Satellite-like W Elements: repetitive, transcribed, and putative mobile genetic factors with potential roles for biology and evolution of Schistosoma mansoni

Maria Stitz, Cristian Chaparro, Zhigang Lu, V Janett Janett Olzog, Christina E Weinberg, Jochen Blom, Alexander Goesmann, Christoph Grunau, Christoph G Grevelding

Satellite-like W elements: repetitive, transcribed, and putative mobile genetic factors with potential roles for biology and evolution of *Schistosoma mansoni*

Maria Stitz[1*], Cristian Chaparro[2*], Zhigang Lu[1], V. Janett Olzog[3], Christina E. Weinberg[3], Jochen Blom[4], Alexander Goesmann[4], Christoph Grunau[2#], Christoph G. Grevelding[1#]

[1]Institute of Parasitology, BFS, Justus Liebig University Giessen, Germany; [2]IHPE, University Montpellier, CNRS, IFREMER, UPVD, IHPE, F-66000 Perpignan, France; [3]Institute for Biochemistry, Leipzig University, Germany; [4]Bioinformatics and Systems Biology, Justus Liebig University Giessen, Germany.

[*] Authors equally contributing to the work

[#] Authors sharing paper correspondence

Correspondence:

Prof. Christoph G. Grevelding
Justus Liebig University
BFS, Institute of Parasitology
Schubertstr. 81
35392 Giessen
Germany
christoph.grevelding@vetmed.uni-giessen.de

Prof. Christoph Grunau
IHPE UMR 5244
Université de Perpignan Via Domitia
58 Avenue Paul Alduy, Bât R
F-66860 Perpignan Cedex
France
christoph.grunau@univ-perp.fr

Abstract

A large portion of animal and plant genomes consists of non-coding DNA. This part includes tandemly repeated sequences and gained attention because it offers exciting insights into genome biology. We investigated satellite-DNA elements of the platyhelminth *Schistosoma mansoni*, a parasite with remarkable biological features. *S. mansoni* lives in the vasculature of humans causing schistosomiasis, a disease of worldwide importance. Schistosomes are the only trematodes that have evolved separate sexes, and the sexual maturation of the female depends on constant pairing with the male. The schistosome karyotype comprises eight chromosome pairs, males are homogametic (ZZ), females heterogametic (ZW). Part of the repetitive DNA of *S. mansoni* are W-elements (WEs), originally discovered as female-specific satellite DNAs in the heterochromatic block of the W-chromosome. Based on new genome and transcriptome data, we performed a reanalysis of the W-element families (WEFs). Besides a new classification of 19 WEFs, we provide first evidence for stage-, sex-, pairing-, gonad-, and strain-specific/preferential transcription of WEs as well as their mobile nature, deduced from autosomal copies of full-length and partial WEs. Structural analyses suggested roles as sources of non-coding RNA like hammerhead ribozymes (HHRs), for which we obtained functional evidence. Finally, the variable WEF occurrence in different schistosome species revealed remarkable divergence. From these results we propose that WEs potentially exert enduring influence on the biology of *S. mansoni*. Their variable occurrence in different strains, isolates, and species suggests that schistosome WEs may represent genetic factors taking effect on variability and evolution of the family Schistosomatidae.

Running Title: W-elements of *S. mansoni*

Key Words: *Schistosoma mansoni*, genome evolution, W-element, mobile genetic element, non-coding RNA, ribozyme.

Significance Statement

Previous studies described W-elements (WE) as repeated, non-coding satellite-DNA of the large heterochromatic part of the female-specific W-chromosome of *S. mansoni* with unknown function. We challenge this view by analysing all W-elements families (WEF), their structures, WEF-dependent transcript profiles (including stage-, sex-, pairing-, gonad-, and strain-specific/preferential expression), and autosomal occurrence, which indicates their mobile nature. Furthermore, we predicted roles of WEs as carriers of genetic information such as non-coding RNA, for which obtained biochemical evidence. Analysing different schistosome strains, isolates and even species finally showed the variable existence of WEF. These finding suggest that W elements might play roles for the biology of *S. mansoni* and might represent one of the driving forces in the evolution of the family Schistosomatidae.

Introduction

Schistosomes are parasitic plathyhelminths and cause schistosomiasis (bilharzia). Listed as neglected tropical disease (NTD) by the WHO, schistosomiasis ranks close to malaria in terms of parasite-induced human morbidity and mortality (Hotez and Kamatha 2009; Colley at al. 2014). The life cycle of the parasite is complex with a fresh-water snail of the genus *Biomphalaria* as intermediate host, producing human-dwelling larvae (cercariae) that develop in the vertebrate host into adults, pair and produce eggs that are excreted with the faeces and hatch snail-infecting miracidia when in contact with water. Schistosomes are evolutionary unique as they are the only digenean parasites that have evolved separate sexes (Moné and Boissier 2004). Heteromorphic sex chromosomes determine the sex of schistosomes, with ZZ in the males and ZW in the females. Earlier studies from us and others has shown that the females-specific region of the W chromosome is essentially composed of repetitive DNA sequences (W-Elements, WE) arranged in large satellite blocks (Lepesant et al. 2012a; Lepesant et al. 2012b). WE sequences are abundant on the W chromosome, but may occasionally be transmitted to autosomes, shown for the female W1 repeat sequence (Grevelding 1999). In most species with Y or W sex chromosomes, it is found that: (i) repetitive sequences accumulate on these chromosomes, (ii) large regions are heterochromatic, and (iii) these chromosomes deteriorate or are completely absent in an extreme case. Among the accepted evolutionary theories is that the accumulation of repetitive sequences on one of sex chromosomes has facilitated recombination suppression between the heterochromosomes thus protecting sexually beneficial loci (Ohno 1967). Another hypothesis is that chromosome rearrangements, sequence accumulation, and amplification may have occurred near the sex-determining locus as a result of suppression of recombination (Charlesworth 1991; Marais and N. Galtier 2003). Heterochromatization of the W chromosome in schistosomes has long been known and has even been used as a marker for sex identification at the cercaria stage, where male and female phenotypes are indistinguishable (Grossman et al. 1980; Grossman et al. 1981; Liberatos & Short 1983). Our earlier results (Lepesant et al. 2012a) showed that the repetitive sequences located in the heterochromatic region of the W chromosome carry a euchromatic signature in the miracidia stage (presence of H3K9ac and H3K4me3), which is gradually lost during development into adults. In most species, the structural change of the chromatin is a highly organised process, and the observed heterochromatin/euchromatin cycle of W-specific repeats suggests functional importance. Full or partial euchromatisation could not only allow transcription of functional WE but may also open space for mobilisation of these elements. The WEs W1 and W2 were originally found in Puerto Rican isolates of *Schistosoma mansoni* (Webster et al. 1989; Drew and Brindley 1995; Spotila et al 1987). Southern blot analysis demonstrated their female-specific occurrence, and i*n situ* hybridization localized W1 and W2 in the heterochromatic region of the W-chromosome (Hirai et al. 1989; Spotila et al. 1989). According to their repetitive nature, both elements were classified as satellite-like DNA (satDNA).

