as the readout of potential post-transcriptional regulation. Using bioinformatics tools we found that members of one miRNA family called miR-277/4989 might be responsible for the change in gene expression observed between juvenile and adult worms. Furthermore, the effect of this miRNA seems to be more prominent in the sexually matured females rather than in the immature “virgin” females.

## Introduction

Schistosomes are parasitic flatworms and causative agents of schistosomiasis (intestinal or urinary depending on the species), and are responsible for more than 200 million cases of human disease across the globe. Three schistosome species infect humans: *Schistosoma haematobium*, *S. japonicum* and *S. mansoni*. Although *S. haematobium* has the greatest clinical impact [1], *S. mansoni* is more amenable to maintenance in laboratory conditions [2,3], and therefore is the most studied. In addition, recent advances in functional genomics have been successfully applied to *S. mansoni* [4–6] and *S. japonicum* [7,8], making functional characterisation of genomic elements more amenable in these parasites.

Unusual among flatworms, schistosomes are dioecious (distinct male and female individuals), and adult worms dwell *in copula* in the blood circulatory system. Females only achieve sexual maturity after pairing with a male worm, lodging in the male’s gynecophoral canal, and laying hundreds of eggs each day. In the absence of a male worm, the female remains sexually immature and stunted in size. The retention of eggs in host tissues drives the pathogenesis of schistosomiasis, mainly characterised by chronic inflammation, fibrosis and formation of granulomas in the liver in the case of intestinal schistosomiasis. About half of the eggs traverse the intestinal wall and are excreted with the faeces into the environment. When in contact with fresh water, mature schistosome eggs hatch releasing free-living ciliated larvae (miracidia) that swim seeking snails, the intermediate host, to infect. Within the snail the parasite undergoes two rounds of asexual reproduction, finally releasing the second free-living human-infective larval stage, called cercariae, into the water environment [9]. The cercariae infect the definitive host by penetrating through the skin [10,11] and once inside, transform into obligate parasitic schistosomula via a rapid and irreversible series of morphological changes. The parasites develop and migrate through the circulatory system and after five to six weeks, adult schistosomes reach the portal circulation where males and females pair (reviewed in [12] and [13]). These extreme environmental changes are associated with rapid and specific physiological transformation—such as the parasite surface membranes and carbohydrate coating changing from cercariae to schistosomula [14,15]—accompanied by transcriptional changes [16,17]. Establishment and development of male and female schistosome pairs has been well characterised at both morphological [13] and transcriptional levels [18,19]. However, the underlying molecular cues that trigger such changes are still unknown.

The non-coding RNA component of the *Schistosoma* genomes (*S. mansoni* and *S. japonicum*) was first described *in silico* [20] followed by reports featuring a combination of

experimental and *in silico* approaches [21–29]. Furthermore, the microRNA (miRNA) pathway in *S. mansoni* has been predicted using computational methods [30] and more recently, progress in the experimental characterisation of individual components of the pathway [31] as well as the role of individual miRNAs have been addressed [32]. With the increasing availability of high-throughput sequencing technologies (mainly Illumina platforms), the number of publications describing the miRNA, small RNA and non-coding RNA complement of *Schistosoma* spp has risen dramatically, in particular for *S. japonicum*. However, the *S. japonicum* genome is highly fragmented (25,048 scaffolds in the WBPS9 release available from Wormbase ParaSite—compared to 885 scaffolds for *S. mansoni*) making the genome localization of miRNAs and the analysis of their genomic context difficult. What is more, the lack of accurately defined untranslated regions (UTR) makes assessing miRNA-target sites unreliable.

At only ~19–22 nucleotides, miRNAs play a central role in post-transcriptional gene regulation. MiRNAs are encoded in the nuclear genome of most eukaryotic organisms and like protein-coding genes are transcribed by RNA polymerase II, poly-adenylated at their 3' ends and capped at their 5' ends. During transcription, a 1kb immature miRNA transcript, called pri-miRNA or primary precursor miRNA, acquires a characteristic stem-loop secondary structure. While still in the nucleus, this structure serves as a target for two proteins, Drosha and DGCR8 (microprocessor complex), which cleave the pri-mRNA at the base of the stem-loop structure producing a pre-miRNA. The pre-miRNA—now 60–80 nt—binds to auxiliary proteins that aid export of the microprocessor complex from the nucleus to the cytoplasm. Once outside the nucleus, Dicer and Argonaut further process the pre-miRNA to form an RNA-induced silencing complex (RISC), which carries the mature miRNA, while the passenger or antisense miRNA (often designated as miRNA\*) is degraded. RISC is responsible for directing the miRNA to its target sequence, often (but not exclusively) located in the 3'-UTR of a mRNA, which results in the repression of protein translation or the degradation of the target mRNA molecule (reviewed in [33] and [34]).

MiRNAs have a pivotal role in organism development. For example, in the larval moults of *Caenorhabditis elegans*, the miRNA *lin-4* progressively accumulates in first-stage larvae, down regulating LIN-14 protein and enabling second-stage larvae to develop. Subsequent larval development from L3 to L4 is controlled in part by miRNAs of the *let-7* family [35].

The development, pairing and consequently sexual maturation of schistosomes are of particular interest because they represent the cause of host pathology and transmission of this devastating parasitic disease. To identify whether miRNAs exert an effect on the transcriptome we adopted a non-conventional approach, instead of simply profiling the miRNAs as is commonplace [21,23,24,28,29], we use Sylamer [36], an algorithm that combines transcript expression changes with the presence of potential miRNA recognition sites in well-annotated 3'-UTRs. Our analyses suggest that the sma-miR-277/4989 family of miRNAs dominates the transcriptional landscape changes during the transition from juvenile to adult worm. We show that most of the targets for these miRNAs encode transcription factors, molecules involved in transcriptional activation/repression as well as signalling and proteins associated with adult stem-cell maintenance. Furthermore, a fraction of the targets are differentially expressed between mature, sexually active females and immature “virgin” females suggesting a role in sexual maturation or sexual reproduction.

## Materials and methods

### 1.1. Ethics statement

All animal experiments were conducted under Home Office Project Licence No. 80/2596. All protocols were presented and approved by the Animal Welfare and Ethical Review Body

(AWERB) of the Wellcome Trust Sanger Institute. The AWERB is constituted as required by the UK Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

## 1.2. Parasite material

Infected *Biomphalaria glabrata* snails were purchased from BioGlab Ltd. Nottingham, UK and cercariae were harvested by exposure of infected snails to light for two hours in aquarium-conditioned water. *S. mansoni* intra-mammalian stages were collected from Balb/c mice previously infected with 300-pooled cercariae. For single sex infections (to retrieve unpaired male and female worms), mono-miracidia snail infections were performed and each snail tested for the production of male or female cercariae, by sex-specific PCR [37]. Balb/c mice were infected separately with either male or female cercariae. Juvenile and adult worms were recovered via perfusion of mouse circulatory system [2] at the indicated times post-infection (see [Results](#)). Male and female worms were separated, washed in DMEM, concentrated and stored in Trizol reagent (Invitrogen, UK) at -80°C.

## 1.3. RNA extraction, RNAseq library preparation and sequencing

RNA was isolated from samples stored in Trizol reagent (Invitrogen, UK) following manufacturer instructions. RNAseq libraries were prepared from 1ug of total RNA using TruSeq RNA Library Preparation Kit (Illumina, UK). All samples were processed with three biological replicates—worms from one mouse representing one biological replicate. Libraries were multiplexed and sequenced on Illumina HiSeq 2500 with 100 bp paired-end reads. Sequencing data was submitted to ArrayExpress under accession number E-ERAD-478.

## 1.4. Processing of poly-A enriched RNAseq sequencing reads and differential expression analysis of paired / unpaired male and female worms

RNAseq data from juvenile and adult, male and female schistosome worms were mapped to the reference *S. mansoni* assembly version 5.2 of the genome assembly [38] using Tophat2 (default parameters except: -g 1—library-type fr-firststrand -a 6 -i 10 -I 40000—microexon-search—min-segment-intron 10—max-segment-intron 40000). The resulting binary alignment mapping (BAM) files were sorted using Samtools [39] and reads per transcript calculated using HTseq-count [40] (parameters: -f bam -a 30 -t CDS -s reverse -m union). The GTF file used to calculate reads per transcript can be found in Supplementary [S1 Dataset](#); only features starting with “Smp” (corresponding to protein coding genes) were taken into consideration. After mapping and counting, differential expression analysis was performed using DESeq2 [41] in the R environment [42]. Pairwise differential expression values for male and female worms (from mixed infections) between juveniles (21 days post infection d.p.i.) and adult worms (38 d.p.i.) were calculated and used as input for Sylamer (see below). Time-course expression data for male and female, paired and unpaired worms was generated using likelihood ratio test incorporated in DESeq2 package.

## 1.5. UTR prediction and experimental validation

RNAseq data generated from polyadenylated-enriched samples was used to improve the annotation of 3'-UTRs in the *S. mansoni* genome. BAM files generated with Tophat2 were merged using Samtools [39] and used as input for Cufflinks [43,44] (using default parameters except: -p 16—library-type fr-firststrand). To generate UTR sequences, RNAseq data and existing annotation were combined using a script developed in-house ([https://github.com/sanger-pathogens/Assembly\\_tools/tree/master/assembly\\_tools/annotate\\_utrs\\_using\\_cufflinks](https://github.com/sanger-pathogens/Assembly_tools/tree/master/assembly_tools/annotate_utrs_using_cufflinks)).

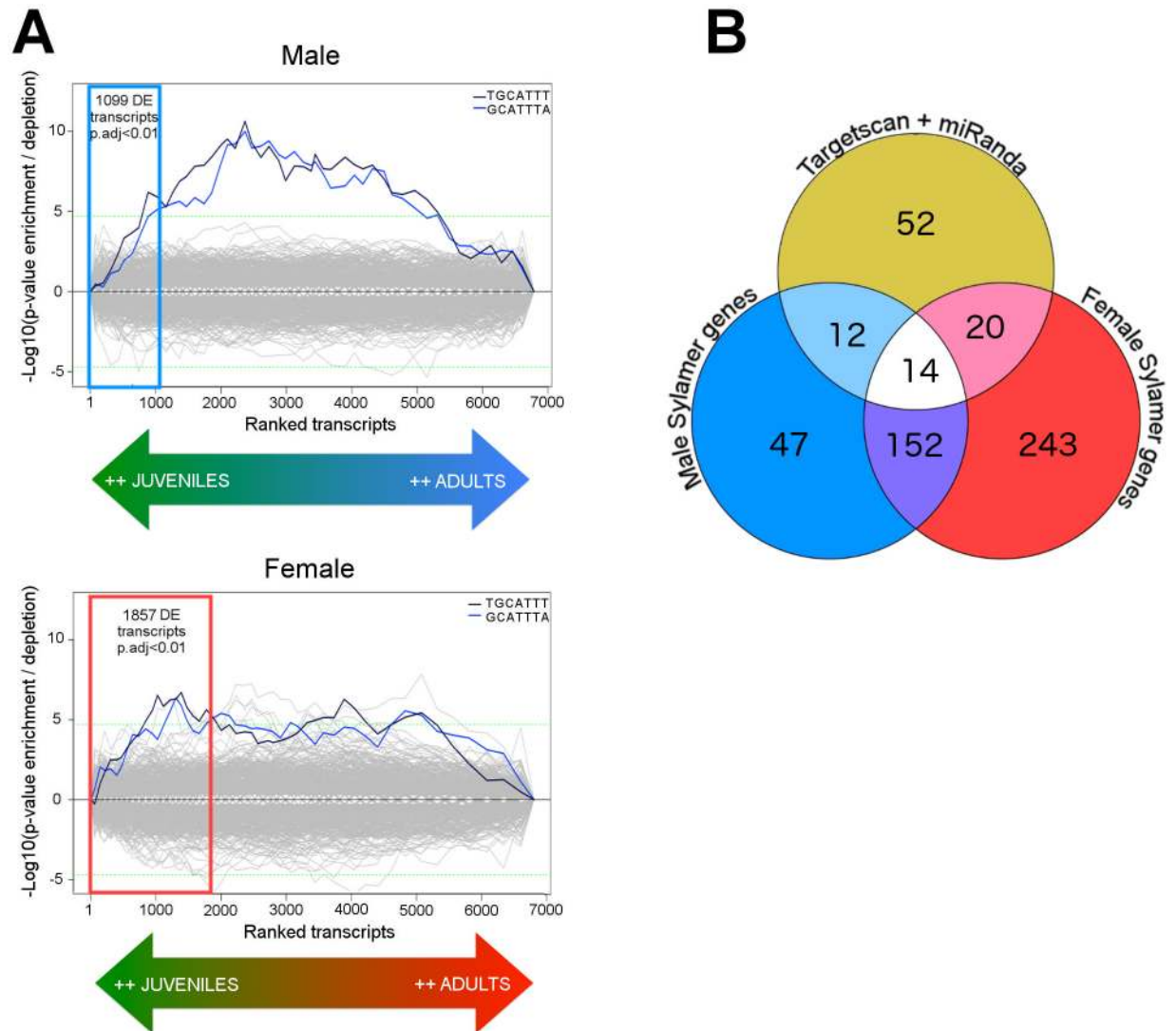
Briefly, for each existing gene, the script takes an intersecting annotation from the cufflinks GTK output file and extends the existing gene model using the provided annotation, labelling the new part of the annotation as a UTR. If the potential UTR happens to overlap a second gene, then it is not used. As a result, 3,321 3'-UTRs from a total of 7,373 3'-UTRs were updated using this method. Likewise, 4,081 5'-UTRs from a total of 7,271 5'-UTRs were updated. The total number of protein-coding genes is 10,841. Note that this version of the annotation is a "snapshot" and may not coincide with the current version in <http://www.genedb.org/Homepage/Smansoni> or [http://parasite.wormbase.org/Schistosoma\\_mansoni\\_prjea36577/Info/Index/](http://parasite.wormbase.org/Schistosoma_mansoni_prjea36577/Info/Index/).

A selection of twelve 3'-UTRs with lengths  $\geq 600$  bp, and a range of RPKM expression values between 9–3,600 in adult worms, were chosen for validation by PCR (Supplementary [S1 Fig](#)). Primers were designed using primer3 batch service [\[45\]](#) using default parameters except amplicon length, which was set to 500 bp. PCR reactions were carried out using Qiagen Fast Cycling PCR Mix (Qiagen, UK) using standard conditions with annealing temperature 55C and mix sex adult cDNA as template. The list of primers is shown in Supplementary [S1 Table](#).

## 1.6. Sylamer analysis

Sylamer [\[36\]](#) is a tool to identify miRNA regulation effects from a list of differentially expressed genes, independently from miRNA measurements. It can either be used to suggest candidate miRNAs for follow-up analysis in the absence of miRNA measurements, or to confirm that putative miRNA target genes are shifted concordantly with miRNA expression changes when miRNA measurements are present. The latter is the case here. It has previously been used, among many more applications, to identify the role of miRNAs in germcell tumors [\[46\]](#), the effect of a miRNA seed mutation on gene expression [\[47\]](#), and identification of miRNA targets in murine Dgcr8-deficient embryonic stem cell lines [\[48\]](#). Sylamer's *modus operandi* is briefly described here, for full details refer to [\[36\]](#). Several nucleotide words in the eight-nucleotide stretch (8-mer) at the 5' end of a miRNA are core determinants of miRNA binding [\[34,49\]](#). These are the 8-mer itself, the two 7-mers, the core 6-mer, and the leading 6-mer. The nucleotide sequences that are complementary to these miRNA seed words are called Seed Complementary Regions (SCR). Sylamer considers a list of differentially expressed genes and tests the hypothesis that SCRs in 3'-UTRs are shifted towards one end of the gene list against the null hypothesis that these sites are homogeneously distributed throughout the gene list. It does so whilst taking into account UTR length and correcting for composition biases, using a hypergeometric model of nucleotide word occurrences in UTRs. Analogously to Gene Set Enrichment Analysis [\[50\]](#), Sylamer employs a moving rank cut-off and can hence detect shifts of SCRs that are dispersed to a greater or lesser extent.

**Sylamer plots construction.** All information produced by a Sylamer run can be summarised in a single plot, as a collection of lines where each line represents a single SCR. In a single Sylamer run, the word length is kept constant, so different plots are obtained for SCRs of length 6, 7 and 8, respectively. A single line in a Sylamer plot describes a single SCR as follows: at different cut-offs in the list of UTRs, a hypergeometric test is performed considering the summed total length of the UTRs up to that cut-off (representing the number of all SCRs), the summed total length of all UTRs, the number of occurrences of the particular SCR up to the cut-off, and the total number of occurrences of the particular SCR in all UTRs. If the SCR is enriched in the subset of UTRs, the  $-\log_{10}(\text{p-value})$  is drawn (hence at the positive y-axis). If it is instead depleted, the  $\log_{10}(\text{p-value})$  is drawn (at the negative y-axis). Multiple cut-offs are tested and a large number of SCRs are tested. On the one hand, this requires multiple testing correction, on the other hand this allows a single SCR result to be evaluated against the



**Fig 1. MiRNA target prediction based on both miRNA-unaware and miRNA-guided approaches.** (A) Sylamer enrichment landscape plots for 7mers in male (top) and female (bottom) expression data. The x-axis represents a list of transcripts, ranked from more expressed in juveniles to more expressed in adults. The y-axis represents the significance values acquired for each 7mer at each position in the ranked list of transcripts. Coloured boxes represent the fraction of transcripts significantly (adjusted p-value < 0.01) differentially expressed between juvenile and adult worm as found using DESeq2. These transcripts were subsequently filtered based on the presence of the 7mers TGCATTT or GCATTTA as found by Sylamer. The resulting sets are referred to as Male and Female Sylamer genes. (B) Venn Diagram showing the intersection of Male and Female Sylamer genes with schistosome-conserved miRNA targets as found using TargetScan with conservation + miRanda. The overlap represents transcripts with highly conserved sma-miR-277 target sites across the three *Schistosoma* spp (*S. mansoni*, *S. haematobium* and *S. japonicum*) that are also significantly down regulated during worm development.

<https://doi.org/10.1371/journal.pntd.0005559.g001>

background of plotted lines for all SCRs. A Bonferroni-adjusted significance threshold of 0.05 (as drawn in Fig 1A and 1B) typically delineates and encloses the background distribution with significant SCRs jutting out.

Additional significant SCRs above the Bonferroni-adjust threshold may occur in a Sylamer plot. This can be due to several causes. Firstly, multiple regulatory elements besides SCR are present in UTRs, such as poly-A and ARE (AU-rich element) signals. MiRNA SCRs may share such motifs. Secondly, repeated sequence fragments in UTRs of related genes may cause sharp

spikes in Sylamer plots, and stretches of low complexity sequence can cause significant results for SCRs matching such a sequence. This is overcome (as done here) by pre-processing the UTRs to remove low complexity stretches with DUST [51] and repeated sequence fragments with the RSAT [52] interface to the Vmatch program (<http://www.vmatch.de>). Finally, a significant SCR signal can cause words that are very close to it to piggyback the signal and achieve elevated significance. This is then evident from the words involved.

**Sylamer result intergration over different word lengths.** Sylamer was run with different word lengths because miRNA determinants of binding can be found among words of different length [34,36]. It is possible to run Sylamer with different word lengths, as we have done here. This provides greater insight but also requires interpretation of multiple resources. We additionally apply the procedure described in [46] to integrate Sylamer results for different word lengths were integrated into a single score for a given miRNA seed region, using a previously described procedure [46]. By integrating results, sensitivity was increased. The procedure assigned miRNAs to groups, each group defined by a common SCR of length eight. A single score was assigned to each group by considering the Sylamer result (line) for each of the constituent 6-mer, 7-mer and the 8-mer itself, obtained from different Sylamer runs. These lines (log10 transformed p-values) represent a normalised view (via the hypergeometric test) of word enrichment. The constituent lines were added together, after which the maximal amplitude of the summed result was taken as a single score. As demonstrated in [46] the resulting scores narrowly follow an extreme value distribution. The extreme value distribution parameters were estimated with the R *evd/* package [53] using the 95% quantile of the data and then obtain empirical p-values using these parameters. In the absence of a null hypothesis these empirical p-values can be interpreted/translated as the percentile range into which a score falls under the modeled distribution, i.e. a p-value of 0.0001 would correspond to the top 0.01 percentile.

### 1.7. miRNA target prediction (miRanda and TargetScan) and gene ontology enrichment

Targets were predicted in the 3'-UTRs of 4,851 genes for previously reported *S. mansoni* and *S. japonicum* miRNAs and the novel miRNA from the present study. First, miRanda (v 3.3a, ref. [54]) was run with default parameters against the predicted set of *S. mansoni* 3'-UTRs. Second, TargetScan (v 6.0, ref. [55]) was used to search the same 3'-UTRs dataset; the script `targets-can_60.pl` was run with default parameters, using the seed sequences of each *Schistosoma* spp. miRNA (from miRBase, release 21, ref. [56]) and alignments of orthologous 3'-UTRs (see below) between *S. mansoni*, *S. japonicum* and *S. haematobium*. Using custom scripts, only *S. mansoni* TargetScan target site predictions conserved among the three *Schistosoma* species were retained, i.e. requiring the target prediction to be present on the same exact aligned location on the all orthologous 3'-UTRs. Further to this, only a high-confidence set of predictions found by both TargetScan and miRanda were retained. Except for *S. mansoni*, current 3'-UTR predictions for the other two *Schistosoma* species used were not reliable; therefore new 3'-UTR predictions for *S. haematobium* and *S. japonicum* were created based on orthology with *S. mansoni*. For each *S. mansoni* gene the mRNA isoform was selected with the longest spliced sequence, and for genes with annotated 3'-UTRs, orthologous genes in *S. japonicum* and *S. haematobium* were identified using a custom-made EnsemblCompara database [57]. For each *S. japonicum* or *S. haematobium* orthologue, a predicted 3'-UTR was created with the same length of the 3'-UTR of the respective *S. mansoni* orthologous gene (up to maximum of 5 kb or to the end of underlying sequence scaffold). These predicted sets of 3'-UTRs were termed "orthologous 3'-UTRs", even though orthology is based on the underlying gene product to

which they correspond. Alignments of orthologous 3'-UTR sets for the TargetScan predictions were produced with MUSCLE v3.8.31 [58], with default parameters. Assemblies and gene sets of the three species used for accessing conservation of miRNA target predictions can be downloaded at WormBase ParaSite release 4 (<http://parasite.wormbase.org/ftp.html>): *S. haematobium* 'S.haematobium.v3.0' from the University of Melbourne, and *S. japonicum* 'ASM15177v1' from the Chinese National Human Genome Center. The *S. mansoni* assembly used here is the most updated version reported previously [38].