Subsequent studies in the Liberian strain and further strains of *S*. *mansoni* demonstrated W1 and W2 occurrence in both sexes of nearly all investigated strains, including another isolate from Puerto Rico (Grevelding 1995; Quack et al. 1998). In contrast to the polymorphic occurrence of W1 and W2, another family of repetitive elements, called D9 (Spotila et al. 1989), repeatedly occurred in the investigated schistosome strains (Quack et al. 1998). Molecular analyses of clonal populations obtained from crossing experiments with characterized parental generations, including males with and without W1 and W2, showed the variable emergence of both WE among 10 generations of male progeny - even when the parental male exhibited no W1 and W2 (Grevelding 1999). These data indicated a meiotic level of (illegitimate) recombination and suggested a Z-chromosomal or autosomal existence of W1 and W2 in the male progeny of these crosses. However, diversity in W1/W2 abundance and presence occurred also among siblings of single offspring generations. This pointed to an additional process generating variable copy numbers of WEs in the genome. Indeed, the analysis of clonal cercariae, generated by monomiracidial snail infections, demonstrated differences in W1 and W2 copy numbers within clonal cercariae obtained from the same snails but at different time points post infection (Grevelding 1999). This was unexpected since clonal cercariae were considered to be genetically identical (Jourdane and Theron 1980; Jones and Kusel 1989). Additionally, the genetic heterogeneity found in individuals of the same clonal cercarial population led to the hypothesis that mitotic recombination may contribute to variable WE copy numbers in males, because only asexual but no sexual reproduction processes take place in the intermediate host (Grevelding 1999). Using an *in vitro* technique that allowed the generation of defined clonal daughter sporocysts originating from a single mother finally confirmed that W1 copy numbers can vary among daughter sporocysts generated by one defined mother (Bayne and Grevelding 1993). These results clearly pointed to the possibility of mitotic recombination or transposition in the intermediate host. The underlying mechanisms of WE variability remained obscure.

A microarray study with RNA of female cercariae provided first evidence for W1 and W2 transcripts (Fitzpatrick et al 2008). The authors hypothesized that these transcripts may be associated to germ-line protection in female schistosomes such as interfering with retrotransposable element activity, as previously proposed for repeat elements in *Drosophila* (Pélisson et al. 2007). In 2012, a comprehensive sequencing analysis to *de-novo* assemble *S. mansoni* repeat elements based on version 5 (V5) of the genome (Berriman et al. 2009) predicted 36 WEFs (Lepesant et al. 2012b).

In addition to their peculiar evolutionary position as the only dioeciously living trematodes, another unique feature of schistosome biology is the essentiality of a constant pairing contact for the sexual maturation of the female. Pairing induces mitotic activity and differentiation processes in the female that finally lead to the development of the female gonads, ovary and vitellarium (Popiel and Basch 1984; Den Hollander and Erasmus 1985; Kunz 2001; Grevelding 2004). This is a prerequisite for egg

production and closely associated with the pathology of schistosomiasis since eggs, which fail to reach the gut lumen, migrate via the blood stream to liver and spleen. Here, these eggs lodge in the tissues causing granuloma formation, inflammatory processes, and fibrosis (Olveda et al. 2014). The sexual maturation status of a female is reversible. Upon separation, egg production stops and females de-differentiate to an immature status. Upon repairing, differentiation of the gonads and egg production start again (Clough 1981). Although the underlying processes have not been completely understood, recent transcriptomics approaches have highlighted a complex scenario of male-female interaction. A comparative RNA-Seq analysis of paired and unpaired adult *S. mansoni* and their gonads showed the occurrence of >7,000 gene transcripts in the gonads of both sexes, of which 243 (testes) and 3,600 (ovaries) were pairing-dependently transcribed. High numbers of differentially expressed genes in the ovary were expected because of the pairing-induced sexual maturation of females (Erasmus 1973; Shaw 1987; Kunz 2001; Neves et al., 2005; Beckman et al., 2010). Among others, evidence was obtained for the participation of neuronal processes, guided by GPCRs (G protein-coupled receptors) and neuropeptides, in male-associated processes of the male-female interaction (Hahnel et al. 2018; Lu et al. 2019). In addition, kinase signaling seem to dominate processes in females leading to gonad differentiation and the maintenance of the sexual maturation status (Grevelding et al. 2018).

In this study, we challenge the classical view that repetitive DNA on the sex chromosomes is simply a by-product of heterochromatization and provide further evidence for their functional importance. We made use of genome sequencing updates in combination with recently obtained transcriptome data sets to re-analyse WEFs, their structures, their chromosomal occurrence, their physical relationships, and their developmentally regulated and strain-associated transcriptional activities. From the obtained results we propose a hypothesis for their functional roles in schistosome biology and evolution.

Results

There are 19 W-Element Families in *S. mansoni*

To get a new overview of WEFs in *S. mansoni*, we performed a local BLAST (Basic Local Alignment Search Tool) search of the published WEF (Lepesant et al. 2012b), which were based on version 5 (V5) of the genome, against version 7 (V7; PRJEA36577, Puerto Rican strain; https://parasite.wormbase.org; Howe et al. 2017). Except for W32, we found all 36 examined WEF with a high percent identity on W-scaffolds, which cover DNA sequences, which are likely derived from/correspond to the female-specific sex chromosome W but have not been exactly assembled yet. We detected differences in DNA-sequence similarity among individual WEs of a single WEF. For instance, the ten most significant hits of W6 elements (monomer repeat length: 310 bp) showed no mismatch and correspond 100%, whereas the ten most significant hits of W23 elements (monomer repeat length: 125 bp) showed deletions and mismatches resulting in 76-100% identity

(Supplementary Figure 1). W1, W8, W10, W13, W17, W18, W20, W22, W31, W34, and W35 exclusively matched to W-scaffolds (Table 1). Some WEs, or partial versions thereof, also occurred on autosomes. BLAST analyses revealed that some WEs of a single WEF clustered in single chromosomal regions, whereas other WEFs occurred as separate clusters in different chromosomal regions. For WEF 26 (W26.2), we detected clusters of different total lengths in two yet undefined regions of W, provisionally designated W003 and W010 in V7 (Table 1). Although WEs of all 36 examined WEFs aligned to V7, repeat units of W9, W15, W19, and W32 showed no clustering of multiple, aligned repeats. Instead, we found a wide distribution of mostly partial fragments of these WEs on sex chromosomes and autosomes (data not shown). Because these WEs presented no multiple repeat character, we excluded them from further analysis.