Gene Ontology enrichment analysis was performed using the TopGO package [59].

### 1.8. Processing of small RNA-seq libraries and genomic localisation of sma-miR-277/4989 cluster

We used previously generated small RNA-seq libraries from schistosomula stages of *S. mansoni* to assist our description of the sma-miR-277 locus (European Nucleotide Archive study PRJEB3190). Kraken [60] was used to remove adapter contamination from libraries and collapse identical reads into single sequences while maintaining annotated depth information (Supplementary S2 Table). Reads were then mapped against each other using BLASTN [61,62] to determine all pairwise similarities between reads allowing up to two mismatches (E-value < = 0.1). A pairwise similarity matrix was used to cluster reads using MCL [60,63]. Multiple sequence alignment of each cluster was performed using Clustal Omega [64]. From each cluster, a 'sentinel' read was chosen with the highest depth and mapped to the reference genome using Bowtie [65], allowing up to two mismatches.

After identifying the genomic location for the candidate and in order to obtain the putative miRNA precursor sequence, the sequences were extended in both directions. The first extension took 50nt upstream and 100nt downstream while the second extension took 100nt upstream and 50nt downstream of the cluster. Secondary structures of these putative miRNA precursors were then assessed using RNAfold from the Vienna package [66], and structures discarded with minimum free energy (MFE) > -20 kcal/mol. For each cluster, the extended sequence with the lowest associated MFE structure was retained.

### 1.9. RT-qPCR / miRNA expression analysis

Real-time quantitative PCRs were performed across two life cycle stages of *S. mansoni* representing juvenile (28 days post-infection, d.p.i.) and adult worms (49 d.p.i.) using TaqMan Small RNA Assays (Supplementary S1 Text) purchased from Applied Biosystems (Life Technologies, UK). Reverse transcription and quantitative PCR experiments were performed according to manufacturer's instructions in a StepOnePlus RT-qPCR machine (Applied Biosystems, Life Technologies, UK). *S. mansoni* U6 was used as the endogenous control [20] (gene entry sma.U6.1.1 in <http://www.genedb.org/Homepage/Smansoni>) and fold changes to miRNA expression were estimated using the delta-delta-Ct method [67].

### 1.10. Whole mount and fluorescence *in situ* hybridisation of miRNAs

Forty-two to 49 d.p.i. male and female worms from mixed infections were retrieved from infected mice, as described above, and processed using previously published protocols for *in situ* hybridization [68]. Parasites were hybridized with 21–22 nt antisense LNA-DNA probes conjugated to Digoxigenin (Exiqon, Denmark) at a final concentration between 6.5–50 nM. *Arabidopsis thaliana* ath-miR-159a or scrambled sequences were used as negative control. For FISH to detect *prohormone convertase 2* (Smp\_077980) or Tegumental cells, we synthesized FITC-conjugated probes generated from *in vitro* transcription of cloned cDNA. To robustly detect tegumental cells we employed a mixture of four FITC-conjugated antisense probes



targeting the following tegument-specific mRNAs: *calpain* (Smp\_214190), *gtp -4* (Smp\_105410), *annexin* (Smp\_077720), and *npp-5* (Smp\_153390) (ref: [69–73]).

## Results

### 2.1. Target-prediction suggests sma-miR-277/4989 are prominent post-regulatory miRNAs in developing juvenile worms

In our analysis, Sylamer [36] was used to search for enriched short sequences, corresponding to potential miRNA target sites, from a list of genes differentially expressed between juvenile and adult male and female worms. Sylamer finds enriched “words” (in our case, Seed Complementary Regions or SCR) in the 3′-UTRs of transcripts with similar expression profiles (i.e. up-regulated or down-regulated). The advantage of this method is that potential targets can be identified without prior knowledge of the specific miRNAs that affect the transcriptome. Sylamer indicated that the 7mers TGCATTT/GCATTTA, corresponding to the SCR of miR-277 family, were significantly enriched in male (p-value = 1.62E-13, Supplementary S3A Table) and female (p-value = 1.22E-06, Supplementary S3B Table) genes that were more highly expressed in juvenile worms compared to adults (Fig 1A).

Genes with 3′-UTRs containing the 7mers TGCATTT/GCATTTA were selected from the list of transcripts significantly upregulated (adjusted p-value  $\leq 0.01$ ) in juveniles compared to adult worms, using genes found in males (n = 1,099) and females (n = 1,857) separately. Among these transcripts, 225 male and 429 female transcripts contained the 7mers TGCATTT/GCATTTA in their 3′-UTRs (as predicted by Sylamer) and therefore were regarded as potential targets for miRNAs of the sma-miR-277 family. We refer to these groups as male and female “Sylamer genes” (Supplementary S2 Fig).

Secondly, we employed miRanda [54] and TargetScan with conservation evidence [55] to perform miRNA-guided predictions of potential targets in the transcriptome. This latter approach is based on curated three-way alignments of 3′-UTR sequences from orthologous genes from *S. mansoni*, *S. haematobium* and *S. japonicum* (see Materials and Methods). Unlike Sylamer, the combined results of miRanda (Supplementary S4E Table) and TargetScan with conservation (Supplementary S4F Table) only take into account known miRNA seeds in the 3′-UTRs and not the expression of their transcripts.

We further refined this list by selecting those Sylamer genes that also have conserved target sites across the three schistosome species (Fig 1B) as identified by combining miRanda [54] and TargetScan with conservation [55]. Notably, miRanda and TargetScan with conservation alone found 98 potential targets for the sma-miR-277 family in the *Schistosoma* transcriptome (Supplementary S4 Table), considerably higher than for any other miRNA in this study. Of these, 46 genes were also “Sylamer genes” (26 male and 34 female gene with 14 shared between the two genders—Fig 1B and Table 1), we called these “high confidence targets”. By performing a hypergeometric test, we calculated the probability that these overlaps are significant, using as background the 1,476 protein-coding genes that possess at least the core sma-miR-277 6mer GCATTT. Our results show that there is significant enrichment in the overlaps between TargetScan+Miranda vs. Male Sylamer genes (p-value < 0.002) and also between TargetScan+Miranda vs. Male (p-value < 1e-50) while the overlap between TargetScan+Miranda vs. Female Sylamer genes was not significant.

Gene Ontology (Biological Processes) enrichment analysis of the male and female high confidence targets shows significant enrichment (p-value < 0.01) for “chromatin assembly or disassembly” (four genes: Smp\_155060, Smp\_041760, Smp\_158610, Smp\_079650), “regulation of transcription, DNA-templated” (11 genes: Smp\_042300, Smp\_009630, Smp\_147950, Smp\_139530, Smp\_041760, Smp\_158610, Smp\_079650, Smp\_021340, Smp\_163240,

**Table 1. High confidence targets of the sma-miR-277 family identified with a combined approach (Sylamer, miRanda and TargetScan with conservation). F = female; M = male.**

| Gene       | Product description                              | gender |
|------------|--|--------|
| Smp_163500 | Protein Sprint                                   | F      |
| Smp_052420 | centaurin gamma 1A                               | F      |
| Smp_072140 | Rho2 GTPase                                      | F      |
| Smp_042300 | PUR alpha protein                                | F      |
| Smp_009630 | homeobox protein SIX6                            | F      |
| Smp_172240 | stress activated protein kinase JNK              | F      |
| Smp_048560 | Four and a half LIM domains protein 2            | F      |
| Smp_031270 | hypothetical protein                             | F      |
| Smp_041760 | chromodomain helicase DNA binding protein 2      | F      |
| Smp_127440 | Serine rich adhesin for platelets                | F      |
| Smp_173900 | pyruvate dehydrogenase phosphatase               | F      |
| Smp_015530 | hypothetical protein                             | F      |
| Smp_156480 | hypothetical protein                             | F      |
| Smp_078270 | mps one binder kinase activator 2 like           | F      |
| Smp_199010 | ras GTP exchange factor son of sevenless         | F      |
| Smp_210730 | prolyl oligopeptidase (S09 family)               | F      |
| Smp_141980 | cAMP specific 3',5' cyclic phosphodiesterase 4D  | F      |
| Smp_194640 | serine:threonine protein kinase MARK2            | F      |
| Smp_198380 | Argonaute  | F      |
| Smp_143070 | kelch 2, Mayven                                  | F      |
| Smp_105970 | Caprin 1   | M      |
| Smp_043100 | Heterogeneous nuclear ribonucleoprotein A1       | M      |
| Smp_155060 | phosphatase 2a inhibitor i2pp2a                  | M      |
| Smp_072110 | programmed cell death involved protein           | M      |
| Smp_211100 | inner nuclear membrane protein man1              | M      |
| Smp_123630 | ubiquitin carboxyl terminal hydrolase 49         | M      |
| Smp_100090 | high mobility group protein B1                   | M      |
| Smp_063040 | actin 6a   | M      |
| Smp_169320 | protein phosphatase 1g                           | M      |
| Smp_006320 | KH domain containing, RNA binding, signal        | M      |
| Smp_044810 | zinc finger CCCH domain containing protein 14    | M      |
| Smp_008830 | SWI:SNF related, matrix associated, actin        | M      |
| Smp_167320 | serine threonine rich antigen                    | M/F    |
| Smp_129840 | hypothetical protein                             | M/F    |
| Smp_173380 | tyrosine kinase                                  | M/F    |
| Smp_147950 | bromodomain containing 2                         | M/F    |
| Smp_139530 | tumor protein p73                                | M/F    |
| Smp_172700 | cyclin dependent kinase 6                        | M/F    |
| Smp_158610 | TATA binding protein associated factor 172       | M/F    |
| Smp_079650 | chromodomain helicase DNA binding protein 4      | M/F    |
| Smp_021340 | max dimerization protein 3                       | M/F    |
| Smp_163240 | t box transcription factor tbx2                  | M/F    |
| Smp_200020 | FAD NAD binding oxidoreductase                   | M/F    |
| Smp_149000 | protein phosphatase 2C                           | M/F    |
| Smp_166070 | ras specific guanine nucleotide releasing factor | M/F    |
| Smp_165280 | protein groucho                                  | M/F    |

<https://doi.org/10.1371/journal.pntd.0005559.t001>

Smp\_100090, Smp\_008830) and “protein dephosphorylation” (three genes: Smp\_173900, Smp\_169320, Smp\_149000) categories (Supplementary [S5 Table](#)).

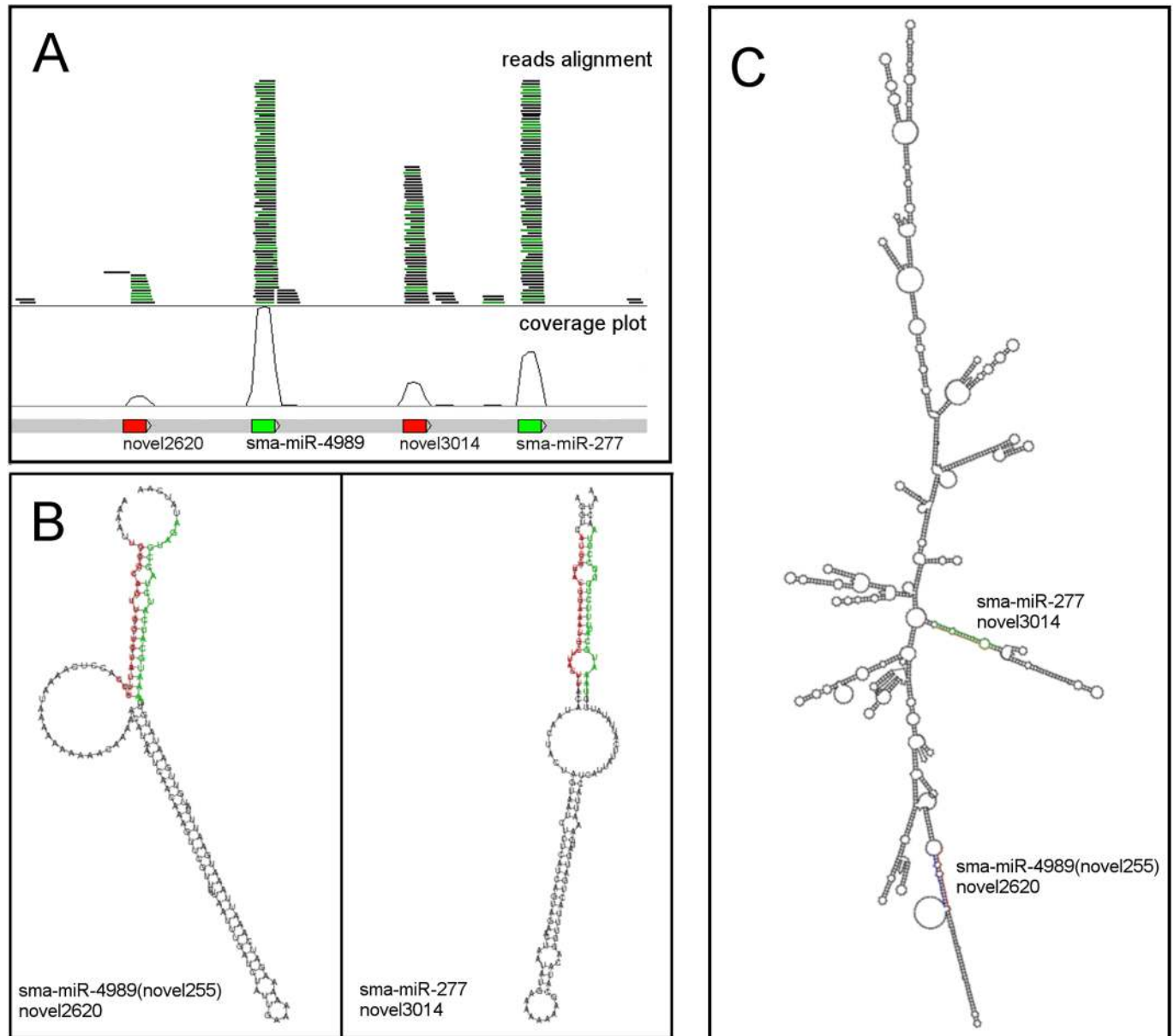
## 2.2. Genomic localisation, structural characterisation of sma-miR-277 and identification of sma-miR-4989

Using *Drosophila melanogaster* dme-miR-277 as a query in the miRBase database (Release 21, [56]) we found full-length sequence similarity and seed sequence (nucleotides 2–7) identity with sja-miR-277 from *S. japonicum* [28] as well as with two *Echinococcus* spp miRNA sequences namely egr- or emu-miR-277 and miR-4989 [74].

Sma-miR-277 has previously been found in vesicles derived from cultured larval stages of *S. mansoni* [75] as well as circulating in naturally infected patients and experimentally infected animals [76]. However, in both reports the genomic location of this miRNA was not described. In tapeworms, miR-277 is located within a cluster with miR-4989 [77], i.e. they are very close together in the genome (less than 10 Kb) and miRNAs encoded in clusters are likely to be transcribed as a single precursor unit [78]. A search for the genome locus of miR-277 in *S. mansoni* showed that it is found in Chromosome 4 (21600667–21600688—reverse). We extended our analysis to include alignments of previously generated small RNA-seq libraries resulting in the identification of another miR-277 family member located in the vicinity of sma-miR-277. Sma-miR-4989 (named by us “novel255”) is located within ~200 bp of sma-miR-277 suggesting that, as is the case in tapeworms, these two miRNAs form a miRNA cluster. Other sequencing reads found in the same region suggested the presence of passenger strands or miRNA\* (Fig 2A). We found that sma-miR-4989 (novel255) folds into a stable structure with its passenger strand (novel2620), the two miRNAs matching to form the lower part in a precursor with an extended stem. The miRNAs sma-miR-277 and its passenger (novel3014) match up similarly, again forming the lower part of an extended stem (Fig 2B). These two extended stem-loop structures were consistently identified among different excisions of the region. They persist when folding the whole (1,700 nt) genomic loci (Fig 2C), where the region folds into a highly stable stem-rich structure with low minimum free energy. To assess the significance of this fold, we randomly shuffled its sequence while preserving dinucleotide frequency. An ensemble of 1,000 randomly shuffled sequences was folded. This yielded a distribution of minimum free energy scores, leading to a Z-score of 3.5 for minimum free energy of the genomic sequence fold, confirming the stability of the stem-rich structure. FASTA sequences for the hairpins, mature and passenger strands are presented in Supplementary [S2 Dataset](#) and an image comparing the folds obtained for different *S. mansoni* predicted miRNAs is presented in Supplementary [S4 Fig](#).

## 2.3. Targets of sma-miR-277/sma-miR-4989 are down regulated in paired females compared with unpaired females

Female worms only achieve sexual maturity when they are coupled with a male worm. Although highly unlikely in natural infections, it is possible to obtain unpaired “virgin” females and unpaired males by infecting mice with single-sex cercariae. Using RNASeq data from paired and unpaired males and females, we found that 16 out of the 34-female high-confidence targets of sma-miR-277 family had significantly (adjusted p-value <0.01) lower expression in paired females compared to unpaired “virgin” females (Fig 3B), while the male targets were not differentially expressed between paired and unpaired males (Fig 3A). Furthermore, the time-course expression analysis indicated that the expression of these targets was gradually reduced as female maturation progressed. Notably, this was not observed in males.

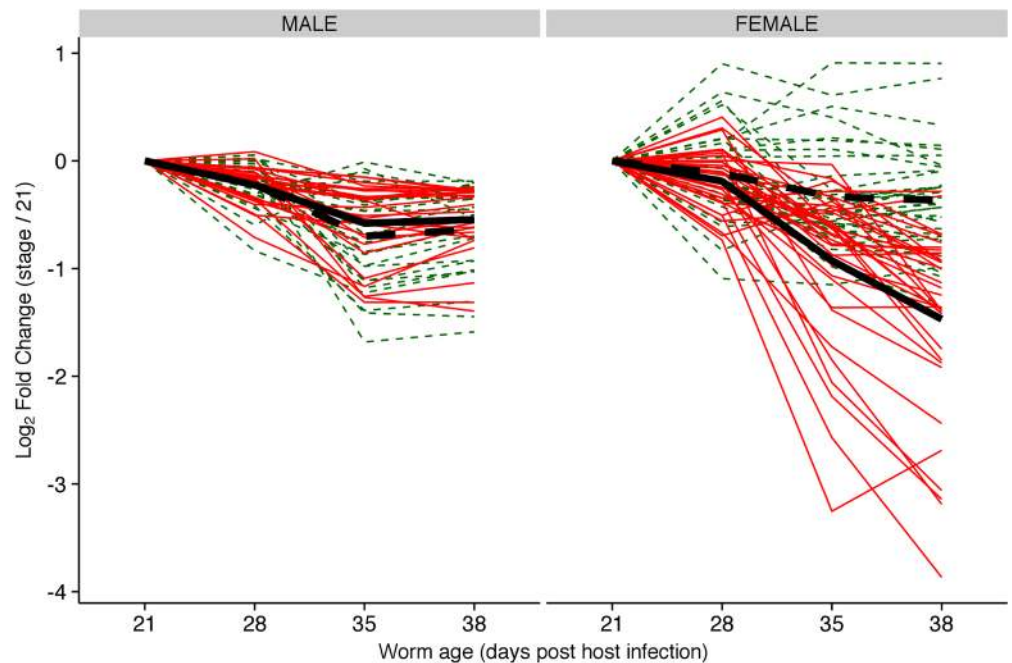


**Fig 2. Sma-miR-277 and sma-miR-4989 belong to a gene cluster.** (A) The genomic locus in Chromosome 4 of sma-miR-277 and sma-miR-4989 suggests they belong to a gene cluster. The average distance between genes (represented by coloured boxes) is 109 bases. Here the mature miRNAs (sma-miR-277 and sma-miR-4989) and passenger miRNAs are represented with coverage plot and aligned reads from one of the small RNA libraries. (B) Predicted stem-loop structures for sma-miR-277 and sma-miR-4989 –individual cases. Mature miRNAs are located in the 3'-end of the hairpin. (C) Due to the cluster organisation of sma-miR-277/sma-miR-4989, it is likely that they are transcribed as one precursor RNA molecule. This figure represents the predicted stem-loop structure for sma-miR-277/sma-miR-4989 when arising from a larger transcript.

<https://doi.org/10.1371/journal.pntd.0005559.g002>

#### 2.4. Sma-miR-4989 is up-regulated during juvenile to adult transition

Tentative miR-277 targets were down-regulated during the juvenile to adult transition. If the miRNA effect is that of causing the decay of its targets, it is expected that the expression of miRNAs from that family would show an opposite pattern to that of its targets, i.e. while the targets of the miR-277 family are down-regulated, sma-miR-277 and sma-miR-4989 should increase their expression.



**Fig 3. Sma-miR-277 family predicted targets downregulated in developing female parasites.** Fold change expression ( $\text{Log}_2$ ) of high confidence targets of sma-miR-277 family during the development of male and female worms in two conditions: paired (solid line red) and unpaired (dashed green). Black lines represent the mean expression of genes in paired (solid black line) and unpaired (dashed black line) worms.

<https://doi.org/10.1371/journal.pntd.0005559.g003>

The expression sma-miR-4989 was quantified by RT-qPCR during juvenile to adult development. This miRNA showed significant up-regulated (28 vs. 49 d.p.i., males t-test p-value = 0.0097; females t-test p-value = 0.01) expression in developing worms (Fig 4) suggesting an association between increased expression in miRNA levels and decreased expression of their targets. Furthermore, sma-miR-277 is also up-regulated in a similar manner, as well as sma-miR-4989 passenger miRNA novel2620 (Supplementary S3 Fig).