Table 1. BLAST Results of WEFs in the Genome Version 7 of *S. mansoni*

No.	WE (V7)	WE (V5)	WE monomer lengths	Number of copies	Scaffold	Start	Stop	Total length of the array
1	W1.2	W1; W23	475	270	SM_V7_W003	320,915	513,415	192,500
2	W2.2	W2; W3	709-711	up to 34.1	SM_V7_W001	422,658	479,629	56,971
3	W4.2	W4; W30	1,206	11.7	SM_V7_W004	1	14,111	14,110
4	W5.2	W5	1,104	41.2	SM_V7_W018	1	45,515	45,514
5	W6.2	W6; W18; W35	715-718	up to 45.7	SM_V7_W014	1	53,884	53,883
6	W7.2	W7	980	29	SM_V7_W016	1	28,424	28,423
7	W8.2	W8	538	60.6	SM_V7_W015	1	32,211	32,210
8	W10.2	W10	671	59.2	SM_V7_W021	1	39,689	39,688
9	W11.2_W002	W11; W14; W28	903	33.8	SM_V7_W002	712,429	755,874	43,445
	W11.2_ZW	W11; W14; W28	1,294	29.5	SM_V7_ZW	1,1804,837	11,863,496	58,659
10	W12.2	W12	475-499	up to 46.1	SM_V7_W004	206,491	279,311	72,820
11	W13.2	W13; W17; W20; W33	524-646	up to 23.8	SM_V7_W003	229,593	276,593	47,000
12	W16.2	W16; W21	317	204.2	SM_V7_W007	57,292	122,025	64,733
13	W22.2	W22	604	106.6	SM_V7_W008	1	64,369	64,368
14	W24.2	W24	636	3.6	SM_V7_W020	1	2,259	2,258
15	W25.2	W25	415-428	up to 115.4	SM_V7_W012	1	49,306	49,305
16	W26.2_W003	W26	399-402	up to 146.7	SM_V7_W003	1	161,189	161,188
	W26.2_W010	W26	400	152.6	SM_V7_W010	1	61,065	61,064
17	W27.2	W27; W29	403	101.3	SM_V7_W017	1	40,810	40,809
18	W31.2	W31; W34	260	205.5	SM_V7_W004	41,351	114,609	73,258
19	W36.2_W001	W36	333-335	173.4	SM_V7_W001	1	104,272	104,271
	W36.2_W005	W36	332	up to 450.5	SM_V7_W005	1	149,646	149,645

Tab. 1: Based on genome version V7, a number of 19 WEF (WE V7) has been newly defined (first column, No. 1 - 19). Each WEF consists of one or more of those repeats that were previously found as individual WEs in version 5 of the genome (WE V5; Lepesant et al. 2012a). In some cases (No. 9, 16 and 19), representing WEF (W11.2, W26.2 and W36.2, respectively) WEs were found on two different scaffolds each associated with the W chromosome and thus split into two subfamilies each. Furthermore, information of the length of WE monomers

within WEFs is given as well as copy numbers (floating points indicate partial sequences), scaffolds with start and stop positions, and the total lengths of WEFs of the array. In some cases, the individual lengths of a WE monomer within a unit varied. For instance, a monomer of WEF 2.2 (No. 2) consists of W2 and W3 elements, which showed length variations between 709 to 711 bp. WEF 2.2 occurred "up to" 34.1 times, which means that the length variant 711 bp occurred 34 times whereas the 709 bp variant in a lower amount (Supplementary Figure 2).

Next, we applied dot plot analyses to determine the structures of WEFs and the total lengths of WE monomers within the respective clusters (Figure 1). In relation to individual monomer repeat units, length variations from 82 bp (W32) to 1,132 bp (W4) occurred in V5 (Supplementary Figure 2). Furthermore, we found that some WEFs analysed in V7 consist of sequences previously defined as separate repeat units in V5.. For example, we found W27 and W29, originally described as monomer repeat units of 110 bp and 97 bp, respectively (Lepesant et al. 2012a), to be part of the same repeat unit (Figure 1). According to V7, the W27/W29 monomers are separated by spacer sequences of 134 by and 62 bp, respectively, which together with the W27/W29 monomers form a new monomer of a total length of 403 bp (Supplementary Figure 2 H). This 403 bp monomer represents a single repeat unit of the newly named W27.2 family. This designation relates to the former nomenclature used for these satellite-like repeat elements in schistosome research (Lepesant et al. 2012a). W27.2 consists of about 100 copies of highly similar monomer repeat units that define this WE family (Figure 1 and Supplementary Figure 2). We made similar findings for the majority of WEs/WEFs.

Figure 1. The New Definition of WEF W27.2

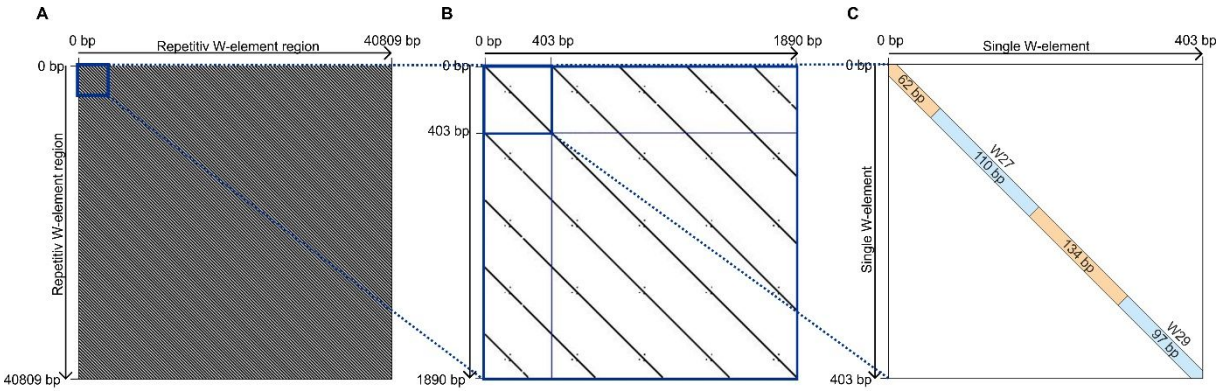


Figure 1: Showcase for the new definition of WEFs in *S. mansoni*. **A**, Result of the dotplot analysis for WE W29 in V7. The regular stripe pattern indicates a tandem repeat structure of this element without noticeable inversions or deletions. **B**, close-up of a small part of A (see blue square in A, upper left corner) showing a length of 403 bp for this WE unit. **C**, a further close-up showing that W29 is closely associated to W27 , which finally led to the new designation WEF W27.2 (Table 1).

Therefore, with respect to the uninterrupted repetitive patterns of WEs within respective chromosome regions as well as small distances or overlaps of different WEs within a genome locus, we

integrated some of the formerly described 36 WEF (Lepesant et al. 2012a), which resulted in newly defined 19 WEFs (Table 1 and Supplementary Figure 2 A and B). These 19 WEFs vary in their copy numbers from 3 (W24.2) to 450 (W36.2) times at one or several chromosomal locations of the V7 genome assembly (Table 1). These numbers may underestimate the actual copy-numbers of the WEF in the genome since large satellite WEF-containing blocks are absent from the assembly.

Some WEs Are Also Present On Autosomes

Former studies reported on the presence of WEs in male *S. mansoni* (Grevelding 1995; Quack et al. 1998). These findings suggested the presence of WE on Z and/or on autosomes. To find evidence for one or both possibilities, we took WE monomers of the newly defined WEFs and WE-transcript sequences detected in males for BLAST analyses against V7 in the NCBI database, which allowed assigning hits to individual chromosomes. For 4 out of 19 WEFs (W1.2, W7.2, W8.2, W10.2), we found no autosomal localization. Represented by either full-length or partial WE sequences, 15 WEFs were distributed among all 7 autosomes (Table 2 and Supplementary Figure 3A).