## 2.5. Sma-miR-4989 mRNA is localised in the male anterior oesophagus gland

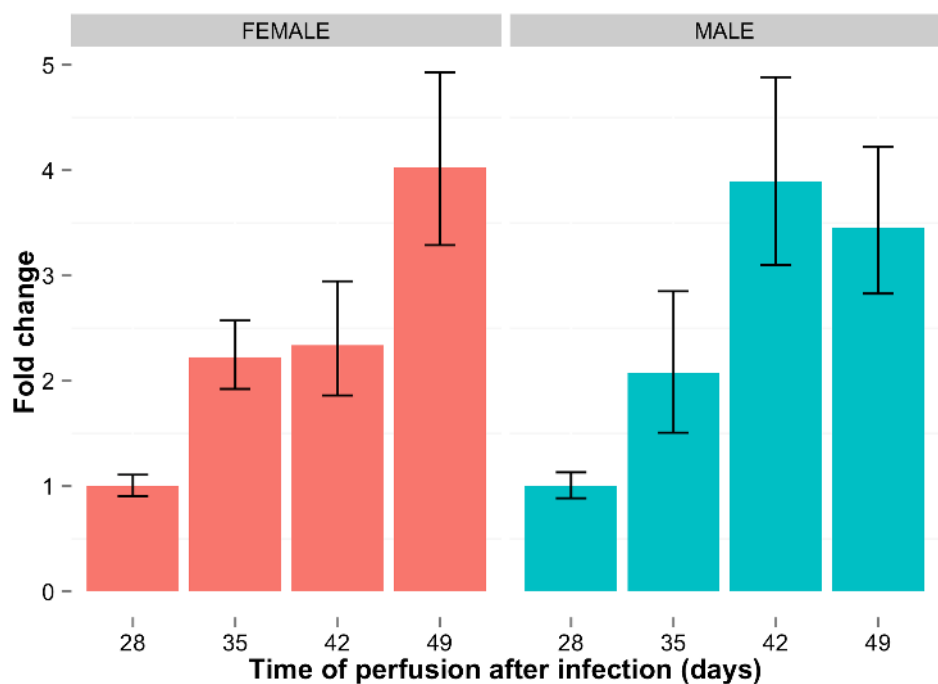
In order to gain insight into the potential functions of sma-miR-4989, we used both whole mount *in situ* hybridisation (WISH), as well as fluorescence *in situ* hybridization (FISH) in adult male and female worms. Given the small size of miRNAs we employed anti-sense Locked Nucleic Acid (LNA) probes to detect the cells expressing these RNAs. To determine the specificity of this approach we performed control experiments with LNA probes targeting a 22nt long sequence of the *CathepsinB* messenger RNA (Smp\_103610), which is detected in the intestine by WISH [68,79]. Consistent with previous reports, we robustly detected *cathepsinB* in the intestine (Fig 5A). To further assess the feasibility to detect Schistosome miRNAs using LNA probes we also examined sma-miR-124a-3p, whose closely related *C. elegans* homolog (miR-124a) is expressed in the nervous system [80]. Consistent with the localization of miR-124a in *C. elegans*, we detected the expression of sma-miR-124a-3p in the schistosome cephalic ganglia and in the nerve chords by WISH (Fig 5B). To determine if this miRNA was broadly expressed in the cephalic ganglia or in subsets of cells in the cephalic ganglia, we performed double FISH with *prohormone convertase 2* (*pc2*), that is expressed in the large number of cells in the schistosome nervous system. Schistosomes possess a pair of cephalic ganglia whose

neuronal cell bodies surround a neuropil composed of a network of neural projections. The individual ganglia are connected via a commissure of neural projections that extend across the midline. We observed that sma-miR-124a-3p and *pc2* were co-expressed in cell bodies that comprise both the cephalic ganglia and the nerve chords (Supplementary S5A Fig). Interestingly, we detected sma-miR-124a-3p not just in *pc2*<sup>+</sup> cell bodies but also broadly throughout the neuropil of the cephalic ganglia (Supplementary S5B Fig). Since the neuropil is comprised of neural projections (axons and dendrites) these data suggest that sma-miR-124a-3p may have functions to regulate its target mRNAs outside the cell bodies.

Given our success in detecting sma-miR-124a-3p, we set out to determine the localization of sma-miR-4989. By WISH, we failed to detect reproducible signal in female parasites. However, we strongly detected sma-miR-4989 in cells surrounding the male oesophagus (Fig 5C) as well as cells throughout in the parenchyma. Since many of the sma-miR-4989-expressing cells in the parenchyma appeared to be quite superficial, just beneath the parasites muscle layer, we explored the possibility these cells comprise the schistosome syncytial epidermis, a structure known as the tegument. To test this, we utilized a mix of probes targeting the mRNAs of four well-characterized tegumental factors: *calpain*, *npp-5*, *annexin* and *gtp-4* (ref. [69–73]). By FISH we observed that the sma-miR-4989 RNA was broadly expressed in cells that make up the schistosome tegument (Fig 5D).

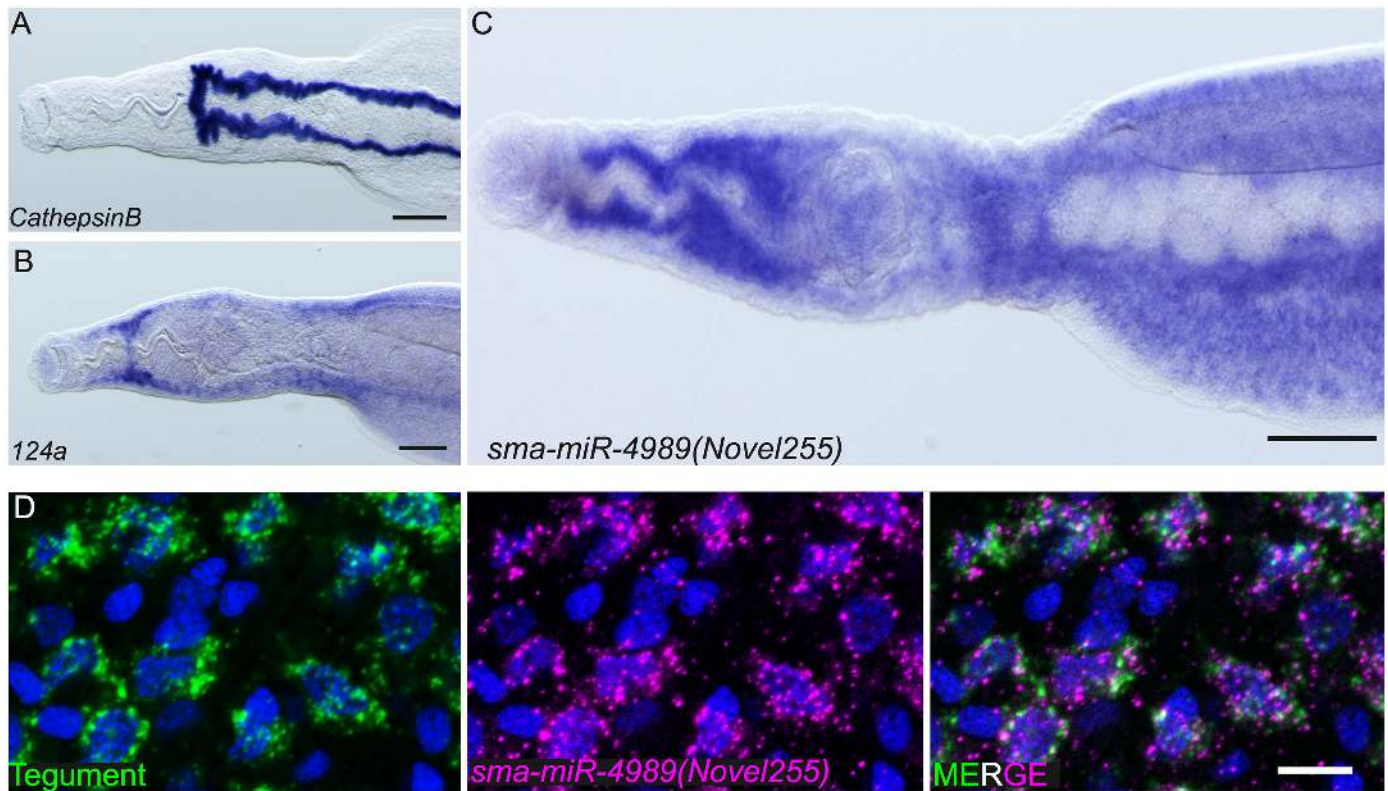
## Discussion

In this study, the contribution of miRNAs in shaping the transcriptional landscape of developing *S. mansoni* juveniles was evaluated by performing a Sylamer analysis [36]. This approach is



**Fig 4. Sma-miR-4989 is significantly up-regulated during male and female maturation.** Fold change expression of sma-miR-4989 during development of juvenile to adult worms in male (blue bars) and females (red bars) as measured by RT-qPCR. Samples were collected at the time points (days post infection) indicated in the x-axis from murine hosts infected with pooled (mixed sex) cercariae. Each barplot represents the mean of three biological replicates. T-tests were performed between 49 d.p.i. and 28 d.p.i. and were both significant with p-value  $\leq 0.01$ . Error bars show the standard error of the mean, based on three biological replicates.

<https://doi.org/10.1371/journal.pntd.0005559.g004>



**Fig 5. Sma-miR-4989 is expressed in the cells surrounding the oesophagus and cells of the tegument in adult worms.** Whole mount *in situ* hybridisation for (A) *cathepsin B*, (B) *sma-miR-124a-3p* (*124a*), and (C) *sma-miR-4989*. (D) Fluorescence *in situ* hybridisation showing the colocalization of *sma-miR-4989* with four co-expressed tegument-specific mRNAs (*calpain*, *npp-5*, *annexin* and *gtp-4*). Nuclei are stained with DAPI and shown in blue. Anterior of worms is to the left in A-C. Scale Bars: A-C 100  $\mu$ m; D 10  $\mu$ m.

<https://doi.org/10.1371/journal.pntd.0005559.g005>

independent from any prior miRNA information available, i.e. miRNA-unaware. Using Sylamer we were able to show that transcripts containing a target site for members of the miR-277 family are highly enriched in early juvenile worms and their expression dramatically decreases towards adulthood (Fig 1A). Remarkably, no other known miRNA seeds were found enriched in this analysis, suggesting that members of the miR-277 family may be the primary miRNAs exerting post-transcriptional regulation in the transition from juvenile to adult worm.

Sylamer results were then coupled with a stringent miRNA target prediction approach that uses miRanda [54] and TargetScan with species conservation [55]. The latter algorithm allows the inclusion of evolutionarily conserved miRNA targets from the three main schistosoma species. The rather low overlap between the Sylamer miRNA-unaware approach and the miRNA-driven TargetScan with conservation and miRanda (12+14+20 targets, Fig 1B) would not represent a disagreement between methods; on the contrary, they complement each other. While the combination of TargetScan with conservation and miRanda provides a highly confident list of miRNA targets, Sylamer provides a statistical approach to the identification of transcripts potentially targeted by a given miRNA based on their co-regulated change in expression. This combined approach, which includes the aforementioned unaware miRNA-target finder component, has never before been applied to schistosomes and our results provide solid evidence for the role of miRNAs during the intra-mammalian development of this parasite. We conclude that the overlap among these methods does identify a group of genes regulated by members of the miR-277. Further *in silico* functional characterisation of the targets showed

that the miR-277 family might be responsible for the down-regulation of important transcription factors, transcription factors associated proteins and signalling molecules (Table 1), for instance a TATA binding box protein associated factor, Tbox transcription factor, a homolog of the *S. mediterranea* p53 protein known to regulate proliferation and self-renewal in stem-cells [81] and a groucho domain-containing protein. The *Drosophila* Groucho (Gro) protein is a corepressor whose action is required, among other processes, for sexual determination [82]. In schistosomes, the 3'-UTR of the gene encoding groucho (Smp\_165280) has three miRNA target recognition sites suggesting close regulation by sma-miR-277/4989 (Supplementary S6 Table). Interestingly, one of the three Argonaute proteins known to be encoded in the *S. mansoni* genome [83] is found among the miR-277 targets. Argonaute is part of the RNA-induced silencing complex (RISC), a key player in the RNA interference pathway [84] and the potential targeting of Argonaute by a miRNAs could suggest a possible feedback loop between the expression of miRNAs and the RNAi effector pathway.

Sixteen of the high confidence targets (potential miR-277 targets that are down-regulated towards adulthood) are also differentially expressed in “virgin” unpaired females compared to sexually active mature females, suggesting that miR-277 family members are involved in reproductive development. Other miRNAs have recently been linked to reproductive development in *S. japonicum* [32]. The post-transcriptional regulation of key players during the transition between developmental stages is consistent with the role predicted for certain miRNAs in other organisms (reviewed in [35]).