Table 2. WEFs on Autosomes

WE	Length	Presence on autosomes	Minimal – maximal length	Copy numbers on autosomes	WE on autosomes
W2.2	709-711	all	37 – 705	68	parts
W4.2	1,206	all	44 – 390	1,101	parts
W5.2	1,104	all	26 – 742	3,454	parts
W6.2	715-718	all	36 – 161	5,382	parts
W7.2	980	2,4,6,7	80 – 156	5	parts
W11.2	903-1,294	all	30 – 514	18,610	parts
W12.2	475-499	all	26 – 102	172	parts
W13.2	524-646	1,2,4,6	38 – 131	19	parts
W16.2	317	all	26 – 314	203	parts
W22.2	604	all	26 – 232	1,897	parts
W24.2	636	all	28 – 450	2,302	parts
W25.2	415-428	all	25 – 428	39,648	parts and full-length (408)
W26.2	399-402	all	25 – 399	25,162	parts and full-length (321)
W27.2	403	all	27 – 403	55,303	parts and full-length (174)
W36.2	333-335	all	25 – 335	42,670	parts and full-length (858)

Tab. 2: Summary of the BLAST analyses of WEs against genome version V7 of *S. mansoni* in the NCBI database. Of the 19 newly defined WEFs, representatives of 15 WEF were found to be distributed among all 7 autosomes, whereas W7.2 and W13.2 were found on 4 autosomes. WEs, their original length (bp), their copy numbers on autosomes, and minimal and maximal lengths (bp) of the WE parts are listed. Next, the numbers of full-length

and partial WEs on autosomes are given. Here, the numbers of full-length WEs for W25.2, W26.2, W27.2, and W36.2, respectively, are given in parentheses.

Next, we analysed the distribution of single-size elements on autosomes and found heterogeneous patterns. In most cases, we detected size distributions patterns typical for individual WEF. The majority of autosomal W5.2 copies were comparably small, between 30-120 bp, with a copy-number bias for 56 bp long variants on chromosome 6 and another bias for 119 bp long variants on chromosome 1. Copy-number variants of W11.2 peaked in 241-244 bp long sequences with a clear bias of 412 bp-sized copies on chromosome 1, which contained also most of the dominating 81-88 bp variants of W25.2. W27.2 copies exhibited the largest distribution of different-sized variants across all chromosomes, whereas 150-165 bp long variants dominated the W36.2 family with a bias of 405 copies of 157 bp on chromosome 1. The majority of the longest size variants across all WEF were found on chromosome 1. Among the few exceptions are W12.2 copies with the dominating size variant of 34 bp on chromosomes 5 (Supplementary Figure 3B).

For some of the parental WEFs of those partial, autosomal WEs, we identified sequence similarities to known mobile genetic elements (Supplementary Figure 4). W2.2 showed 65% sequence identity over a stretch of 607 bp to the LTR retrotransposon Saci-1 of *S. mansoni*, which is 5,980 bp in length (DeMarco et al. 2004). Open reading frame (ORF) analysis indicated that this sequence stretch covered the complete protease-coding part of Saci-1. Within WEF W4.2 (1,206 bp) a short but significant sequence homology of 44 bp (89% identity) occurred to the LTR retrotransposon Boudicca (5,858 bp) (Copeland et al. 2003). A 120 bp fragment of W5.2 (1,104 bp) showed 96% identity to the non-LTR retrotransposon Perere-2 (4,544 bp) (DeMarco et al. 2005). This fragment covered part of the reverse transcriptase gene. In W11.2_ZW (1,294 bp), a partial sequence of 150 bp showed 84% identity to the non-LTR retrotransposon Perere-3 (DeMarco et al. 2005). Finally, in W16.2_ZW (1,294 bp) a partial sequence of 317 bp showed 81% identity to the DNA transposon Curupira-1 (4,878 bp) (Jacinto et al. 2011). An analysis of potential ORFs indicated partial sequence homologies of W2.2 to *gag* (Group antigen) and *pol* (a reverse transcriptase), and of W5.2 to endonuclease-reverse transcriptase and *pol* (data not shown). Genes like *gag, pol,* and endonuclease-reverse transcriptase are parts of mobile genetic elements such as retrotransposons or retroviruses. However, we did not find a complete ORF for a transposase or an endonuclease-reverse transcriptase in one of the investigated WEFs. The exception is W36.2_W001 (335 bp), which showed strong sequence similarity to the SMalpha family of SINE-like retrotransposons as reported in an independent study (Ferbeyre et al. 1998). Furthermore, a focused structural analysis on one element showed direct repeat sequences as part of a monomer of WEF W25.2. In this case, duplicate sequence stretches were identified flanking this WE at the presumptive target site. The duplicated sequences differed from the W25.2 sequence and could have originated from target site duplication (TSD) (Supplementary fig 5; data not shown).

WE Transcripts Occur in All Investigated Strains, Life stages, Sexes, and the Gonads of *S. mansoni*

To get a first overview of WEF transcript occurrence in the different biological samples we performed a sample-distance matrix analysis. By pairwise comparisons, this approach provided information about WEF transcript amounts based on the summation of all reads of transcribed WEs - independent of the composition of each WEF that contributed to the WE transcript pool of each sample. These samples included miracidia, cercariae, sporocysts, female and male schistosomula, paired and unpaired males and females, as well as testes and ovaries of paired and unpaired females and males, respectively, and also samples of other strains, depending on availability of sequencing data. We selected data sources with at least two biological replicates for the individual life stages, both sexes, and the gonads (Supplementary Table 1) and normalized the data using DESeq2 prior to further analysis. When available, we also included replicates.

The result indicated remarkable differences in WE transcript amounts among the investigated samples. Furthermore, we observed clustering of samples with similar or dissimilar levels of the total amount of WE transcripts (Supplementary figs. 6 and 7). In the Liberian strain, miracidia (Mir_1, Mir_3, Mir_6) showed high deviations of WE transcripts compared to other stages but these levels differed from miracidia of the Puerto Rican strain (Mir). We detected similar deviations for cercariae of the Puerto Rican strain (Cer1-3) and the Liberian strain (CerM1-CerM3; CerF1-CerF3). Generally, schistosomula samples (SomF/M) and samples from males and testes (sM1-sM3; bM1-sM3; sT1-sT3, bT1-bT3) seemed to be homogeneous with respect to transcript amounts. Bigger differences occurred between unpaired females (sF1-sF3), which revealed an overall higher WE transcript level compared to paired females (bF1-bF3). Finally, we found differences also among biological replicates as exemplified by samples from females and ovaries (bF1-bF3; bO1-bO3; sO1-sO2), and even among replicates as observed for schistosomula (SomM/SomF samples). In summary, this first overview indicated a high degree of variability in overall WE transcript amounts between life stages, strains, and sexes, but also between biological replicates.

WEF Expression Levels Differ Among and Within Strains, Life Stages, Sexes, Pairing-Status and Gonads of *S. mansoni*

Next, we performed differential expression analysis to investigate WEF expression profiles across strains, different life stages (from miracidia to adults), sex-, gonad-, and/or pairing-dependent expression. Again, we discovered remarkable differences (Figure 2 and Supplementary figs. 8 and 9). Strain-dependent expression occurred, among others, for W7.2, with transcripts in schistosomula of the Puerto Rican strain but not in schistosomula of the Guadeloupean strain. We found strain-dependent expression differences also between the Guadeloupean and Liberian strains. For example, W11.2 expression in males of the Guadeloupean strain was higher than the respective levels of males

of the Liberian strain. Also with respect to schistosomula samples, we discovered differences between strains, here Guadeloupean and Puerto Rican. This applied also to miracidia samples from the Liberian and Puerto Rican strains, although here only one biological replicate was available for analysis. Furthermore, differences in stage-dependent expression patterns occurred, within and among strains. Within the Liberian strain, for example, W13.2 appeared to be expressed in miracidia and paired females but not in sporocysts and males, respectively. In contrast, there was no W13.2 expression in females of the Guadeloupean strain. In the latter strain, no W13.2 transcripts occurred in schistosomula in contrast to schistosomula of the Puerto Rican strain, which expressed W13.2. We identified comparable differences for WEF expression comparing cercariae of both strains.

Figure 2. WE Transcript Profiles Across Different *S. mansoni* Strains, Life Stages, Sexes, and Gonads

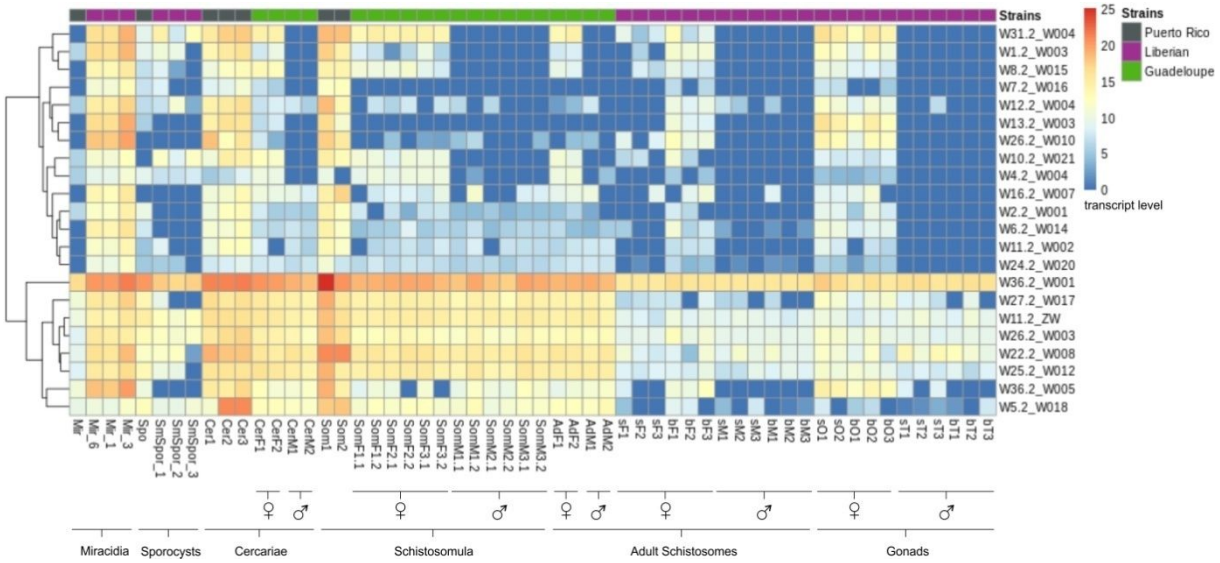


Figure 2: Heatmap generated by Deseq2 analysis showing the hierarchical clustering of transcript profiles of WEs of all 19 WEFs in the following life stages: Mir, miracidia; Cer, cercariae; (Sm)Spo(r), sporocysts; Cer(M/F), cercariae (male/female); Som(F/M), schistosomula (female/male); AdM, adult paired males; sF, unpaired (single-sex) females, bF, paired (bisex) females; sM, unpaired (single-sex) males; bM, paired (bisex) males; sO, ovaries of unpaired females; bO, ovaries of paired females; sT, testes of unpaired males; bT, testes of paired males. When available, biological and technical replicates were included. The first number behind a sample abbreviation indicates the number of the biological replicate. The second number indicates the technical replicate. For example, SomF1.1 indicates the first biological and first technical replicate of a female schistosomula, sample. SomF1.2 is the second technical replicate of this schistosomula sample. Samples without numbers had no replicate (Supplementary Table 1). Furthermore, biological symbols indicate female and male samples. In cases without symbol, the sample origin was mixed-sex. The horizontal line at the top of this figure shows a color code for the different schistosome strains, which were the basis for generating RNA-Seq data (Supplementary Table 1). We used strains from Guadeloupe (beige), Liberia (purple), and Puerto Rico (grey). The dendrogram at the outer left side indicates relationships between appropriate WEs/WEFs as labelled at the outer right side. The

color code indicates various levels of expression (from dark blue = no transcripts (0) to deep red = high transcript level (25) in the form of log2-transformed normalized counts of all data sets used.

Furthermore, we found sex- and gonad-dependent expression patterns such as for W13.2 but also for W31.2. In these and other cases, expression levels were high in paired females and even higher in their ovaries. No W13.2 or W31.2 expression occurred in males and testes, a finding which applied to the majority of WE. In case of W31.2, an additional finding was the high expression level in miracidia and in ovaries, independent of the pairing status. Sex- and pairing-dependent expression occurred, among others, for W11.2 and W13.2 with transcripts in paired females but not in unpaired ones or in males. Also within biological replicates, independent of the strain, we identified in part differences in expression levels of some WEFs. For example, the expression level of W31.2 varied within three biological replicates of cercariae from the Puerto Rican strain. Also in the Liberian strain, WEF 31.2 expression varied among three biological replicates of miracidia, sporocysts, unpaired and paired females and their ovaries, respectively.

Finally, we observed the most persistent profile in these analyses for W36.2_W001, which appeared to be transcribed in all strains, life stages, sexes, and gonads. Overall expression levels of W36.2_W001 were higher in larval compared to adult stages, independent of the strain.

Evidence for Non-Coding RNAs in WEFs

Since WEF transcripts occurred, and since we detected nearly no continuous ORFs indicating protein-coding information, we next investigated whether WEFs may contain ncRNAs. These RNAs encode no proteins, but as regulatory RNAs they can directly influence cellular processes (Mattick and Makunin 2006; Hombach and Kretz 2016). To search for ncRNAs, we analysed the 19 WEFs using StructRNAFinder, which predicts and annotates RNA families in transcripts or genome sequences (Arias-Carrasco et al. 2018).

For sequence parts of WEF W1.2, W5.2, W7.2, W8.2, W11.2_ZW, and W12.2, StructRNAFinder predicted similarities to micro RNA (miRNA) (Supplementary Table 2; Supplementary Figure 10). According to the output of sequence similarities and structural predictions, some seemed more likely than others (Supplementary Figure 11). This class of short ncRNAs is processed from stem-loop regions of longer RNA transcripts and can influence post-transcriptional processes during gene expression (Bartel 2018). Transcripts of WEF possibly encoding miRNAs showed varying transcript amounts in life-cycle stages, strains, sex and tissue. For example, miRNA candidates mir-785 and mir-891 appeared to form stable, miRNA-like hairpin structures (Supplementary Figure 11). Transcripts of mir-785 sequence-containing W12.2 occurred in a strain-influenced (lower expression levels in the Guadeloupean strain compared to the other two strains; Figure 2 and Supplementary Figure 8) as well as stage-restricted and sex/gonad-influenced manner (high expression levels only in miracidia, paired

females and ovaries compared to males and testes of the Liberian strain; Supplementary Figure 9). For mir-891 sequence-containing W8.2, we found sex-, gonad-, strain-, and stage-influenced patterns (Figure 2). Interestingly, in the Guadeloupean strain a clear sex-biased expression level of W8.2 occurred. We detected W8.2 transcripts in female cercariae, female schistosomula and adult females, whereas no W8.2 transcripts occurred in the male-sample counterparts of this strain (Supplementary Figure 8). Among other possibilities, these observations suggest roles for W12.2 and W8.2 in sex-related processes.