The miR-277 family is of particular interest in that, to date, it has only been found in protostomes [85]. In *Schmidtea mediterranea*, a model free-living flatworm for the study of tissue regeneration, Sasidharan *et al.* identified four members of the miR-277 family [86]. In three cestodes (*E. granulosus*, *E. multilocularis* and *Hymenolepis microstoma*), the gene loci encoding miR-277 and miR-4989 (another member of the miR-277 family) are ~ 430 bases apart, comprising a miRNA cluster [87] with gene expression potentially co-regulated. In schistosomes, sma-miR-277 has been detected in sera from experimentally-infected animals and humans naturally infected with schistosomes [76] as well as in secreted vesicles of schistosomula larvae [75]. A recent publication featuring miRNA profiling in developing *S. japonicum* worms identified both sja-miR-277 and sja-miR-277b [32], the latter being identical to our sma-miR-4989. Our results suggest that sma-miR-4989 is the *S. mansoni* homolog of sja-miR-277b.

The high quality of the *S. mansoni* genome assembly allowed us to localise both sma-miR-277 and sma-miR-4989 within a locus containing clustered miRNA genes (Fig 2A) in Chromosome 4. Given the conserved architecture of this miRNA genomic locus when compared to that of tapeworms, and the sequence conservation, we have therefore named novel255 as sma-miR-4989. The miRNA precursor structure results for sma-miR-4989 and sma-miR-277 were found to require longer flanking regions to achieve a thermodynamically stable stem-loop structure (Fig 2B and Fig 2C). This structural difference found in these *Schistosoma* miRNA precursors (Supplementary S4B Fig and Supplementary S4C Fig), when compared to canonical structures (i.e. Supplementary S4A Fig) described for model organisms, could potentially explain previous claims that flatworms might have lost many of the conserved miRNA families [85]. In addition, the highly fragmented nature of the current *S. japonicum* [88] and previous *S. mansoni* genome assemblies [89] may explain the difficulties previously encountered in identifying these particular miRNAs.

Regarding miRNA expression, our results show that sma-miR-4989 displays the same pattern of expression in male and female worms during development (Fig 4). These results are in agreement with recently published work which showed that *S. japonicum* sja-miR-277 and sja-miR-277b (homolog to sma-miR-4989) are both differentially expressed in the transition from juvenile to adult as well as in between male and female *S. japonicum* worms [32]. Given that



the expression pattern of sma-miR-277/4989 and their predicted targets show an inverse correlation, we speculate that the post-transcriptional effect of these miRNAs might be exerted by degradation of the target mRNA [34].

As part of our functional characterisation of sma-miR-4989, we performed whole-mount *in situ* hybridisation (WISH) experiments to localise the site of expression of this miRNA in the worms. Our results show that sma-miR-4898 is expressed in tegumental cells along the whole of the male worm body but its expression is most pronounced in the cells surrounding oesophagus in proximity to the oesophageal glands [90,91]. Detailed transmission electron microscopy of the oesophageal region has shown that the schistosome oesophagus has two distinct regions [92]: the posterior region that is involved in initialling the blood meal while the cellular structure of the anterior region has been described as similar to the tegument, with cell bodies located beneath the muscle fibres and projecting cytoplasmatic extensions that end in the oesophageal lumen. The discoid bodies and multilaminar vesicles that are exported to the oesophageal lumen reveal the glandular nature of the anterior oesophagus and suggest that their secretions might constitute the building blocks of the membranous lining of the oesophagus [90,91,93]. At present, it is difficult to determine if sma-miR-4898 is expressed in these oesophageal glands or in the tegument-like tissue that lines the oesophagus. Although RNA contents of adult oesophagus gland secretions have yet to be analysed in detail, a study of the *S. mansoni* schistosomula vesicles and exosomes [75] identified several known and novel miRNAs, including sma-miR-277. Given the localisation of sma-miR-4989 in the tegumental cells, we speculate that these molecules could be reaching the exterior of the worm in vesicle-like structures. Further work is needed to test this hypothesis.

In summary, the high quality nature of the *S. mansoni* genome and gene annotations allowed us to query 3'-UTR regions for miRNAs target recognition sites that are conserved across three Schistosoma species. By incorporating gene expression information, we were able to conclude that sma-miR-277 and sma-miR-4989 may be the primordial miRNAs driving an effect on the transcriptional landscape regulating gene expression changes underlying the development of juveniles and sexual maturation. Further, we showed that sma-miR-277 and sma-miR-4989 expression is increased towards adulthood, this is consistent with evidence that their targets are down-regulated during the same transition. Finally, we report the first use of a standard WISH technique to the localisation of miRNAs in schistosomes, demonstrating that sma-miR4989 is expressed in tegumental cells and raising many questions about the potential roles of this miRNA.

## Supporting information

**S1 Fig. Reverse transcription PCR reactions for 12 UTRs.** Lanes from left to right: Ladder; 1, 2: Smp\_152790.1; 3, 4: Smp\_149640.1, 5, 6: Smp\_147920.1; 7, 8: Smp\_144140.1; 9, 10: Smp\_200240.1; 11, 12: Smp\_159780.1; 13, 14: Smp\_186020.1; 15, 16: Smp\_159570.1; 17, 18: Smp\_133210.1; 19, 20: Smp\_213910.1; 21, 22: Smp\_154340.1; 23–30: empty; ladder; positive control. Odd numbers represent the test PCR while even numbers represent the “-RT” (without reverse transcriptase) control. (DOCX)

**S2 Fig. Distribution of targets (mRNA) of sma-miR-277 miRNA family across the expression spectrum of all transcripts in adult worms.** Sylamer and high confidence targets of sma-miR-4989(novel255) are spread across the differential expression ranking set of genes. Each point represents a transcript that is significantly higher expressed in juveniles. Highlighted in orange are those transcripts that, in addition, contained a sma-miR-277 miRNA family target in their UTR according to Sylamer. Blue dots represent the subset of Sylamer genes whose

orthologs in *S. haematobium* and *S. japonicum* also have a sma-miR-277 miRNA family target. These are regarded as high confidence targets as they are confirmed independently by three methods (i.e. Sylamer, miRanda and TargetScan with conservation).

(DOCX)

**S3 Fig. Expression of selected miRNAs during juvenile to adult development.** Fold change expression of sma-miR-277, novel2620 and sma-miR-4989(novel255) during development of juvenile to adult worms in male (blue bars) and females (red bars) as measured by RT-qPCR. Samples were collected at the time points (days post infection) indicated in the x-axis from murine hosts infected with pooled (mixed sex) cercariae. In the case of sma-miR-4989 (novel255) these data were independently collected from that shown in Fig 4 of the main text. Error bars represent the standard error of the mean, based on three biological replicates.

(DOCX)

**S4 Fig. Stem-loop structures predicted for known and novel small RNAs in *Schistosoma mansoni*.** (A) Examples of “canonical” hairpins. (B) and (C) Examples of hairpins that required longer flanking regions to achieve a thermodynamically stable stem-loop structure. Part (C) depicts those with a characteristic side “bulge”.

(DOCX)

**S5 Fig. Fluorescence *in situ* hybridization of sma-miR-124.** Fluorescence *in situ* hybridization (FISH) in adult male worms using a Locked Nucleic Acid (LNA) probe to detect sma-miR-124a-3p (instead of antisense mRNA probes) and *prohormone convertase 2* (*pc2*), that is expressed in the large number of cells in the schistosome nervous system. (A) shows detection of the expression of sma-miR-124a-3p and *pc2* in the schistosome cephalic ganglia and in the nerve chords. (B) zoomed-in view of (A) showing the presence of sma-miR-124a-3p in the neural projections that connect the two cephalic ganglia (indicated with arrows) while *pc2* is restricted to the cell bodies in the cephalic ganglia.

(DOCX)

**S1 Table. List of primers (gene id and sequence) used for UTR validation.**

(XLS)

**S2 Table. Kraken output from small RNA sequencing processing.**

(XLS)

**S3 Table. Sylamer table output listing identified 7-mers and their respective p-value.**

(XLS)

**S4 Table. List of miRNA targets found using MiRanda and TargetScan with conservation in the three *Schistosoma* species (*S. mansoni*, *S. japonicum* and *S. haematobium*).**

(XLS)

**S5 Table. Gene Ontology (GO) term enrichment (Biological processes only) of high confidence sma-miR-277 and sma-miR-4989(novel255) targets.**

(XLS)

**S6 Table. Location of sma-miR-4989(novel255) target sites (SCR) in *S. mansoni* transcripts.**

(XLS)

**S1 Text. RNA isolation using phase extraction and ETOH precipitation—suitable for samples with high carbohydrate content.**

(DOCX)

**S1 Dataset. GTF (annotation) file used for transcriptome analysis.**

(ZIP)

**S2 Dataset. FASTA sequences for the hairpins, mature and passanger strands.**

(TXT)

## Acknowledgments

We thank Professor Mike Doehhoff (University of Nottingham, UK) for providing parasite materials. We are grateful to colleagues at the Wellcome Trust Sanger Institute: Dr Simon Clare for assistance with animal infections and Dr Caroline Durrant for assistance with the statistical analysis of rt-qPCR data. We also thank Drs Steve Doyle (Wellcome Trust Sanger Institute) and Jeffrey Philippon for manuscript discussions and suggestions, and George Wendt (UT Southwestern) for providing probes used for detecting tegumental cells.

## Author Contributions

**Conceptualization:** AVP MB.

**Formal analysis:** AVP SvD LQ DMR MH AJE.

**Funding acquisition:** JJC MB.

**Investigation:** AVP JC LQ DMR FS.

**Project administration:** AVP.

**Resources:** GR.

**Software:** AVP SvD LQ MH DMR.

**Supervision:** AVP MB AJE.

**Validation:** AVP SvD.

**Visualization:** AVP SvD JJC.

**Writing – original draft:** AVP SvD MB.

**Writing – review & editing:** AVP GR JJC MB.

## References

1. Rollinson D (2009) A wake up call for urinary schistosomiasis: reconciling research effort with public health importance. *Parasitology* 136: 1593–1610. <https://doi.org/10.1017/S0031182009990552> PMID: [19627633](https://pubmed.ncbi.nlm.nih.gov/19627633/)
2. Mann VH, Morales ME, Rinaldi G, Brindley PJ (2010) Culture for genetic manipulation of developmental stages of *Schistosoma mansoni*. *Parasitology* 137: 451–462. <https://doi.org/10.1017/S0031182009991211> PMID: [19765348](https://pubmed.ncbi.nlm.nih.gov/19765348/)
3. Lewis FA, Liang YS, Raghavan N, Knight M (2008) The NIH-NIAID schistosomiasis resource center. *PLoS Negl Trop Dis* 2: e267. <https://doi.org/10.1371/journal.pntd.0000267> PMID: [18665228](https://pubmed.ncbi.nlm.nih.gov/18665228/)
4. Hagen J, Scheerlinck JP, Gasser RB (2015) Knocking down schistosomes—promise for lentiviral transduction in parasites. *Trends Parasitol* 31: 324–332. <https://doi.org/10.1016/j.pt.2015.03.009> PMID: [25933926](https://pubmed.ncbi.nlm.nih.gov/25933926/)
5. Mann VH, Suttiprapa S, Skinner DE, Brindley PJ, Rinaldi G (2014) Pseudotyped murine leukemia virus for schistosome transgenesis: approaches, methods and perspectives. *Transgenic Res* 23: 539–556. <https://doi.org/10.1007/s11248-013-9779-3> PMID: [24474164](https://pubmed.ncbi.nlm.nih.gov/24474164/)