For sequence parts of WEFs W2.2, W5.2, W11.2_ZW, W22.2, W24.2, and W26.2_W003, StructRNAFinder predicted similarities to small nucleolar RNAs (snoRNAs) (Supplementary Table 3). Together with associated proteins, snoRNAs form ribonucleoprotein complexes directing the post-transcriptional modification of target RNAs (Lui and Lowe 2013). WEFs potentially encoding snoRNAs showed varying transcript amounts in different strains, life-cycle stages, sexes, and tissues. Among these is a candidate for SNORD59. It is potentially encoded by WEF W2.2, which contains the snoRNA C/D family-specific C-box (UGAUGA) and D-box (CUGA) motifs (Galardi et al. 2002) within the SNORD59 region of W2.2 (Supplementary Figure 12). The W2.2 expression level was higher in cercariae and schistosomula of the Puerto Rican strain compared to the Guadeloupean strain (Supplementary Figure 8). In the Liberian strain, W2.2 showed preferential expression in miracidia, paired females, and ovaries (Supplementary Figure 9). Expression levels of WEFs containing snoRNA candidates TB11Cs2H1 (W26.2_W003) and GlsR19 (W36.2_W001) in adults of the Guadeloupean strain exceeded those from adults of the Liberian strain (Figure 2). In the latter strain, W26.2_W003 and W36.2_W001 expression dominated in miracidia compared to other samples (Supplementary Figure 9). Within the sequence of W22.2, different snoRNA candidates were predicted, sR11, snoZ178, and SCARNA7 (Supplementary Table 3 and Supplementary Figures 10 and 11). W22.2 appeared expressed at higher levels in miracidia and schistosomula of the Puerto Rican strain compared to the Guadeloupean strain. In the Liberian strain, W22.2 expression is higher in miracidia than in other samples (Supplementary Figure 9).

In case of W5.2 and W11.2_ZW, StructRNAFinder predicted sequence parts for both miRNA and snoRNA, which partly overlapped (Supplementary figs. 10 and 11). W5.2 and W11.2_ZW expression levels appeared to be higher in the Puerto Rican strain compared to the Guadeloupean and Liberian strains (Figure 2 and Supplementary Figure 8). However, compared to W11.2_ZW showing constant expression among all samples of the Liberian strain, W5.2 expression was higher in miracidia and sporocysts than all other stages and tissues of this strain (Supplementary Figure 9).

At this stage of the analysis, however, it is unclear whether these predictions of miRNA and snoRNA correspond to biologically relevant ncRNAs. In addition, we found four WEF that contain sequences reminiscent of the HHR class of ribozymes. These are catalytic RNAs and defined as ncRNA molecules that can catalyse chemical reactions (Lilley 2019; see next section).

Some WEF Contain Functional Hammerhead Ribozyme Sequences

To date, there are 14 natural ribozyme classes known that differ by their conserved secondary and tertiary structure. They are grouped according to the chemical reaction they catalyse. Of the 14 classes, 9 cleave their own phosphate backbone at a specific site by catalysing a phosphoester transfer reaction and are therefore called self-cleaving ribozymes (Jimenez 2015, Weinberg et al. 2019). The first HHRs were discovered in plant virus-like satellite RNAs and viroids (Prody et al. 1986). Among the self-cleaving ribozymes, HHRs are abundant and can be found in all domains of life (de la Peña et al. 2010; Perreault et al. 2011; Jimenez et al. 2011; Seehafer et al. 2011; Hammann et al. 2012) including representatives in schistosomes (Ferbeyre et al. 1998).

The HHR class is characterized by three helices (stem I-III) forming a junction that includes 12 highly conserved nucleotides. Together, these elements build the catalytic core of the HHR. If the transcription start and end lies within stem I, these ribozymes are referred to as type I HHRs. In *S. mansoni*, a type I HHR occurs as part of the SMalpha family of SINE-like retrotransposons (Ferbeyre et al. 1998), which is represented by the WEF W36.2_W001 in our study. To investigate the similarity between the HHR sequence of the W36.2_W001 and the HHRs found in other WEFs, we created a multiple sequence alignment using Infernal (Nawrocki et al. 2013). The results showed weak sequence similarity among W25.2, W26.2_W003, W27.2, and W36.2_W001 (Figure 3).

Figure 3. Alignment of WE Example Sequences from Families W36.2, W25.2, W26.2 and W27.2 Compared to SMalpha

```
Smalpha          atcgcacaagcaagtggct--------atcaggactcagtggccgagtggataacgcgat
W36.2_335bp      attgcacaagcaagtggct--------atcaggactcagtggtcgagtggataacgcgat
W25.2_415bp      gttggaaatagacataaacaccatgggatgccgtctcagtggtgtatgtgttgagtgctc
W26.2_400_bp     gttgaagatggaaatgaacacggttggatgccgactcagtgttcttgatgttctcagctt
W27.2_403bp      tttgaagtaaggcattgagagcgttggatggtgacccagtggtctagtgtttaagtgctc
                 *.* *   . . .* .          **   * *.***** .       *       .

Smalpha          ggtgtttgaagcgaggg-tactgggttggagtcccagagtggatatcaactctgagatgc
W36.2_335bp      ggcgtttgaagcgaaagttactgggttcgagtgccagaatggacatcaactctgagatgc
W25.2_415bp      tattgagaaagttgttattcctgtgttggaatctgg----------tgaggcgaaatcgt
W26.2_400_bp     gcgcacgaaaatgataagcgcagttttcgaatctgg----------agaggcggtatcgt
W27.2_403bp      acgcgcgagactgataggtgctgggttggaatctcg----------cgagactgcatcgt
                 ..* . .  . . * *  ** **.* . .          .*  * . . .*.
Smalpha          aggtacatccagctgacgagtccca-aataggacgaaacgcgcgtcctggattccactgc
W36.2_335bp      aggtacatccagctgacgagtccca-aataggacgaaacgcgcgtcctccattccactgc
W25.2_415bp      ggattggaaatgctgaggagtcgcacaatagaaccaaatgaccgtccagtgtttccagga
W26.2_400_bp     ggatgcacactgctgagaaaagatattaatgaagttccaaggctgctcag-attgtcgagg
W27.2_403bp      ggatgcgcactgctgagcagtcccacagtacgacgaaacggccatccagtgcttccaggt
                 .*.*  .    *****  *.      *.*. ...   * *  ....*   ..* .   *

Smalpha          tatccactattcatc--------tttgcttatcatgcttgtgaatcaaggctatatcgag
W36.2_335bp      tagccactatccatc--------tttgcctaccatgcttgtcaatttaggctatatcgag
W25.2_415bp      tttccctagtggtctaccttcaatccactcatgatttcaaata------taaata-cact
W26.2_400_bp     ttaacatgacggtctctcaacaattgattcatcatcataacca------ttaacattacg
W27.2_403bp      tttccatcgtgctccagcttcaattcattcatcatctcaacta------tcaacattact
                 *   * . ..   ..        *. ....*. **  . .  *         *.* ..
```

```
Smalpha         gcaatacgcacagt-----------------atgcacata-tgccaattagagactgacc
W36.2_335bp     gcaatacgcacagt-----------------atgcacaca-tgccaattacagactgacc
W25.2_415bp     caaatctccacaaatgccactcctcataataataataataataataatcaaatgctcacc
W26.2_400_bp    agattatccacaaa-accacttctcatattagtcaacacg-----------agctcact
W27.2_403bp     ataatatctacgaa-acgccttctcatcactgtcaacatt-----------tgctcaac
                * * . .**..                  .*    *.             .** * .

Smalpha         agttgcagtcct-aacacat--------------cgatgggaagattcaaacaaacaata
W36.2_335bp     agttgctgtcctaaaaacat--------------caatgggaagatccaaacaaacaata
W25.2_415bp     aatgactgaattgaagagatatttcgtggagttgtagtgagaagcagtgaccagtggagt
W26.2_400_bp    agtgactgatttcaacaggtatttcctggagttctggtgagaagcagagaccagtgcagt
W27.2_403bp     agtgactggcttgaagaggcatttcctgaagttctcgtgagaagcggtgaccagtggatt
                *.* .* *  .* ** * ..             . .**.****    .* **.   *

Smalpha         ctaagtg-----------------------------------------------------
W36.2_335bp     ctaagta-------------aatttcaact-------------------------------
W25.2_415bp     tgaagccacgtctgttgtgagatgtcaactgactgaagacaattgtgaacg-gttggtga
W26.2_400_bp    tgaagcggtttgttttgaaaaatat---ctcactgaagacattggtggatgtgtcgctga
W27.2_403bp     ggaagcag-gtgtgatgtgagatggtcactcaatgaagacattggtggatgtgtcggtca
                ***.

Smalpha         -------------
W36.2_335bp     ------tcacccc
W25.2_415bp     acttggtggattg
W26.2_400_bp    atatcgtggatcg
W27.2_403bp     tttcagtgaatcg
```

Figure 3: Clustal-based alignment of the WEF W36.2, W25.2, W26.2 and W27.2 compared to the SMalpha family of SINE-like retrotransposons (Ferbeyre et al. 1998), which contain the HHR sequence. The chosen examples show sequence similarity to the highly conserved core sequences CUGANGA and GAAA (grey background), but also deviations occurred (marked in yellow) (Ruffner et al. 1990; Perreault et al. 2011).

Varying sequence similarity for HHRs occurred within these WEF (Figure 4 and Supplementary Figure 13), which we also observed for W-chromosomal and autosomal copies of these WEF showing incomplete or altered HHR consensus sequences with deletions, insertions, and point mutations. As shown before, mutations in the CUGANGA and GAAA consensus motifs can lead to inactivity of HHRs or diminished cleavage speeds (Ruffner et al. 1990).

Previous studies already confirmed the activity of the SMalpha-encoded type I HHR, represented by W36.2_W001 (Ferbeyre et al. 1998; Canny et al. 2004). Prior to experimental validation of activities of predicted HHRs of other WEFs, we analysed the appropriate sequences for intact three-stem junctions and perfect conservation in the CUGANGA and GAAA sequence motifs to exclude presumably inactive HHRs (Supplementary Table 4) (Ruffner et al. 1990). To investigate additional HHRs for their self-cleavage activity *in vitro*, we selected predicted candidates from different WEs (Figure 4). Because of the high similarity between all candidates (Supplementary Figure 13), we selected only five representatives. These candidates descend from WEF copies on autosomes and represent variants of W25.2 and W36.2 (Figure 4C and Supplementary Table 4). Although the HHR core mainly comprised the three-stem junction with its conserved CUGANGA and GAAA sequences (Figure 4A), additional interactions outside the catalytic core increased ribozyme cleavage speed and structural stability

(Khvorova et al. 2003; Uhlenbeck et al. 2003; Martick et al. 2006; Perreault et al. 2011). Therefore, we extended predicted HHR sequences by 11-13 nucleotides, which naturally occur at the 5'- and 3'-ends of the motifs.

Our analysis confirmed the activity of HHR candidates 1-3, which cleave into the expected 5'- and 3'-subfragments (Figure 4D). Due to the imprecise run-off transcription of the T7 RNA polymerase, we observed additional bands for the 3'-cleavage fragment (Chamberlin et al. 1973). Candidate 4 contains a mutation in the catalytic core (CUGANAA), and candidate 5 has a deletion at the cleavage site. Both alterations likely rendered these candidates inactive based on detailed studies on the effect of mutations to the HHR core on cleavage (Ruffner et al. 1990). In summary, we confirm all predictions in this co-transcriptional cleavage assay.

Figure 4. Active Hammerhead Ribozymes are Part of Autosomal WEs

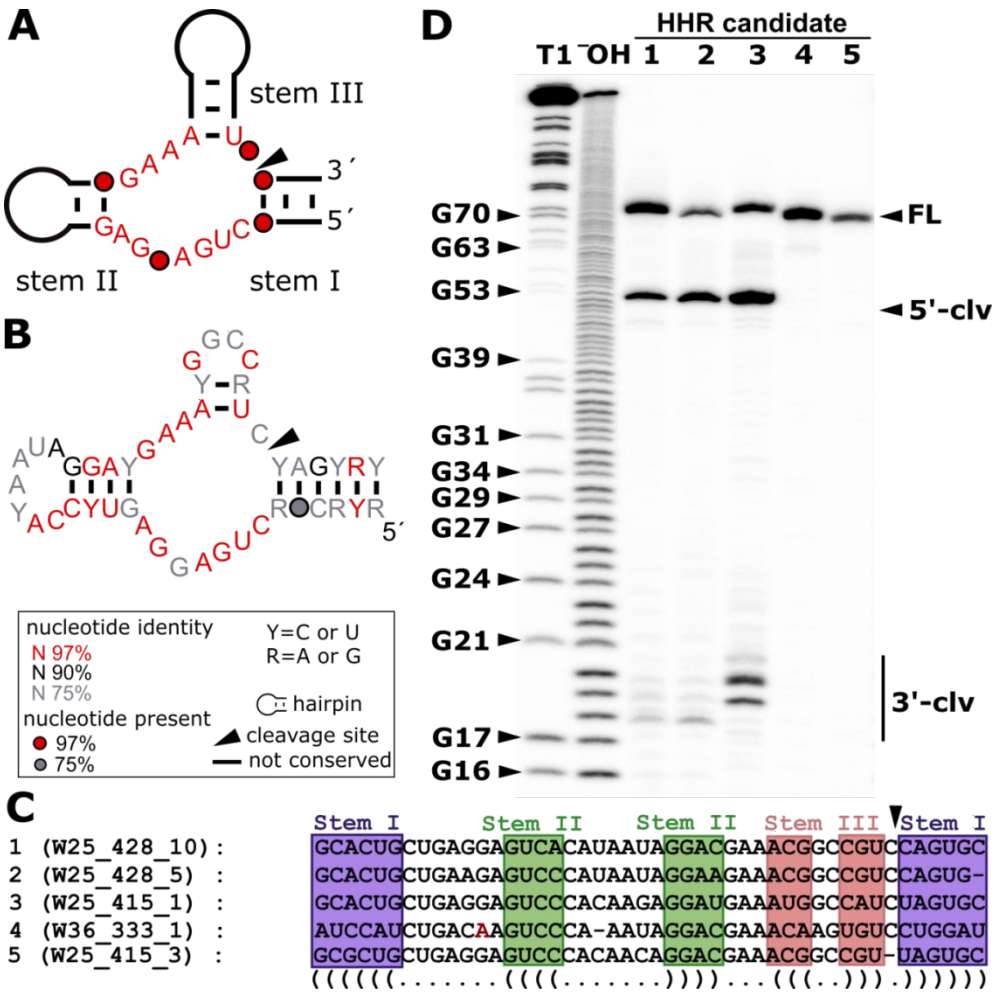


Figure 4: HHR candidates as part of WEs were selected for functional analysis. **A**, Consensus sequence and secondary structure of the HHR type I catalytic core. The stems are labeled I-III, and highly conserved nucleotides are shown in red. **B**, Consensus sequence and secondary structure of HHR candidates found as part of WEs. Only those HHR candidates that contain all conserved sequence and structural features were included into this