6. Suttiprapa S, Rinaldi G, Tsai IJ, Mann VH, Dubrovsky L, et al. (2016) HIV-1 Integrates Widely throughout the Genome of the Human Blood Fluke *Schistosoma mansoni*. *PLoS Pathog* 12: e1005931. <https://doi.org/10.1371/journal.ppat.1005931> PMID: 27764257
7. Yang S, Brindley PJ, Zeng Q, Li Y, Zhou J, et al. (2010) Transduction of *Schistosoma japonicum* schistosomules with vesicular stomatitis virus glycoprotein pseudotyped murine leukemia retrovirus and expression of reporter human telomerase reverse transcriptase in the transgenic schistosomes. *Mol Biochem Parasitol* 174: 109–116. <https://doi.org/10.1016/j.molbiopara.2010.07.007> PMID: 20692298
8. You H, Gobert GN, Cai P, Mou R, Nawaratna S, et al. (2015) Suppression of the Insulin Receptors in Adult *Schistosoma japonicum* Impacts on Parasite Growth and Development: Further Evidence of Vaccine Potential. *PLoS Negl Trop Dis* 9: e0003730. <https://doi.org/10.1371/journal.pntd.0003730> PMID: 25961574
9. Davis A (2002) Schistosomiasis. In: Manson PS, Cook GC, Zumla A, editors. *Manson's tropical diseases*. 21st ed. Edinburgh: Saunders. pp. 1434–1469.
10. Paveley RA, Aynsley SA, Cook PC, Turner JD, Mountford AP (2009) Fluorescent imaging of antigen released by a skin-invading helminth reveals differential uptake and activation profiles by antigen presenting cells. *PLoS Negl Trop Dis* 3: e528. <https://doi.org/10.1371/journal.pntd.0000528> PMID: 19829705
11. Hansell E, Braschi S, Medzihradsky KF, Sajid M, Debnath M, et al. (2008) Proteomic analysis of skin invasion by blood fluke larvae. *PLoS Negl Trop Dis* 2: e262. <https://doi.org/10.1371/journal.pntd.0000262> PMID: 18629379
12. Wilson RA (2009) The saga of schistosome migration and attrition. *Parasitology* 136: 1581–1592. <https://doi.org/10.1017/S0031182009005708> PMID: 19265564
13. Basch PF (1991) *Schistosomes: development, reproduction, and host relations*. New York: Oxford University Press. vii, 248 p. p.
14. Fishelson Z, Amiri P, Friend DS, Marikovsky M, Petitt M, et al. (1992) *Schistosoma mansoni*: cell-specific expression and secretion of a serine protease during development of cercariae. *Exp Parasitol* 75: 87–98. PMID: 1639166
15. Stirewalt MA, Minnick DR, Fregeau WA (1966) Definition and collection in quantity of schistosomules of *Schistosoma mansoni*. *Trans R Soc Trop Med Hyg* 60: 352–360. PMID: 5919625
16. Protasio AV, Dunne DW, Berriman M (2013) Comparative study of transcriptome profiles of mechanical- and skin-transformed *Schistosoma mansoni* schistosomula. *PLoS Negl Trop Dis* 7: e2091. <https://doi.org/10.1371/journal.pntd.0002091> PMID: 23516644
17. Parker-Manuel SJ, Ivens AC, Dillon GP, Wilson RA (2011) Gene expression patterns in larval *Schistosoma mansoni* associated with infection of the mammalian host. *PLoS Negl Trop Dis* 5: e1274. <https://doi.org/10.1371/journal.pntd.0001274> PMID: 21912711
18. Grevelding CG, Sommer G, Kunz W (1997) Female-specific gene expression in *Schistosoma mansoni* is regulated by pairing. *Parasitology* 115 (Pt 6): 635–640.
19. Anderson L, Amaral MS, Beckedorff F, Silva LF, Dazzani B, et al. (2015) *Schistosoma mansoni* Egg, Adult Male and Female Comparative Gene Expression Analysis and Identification of Novel Genes by RNA-Seq. *PLoS Negl Trop Dis* 9: e0004334. <https://doi.org/10.1371/journal.pntd.0004334> PMID: 26719891
20. Copeland CS, Marz M, Rose D, Hertel J, Brindley PJ, et al. (2009) Homology-based annotation of non-coding RNAs in the genomes of *Schistosoma mansoni* and *Schistosoma japonicum*. *BMC Genomics* 10: 464. <https://doi.org/10.1186/1471-2164-10-464> PMID: 19814823
21. Cai P, Hou N, Piao X, Liu S, Liu H, et al. (2011) Profiles of small non-coding RNAs in *Schistosoma japonicum* during development. *PLoS Negl Trop Dis* 5: e1256. <https://doi.org/10.1371/journal.pntd.0001256> PMID: 21829742
22. Cai P, Piao X, Hao L, Liu S, Hou N, et al. (2013) A deep analysis of the small non-coding RNA population in *Schistosoma japonicum* eggs. *PLoS One* 8: e64003. <https://doi.org/10.1371/journal.pone.0064003> PMID: 23691136
23. de Souza Gomes M, Muniyappa MK, Carvalho SG, Guerra-Sa R, Spillane C (2011) Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. *Genomics* 98: 96–111. <https://doi.org/10.1016/j.ygeno.2011.05.007> PMID: 21640815
24. Hao L, Cai P, Jiang N, Wang H, Chen Q (2010) Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics* 11: 55. <https://doi.org/10.1186/1471-2164-11-55> PMID: 20092619
25. Marco A, Kozomara A, Hui JH, Emery AM, Rollinson D, et al. (2013) Sex-biased expression of microRNAs in *Schistosoma mansoni*. *PLoS Negl Trop Dis* 7: e2402. <https://doi.org/10.1371/journal.pntd.0002402> PMID: 24069470

26. Simoes MC, Lee J, Djikeng A, Cerqueira GC, Zerlotini A, et al. (2011) Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics* 12: 47. <https://doi.org/10.1186/1471-2164-12-47> PMID: [21247453](https://pubmed.ncbi.nlm.nih.gov/21247453/)
27. Sun J, Wang S, Li C, Ren Y, Wang J (2014) Novel expression profiles of microRNAs suggest that specific miRNAs regulate gene expression for the sexual maturation of female *Schistosoma japonicum* after pairing. *Parasit Vectors* 7: 177. <https://doi.org/10.1186/1756-3305-7-177> PMID: [24721600](https://pubmed.ncbi.nlm.nih.gov/24721600/)
28. Wang Z, Xue X, Sun J, Luo R, Xu X, et al. (2010) An "in-depth" description of the small non-coding RNA population of *Schistosoma japonicum schistosomulum*. *PLoS Negl Trop Dis* 4: e596. <https://doi.org/10.1371/journal.pntd.0000596> PMID: [20161724](https://pubmed.ncbi.nlm.nih.gov/20161724/)
29. Xue X, Sun J, Zhang Q, Wang Z, Huang Y, et al. (2008) Identification and characterization of novel microRNAs from *Schistosoma japonicum*. *PLoS One* 3: e4034. <https://doi.org/10.1371/journal.pone.0004034> PMID: [19107204](https://pubmed.ncbi.nlm.nih.gov/19107204/)
30. Gomes MS, Cabral FJ, Jannotti-Passos LK, Carvalho O, Rodrigues V, et al. (2009) Preliminary analysis of miRNA pathway in *Schistosoma mansoni*. *Parasitol Int* 58: 61–68. <https://doi.org/10.1016/j.parint.2008.10.002> PMID: [19007911](https://pubmed.ncbi.nlm.nih.gov/19007911/)
31. Cai P, Piao X, Hou N, Liu S, Wang H, et al. (2012) Identification and characterization of argonaute protein, Ago2 and its associated small RNAs in *Schistosoma japonicum*. *PLoS Negl Trop Dis* 6: e1745. <https://doi.org/10.1371/journal.pntd.0001745> PMID: [22860145](https://pubmed.ncbi.nlm.nih.gov/22860145/)
32. Zhu L, Zhao J, Wang J, Hu C, Peng J, et al. (2016) MicroRNAs Are Involved in the Regulation of Ovary Development in the Pathogenic Blood Fluke *Schistosoma japonicum*. *PLoS Pathog* 12: e1005423. <https://doi.org/10.1371/journal.ppat.1005423> PMID: [26871705](https://pubmed.ncbi.nlm.nih.gov/26871705/)
33. Ha M, Kim VN (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 15: 509–524. <https://doi.org/10.1038/nrm3838> PMID: [25027649](https://pubmed.ncbi.nlm.nih.gov/25027649/)
34. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215–233. <https://doi.org/10.1016/j.cell.2009.01.002> PMID: [19167326](https://pubmed.ncbi.nlm.nih.gov/19167326/)
35. Rougvié AE, Moss EG (2013) Developmental transitions in *C. elegans* larval stages. *Curr Top Dev Biol* 105: 153–180. <https://doi.org/10.1016/B978-0-12-396968-2.00006-3> PMID: [23962842](https://pubmed.ncbi.nlm.nih.gov/23962842/)
36. van Dongen S, Abreu-Goodger C, Enright AJ (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nat Methods* 5: 1023–1025. <https://doi.org/10.1038/nmeth.1267> PMID: [18978784](https://pubmed.ncbi.nlm.nih.gov/18978784/)
37. Grevelding CG, Kampkotter A, Kunz W (1997) *Schistosoma mansoni*: sexing cercariae by PCR without DNA extraction. *Exp Parasitol* 85: 99–100. <https://doi.org/10.1006/expr.1996.4129> PMID: [9024208](https://pubmed.ncbi.nlm.nih.gov/9024208/)
38. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, et al. (2012) A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis* 6: e1455. <https://doi.org/10.1371/journal.pntd.0001455> PMID: [22253936](https://pubmed.ncbi.nlm.nih.gov/22253936/)
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
40. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169. <https://doi.org/10.1093/bioinformatics/btu638> PMID: [25260700](https://pubmed.ncbi.nlm.nih.gov/25260700/)
41. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
42. "R: A language and environment for statistical computing." (2015). R Foundation for Statistical Computing, Vienna, Austria. <http://www.r-project.org/>.
43. Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27: 2325–2329. <https://doi.org/10.1093/bioinformatics/btr355> PMID: [21697122](https://pubmed.ncbi.nlm.nih.gov/21697122/)
44. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511–515. <https://doi.org/10.1038/nbt.1621> PMID: [20436464](https://pubmed.ncbi.nlm.nih.gov/20436464/)
45. You FM, Huo N, Gu YQ, Luo MC, Ma Y, et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9: 253. <https://doi.org/10.1186/1471-2105-9-253> PMID: [18510760](https://pubmed.ncbi.nlm.nih.gov/18510760/)
46. Palmer RD, Murray MJ, Saini HK, van Dongen S, Abreu-Goodger C, et al. (2010) Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets. *Cancer Res* 70: 2911–2923. <https://doi.org/10.1158/0008-5472.CAN-09-3301> PMID: [20332240](https://pubmed.ncbi.nlm.nih.gov/20332240/)

47. Lewis MA, Quint E, Glazier AM, Fuchs H, De Angelis MH, et al. (2009) An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat Genet* 41: 614–618. <https://doi.org/10.1038/ng.369> PMID: [19363478](https://pubmed.ncbi.nlm.nih.gov/19363478/)
48. Davis MP, Abreu-Goodger C, van Dongen S, Lu D, Tate PH, et al. (2012) Large-scale identification of microRNA targets in murine Dgcr8-deficient embryonic stem cell lines. *PLoS One* 7: e41762. <https://doi.org/10.1371/journal.pone.0041762> PMID: [22912678](https://pubmed.ncbi.nlm.nih.gov/22912678/)
49. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120: 15–20. <https://doi.org/10.1016/j.cell.2004.12.035> PMID: [15652477](https://pubmed.ncbi.nlm.nih.gov/15652477/)
50. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545–15550. <https://doi.org/10.1073/pnas.0506580102> PMID: [16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
51. Morgulis A, Gertz EM, Schaffer AA, Agarwala R (2006) A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol* 13: 1028–1040. <https://doi.org/10.1089/cmb.2006.13.1028> PMID: [16796549](https://pubmed.ncbi.nlm.nih.gov/16796549/)
52. Medina-Rivera A, Defrance M, Sand O, Herrmann C, Castro-Mondragon JA, et al. (2015) RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic Acids Res* 43: W50–56. <https://doi.org/10.1093/nar/gkv362> PMID: [25904632](https://pubmed.ncbi.nlm.nih.gov/25904632/)
53. Stephenson AG (2002) evd: Extreme Value Distributions. *R News* 2: 31–32.
54. Enright AJ, John B, Gaul U, Tuschl T, Sander C, et al. (2003) MicroRNA targets in *Drosophila*. *Genome Biol* 5: R1. <https://doi.org/10.1186/gb-2003-5-1-r1> PMID: [14709173](https://pubmed.ncbi.nlm.nih.gov/14709173/)
55. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, et al. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of *Isy-6* and other microRNAs. *Nat Struct Mol Biol* 18: 1139–1146. <https://doi.org/10.1038/nsmb.2115> PMID: [21909094](https://pubmed.ncbi.nlm.nih.gov/21909094/)
56. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42: D68–73. <https://doi.org/10.1093/nar/gkt1181> PMID: [24275495](https://pubmed.ncbi.nlm.nih.gov/24275495/)
57. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, et al. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 19: 327–335. <https://doi.org/10.1101/gr.073585.107> PMID: [19029536](https://pubmed.ncbi.nlm.nih.gov/19029536/)
58. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
59. Alexa A, Rahnenfuhrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22: 1600–1607. <https://doi.org/10.1093/bioinformatics/btl140> PMID: [16606683](https://pubmed.ncbi.nlm.nih.gov/16606683/)
60. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63: 41–49. <https://doi.org/10.1016/j.ymeth.2013.06.027> PMID: [23816787](https://pubmed.ncbi.nlm.nih.gov/23816787/)
61. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 39: D152–157. <https://doi.org/10.1093/nar/gkq1027> PMID: [21037258](https://pubmed.ncbi.nlm.nih.gov/21037258/)
62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
63. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584. PMID: [11917018](https://pubmed.ncbi.nlm.nih.gov/11917018/)
64. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, et al. (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* 41: W597–600. <https://doi.org/10.1093/nar/gkt376> PMID: [23671338](https://pubmed.ncbi.nlm.nih.gov/23671338/)
65. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/)
66. Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, et al. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol* 6: 26.
67. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. *Methods* 25: 402–408. <https://doi.org/10.1006/meth.2001.1262> PMID: [11846609](https://pubmed.ncbi.nlm.nih.gov/11846609/)
68. Collins JJ 3rd, Wendt GR, Iyer H, Newmark PA (2016) Stem cell progeny contribute to the schistosome host-parasite interface. *Elife* 5: e12473. <https://doi.org/10.7554/eLife.12473> PMID: [27003592](https://pubmed.ncbi.nlm.nih.gov/27003592/)
69. Skelly PJ, Shoemaker CB (1996) Rapid appearance and asymmetric distribution of glucose transporter SGTP4 at the apical surface of intramammalian-stage *Schistosoma mansoni*. *Proc Natl Acad Sci U S A* 93: 3642–3646. PMID: [8622989](https://pubmed.ncbi.nlm.nih.gov/8622989/)

70. van Balkom BW, van Gestel RA, Brouwers JF, Krijgsveld J, Tielens AG, et al. (2005) Mass spectrometric analysis of the *Schistosoma mansoni* tegumental sub-proteome. *J Proteome Res* 4: 958–966. <https://doi.org/10.1021/pr050036w> PMID: 15952743
71. Braschi S, Wilson RA (2006) Proteins exposed at the adult schistosome surface revealed by biotinylation. *Mol Cell Proteomics* 5: 347–356. <https://doi.org/10.1074/mcp.M500287-MCP200> PMID: 16269422
72. Rofatto HK, Tararam CA, Borges WC, Wilson RA, Leite LC, et al. (2009) Characterization of phosphodiesterase-5 as a surface protein in the tegument of *Schistosoma mansoni*. *Mol Biochem Parasitol* 166: 32–41. <https://doi.org/10.1016/j.molbiopara.2009.02.006> PMID: 19428670
73. Wilson RA (2012) Proteomics at the schistosome-mammalian host interface: any prospects for diagnostics or vaccines? *Parasitology* 139: 1178–1194. <https://doi.org/10.1017/S0031182012000339> PMID: 22717150
74. Cucher M, Prada L, Mourglia-Ettlin G, Dematteis S, Camicia F, et al. (2011) Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. *Int J Parasitol* 41: 439–448. <https://doi.org/10.1016/j.ijpara.2010.11.010> PMID: 21219906
75. Nowacki FC, Swain MT, Klychnikov OI, Niazi U, Ivens A, et al. (2015) Protein and small non-coding RNA-enriched extracellular vesicles are released by the pathogenic blood fluke *Schistosoma mansoni*. *J Extracell Vesicles* 4: 28665. <https://doi.org/10.3402/jev.v4.28665> PMID: 26443722
76. Hoy AM, Lundie RJ, Ivens A, Quintana JF, Nausch N, et al. (2014) Parasite-derived microRNAs in host serum as novel biomarkers of helminth infection. *PLoS Negl Trop Dis* 8: e2701. <https://doi.org/10.1371/journal.pntd.0002701> PMID: 24587461
77. Cucher M, Macchiaroli N, Kamenetzky L, Maldonado L, Brehm K, et al. (2015) High-throughput characterization of *Echinococcus* spp. metacestode miRNomes. *Int J Parasitol* 45: 253–267. <https://doi.org/10.1016/j.ijpara.2014.12.003> PMID: 25659494
78. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297. PMID: 14744438
79. Collins JJ 3rd, Wang B, Lambrus BG, Tharp ME, Iyer H, et al. (2013) Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature* 494: 476–479. <https://doi.org/10.1038/nature11924> PMID: 23426263
80. Clark AM, Goldstein LD, Tevlin M, Tavare S, Shaham S, et al. (2010) The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. *Nucleic Acids Res* 38: 3780–3793. <https://doi.org/10.1093/nar/gkq083> PMID: 20176573
81. Pearson BJ, Sanchez Alvarado A (2010) A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development* 137: 213–221. <https://doi.org/10.1242/dev.044297> PMID: 20040488
82. Chen G, Courey AJ (2000) Groucho/TLE family proteins and transcriptional repression. *Gene* 249: 1–16. PMID: 10831834
83. Skinner DE, Rinaldi G, Koziol U, Brehm K, Brindley PJ (2014) How might flukes and tapeworms maintain genome integrity without a canonical piRNA pathway? *Trends Parasitol* 30: 123–129. <https://doi.org/10.1016/j.pt.2014.01.001> PMID: 24485046
84. Cenik ES, Zamore PD (2011) Argonaute proteins. *Curr Biol* 21: R446–449. <https://doi.org/10.1016/j.cub.2011.05.020> PMID: 21683893
85. Fromm B, Worren MM, Hahn C, Hovig E, Bachmann L (2013) Substantial loss of conserved and gain of novel MicroRNA families in flatworms. *Mol Biol Evol* 30: 2619–2628. <https://doi.org/10.1093/molbev/mst155> PMID: 24025793
86. Sasidharan V, Lu YC, Bansal D, Dasari P, Poduval D, et al. (2013) Identification of neoblast- and regeneration-specific miRNAs in the planarian *Schmidtea mediterranea*. *RNA* 19: 1394–1404. <https://doi.org/10.1261/rna.038653.113> PMID: 23974438
87. Jin X, Lu L, Su H, Lou Z, Wang F, et al. (2013) Comparative analysis of known miRNAs across platyhelminths. *FEBS J* 280: 3944–3951. <https://doi.org/10.1111/febs.12395> PMID: 23777576
88. Zhou Y, Z H., Chen Y., Zhang L., Wang K., Guo J., Huang Z., Zhang B., Huang W., Jin K., Dou T., Hasegawa M., Wang L., Zhang Y., Zhou J., Tao L., Cao Z., Li Y., Vinar T., Brejova B., Brown D., Li M., Miller DJ., Blair D., Zhong Y., Chen Z., Liu F., Hu W., Wang ZQ., Zhang QH., Song HD., Chen S., Xu X., Xu B., Ju C., Huang Y., Brindley PJ., McManus DP., Feng Z., Han ZG., Lu G., Ren S., Wang Y., Gu W., Kang H., Chen J., Chen X., Chen S., Wang L., Yan J., Wang B., Lv X., Jin L., Wang B., Pu S., Zhang X., Zhang W., Hu Q., Zhu G., Wang J., Yu J., Wang J., Yang H., Ning Z., Beriman M., Wei CL., Ruan Y., Zhao G., Wang S., Liu F., Zhou Y., Wang ZQ., Lu G., Zheng H., Brindley PJ., McManus DP., Blair D., Zhang QH., Zhong Y., Wang S., Han ZG., Chen Z., Wang S., Han ZG., Chen Z. (2009) The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* 460: 345–351.

89. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, et al. (2009) The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352–358. <https://doi.org/10.1038/nature08160> PMID: [19606141](https://pubmed.ncbi.nlm.nih.gov/19606141/)
90. Li XH, de Castro-Borges W, Parker-Manuel S, Vance GM, Demarco R, et al. (2013) The schistosome oesophageal gland: initiator of blood processing. *PLoS Negl Trop Dis* 7: e2337. <https://doi.org/10.1371/journal.pntd.0002337> PMID: [23936568](https://pubmed.ncbi.nlm.nih.gov/23936568/)
91. Li XH, Stark M, Vance GM, Cao JP, Wilson RA (2014) The anterior esophageal region of *Schistosoma japonicum* is a secretory organ. *Parasit Vectors* 7: 565. <https://doi.org/10.1186/s13071-014-0565-8> PMID: [25490864](https://pubmed.ncbi.nlm.nih.gov/25490864/)
92. Spence IM, Silk MH (1970) Ultrastructural studies of the blood fluke—*Schistosoma mansoni*. IV. The digestive system. *S Afr J Med Sci* 35: 93–112. PMID: [5520361](https://pubmed.ncbi.nlm.nih.gov/5520361/)
93. Wilson RA, Li XH, MacDonald S, Neves LX, Vitoriano-Souza J, et al. (2015) The Schistosome Esophagus Is a 'Hotspot' for Microexon and Lysosomal Hydrolase Gene Expression: Implications for Blood Processing. *PLoS Negl Trop Dis* 9: e0004272. <https://doi.org/10.1371/journal.pntd.0004272> PMID: [26642053](https://pubmed.ncbi.nlm.nih.gov/26642053/)