consensus (Supplementary Table 4 and Supplementary Figure 13). R2R was used to draw the model (Weinberg et al. 2011). **C**, Alignment of HHR sequences from selected candidates tested for self-cleavage activity *in vitro*. The stems are highlighted, and mutations in highly conserved regions shown in red (i.e. Candidate 4-W36_333). At the cleavage site (arrow), candidate 5 W25_415_3 has a deletion and served as control. **D**, Co-transcriptional cleavage analysis of selected HHR candidates. Full-length transcript (FL), 5'- (5'-clv), and 3'-cleavage (3'-clv) fragments are indicated by arrows. "T1" indicates partial digestion after G nucleotides by RNase T1 and "$^{-}$OH" partial alkaline hydrolysis. Samples were separated by 20% PAGE.

WEFs Vary Among Different Schistosome Species

Previous studies discussed the existence of WEFs and mobile genetic elements such as retrotransposons in different schistosome species. In their original article on W1 elements in *S. mansoni*, now WEF W1.2, Webster et al. (1989) reported the absence of W1 from *S. matthei, S. haematobium, S. japonicum, S. douthitti*, and *Fasciola hepatica* based on Southern blot results in the pre-genomic era. Furthermore, the SMalpha family (represented by W36.2), based on PCR data originally assumed by Ferbeyre et al. (2000) to be present in *S. haematobium* and *S. douthitti* but absent from *S. japonicum*, was later found in this species in high copy numbers and with the HHR core motif (Laha et al. 2000). Based on genome data for schistosomes (Howe et al. 2017), we searched for WEFs in *S. rodhaini*, which belongs to the same clade as *S. mansoni*, and in *S. haematobium* and *S. japonicum*, which represent different clades, respectively (Lawton et al. 2011).

BLASTn analysis showed the presence of the majority of WEFs in the other species. Compared to *S. mansoni*, however, lengths and structures of the elements varied. Most of the WEFs found in the other species revealed reduced sizes (Supplementary Table 5). One example is W11.2_ZW, which in *S. mansoni* has a length of 1,294 bp, whereas in *S. japonicum* its size is 251 bp, in *S. haematobium* 375 bp, and in *S. rodhaini* 275 bp, respectively. Partial versions of five WEF, W2.2_709, W4.2, and W7.2, occurred in all investigated species except *S. japonicum*. In *S. rodhaini*, a nearly complete version of WEF W31.2 exists, which is absent from *S. japonicum* and *S. haematobium*. Remarkably, full-length and partial versions of WEF W8.2 and W10.2 exclusively occurred in *S. mansoni*. Only for WEF W36.2, size differences of the repeat units were lower among the different species. For W36.2_W005 and W36_W001, we found slightly larger unit sizes in *S. rodhaini* (347 bp and 337 bp, respectively) compared to *S. mansoni* (332 bp and 335 bp, respectively) (Supplementary Table 5).

To investigate whether the size variations structurally affected regions with hypothesized regulatory functions, we performed alignment analyses using the best BLASTn hits focusing on selected WEFs that contain regions potentially coding for regulatory RNAs, W1.2 (mir-279, mir-2587), W5.2 (mir-598, sR36), W11.2_ZW (mir-232; DdR16), and W22.2 (SCARNA7, sR11, snoZ178). This analysis indicated a mosaic pattern with respect to the presence and integrity of these parts of the WEF sequences that potentially code for regulatory RNAs. Whereas the part of W1.2 coding for the miRNAs mir-279 and

mir-2587 was completely present in all investigated species, other *S. mansoni* miRNAs or snoRNAs were either absent or only partially preserved, or the W1.2 sequence was completely present in one or two further species (Figure 5).

Figure 5. Structural Comparison of Selected WEFs in Different Schistosome Species

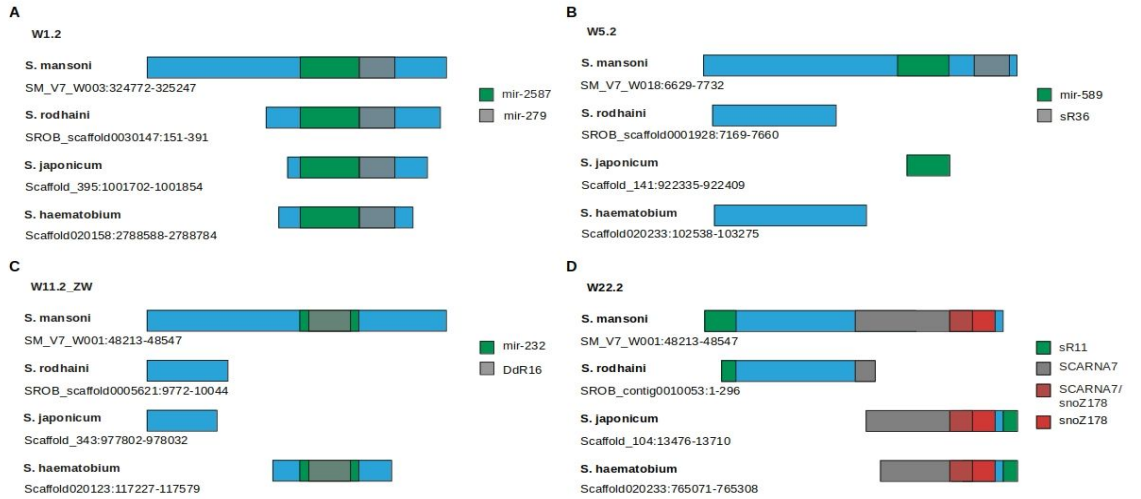


Figure 5: Diagram of structural differences of selected WEF in *S. mansoni*, *S. rodhaini*, *S. japonicum*, and *S. haematobium* that potentially encode regulatory RNAs. **A**, W1.2 (blue) harbours coding sequences for miRNAs mir-2587 (green) and mir-279 (grey). Compared to *S. mansoni*, best W1.2 hits upon BLASTn analysis showed shorter orthologs in *S. rodhaini*, *S. haematobium*, and *S. japonicum*. Although shorter, these orthologous sequences preserved full-length mir-279 and mir-2587. **B**, W5.2 (blue) harbours coding sequences for mir-598 (green) and snoRNA sR36 (grey). We found none of these regulatory RNAs in orthologous sequences of *S. rodhaini* and *S. haematobium*. In *S. japonicum*, a partial mir-2587 sequence occurs, however, not flanked by W5.2 sequence areas. In this case, we cannot exclude a false positive hit. **C**, W11.2_ZW harbours overlapping coding sequences for mir-232 (green) and DdR16 (grey), which is part of the mir-232 sequence. This overlapping RNA-coding part appears to be completely preserved in *S. haematobium*, although the flanking W11.2_ZW sequence areas are shorter compared to *S. mansoni*. In *S. rodhaini* and *S. japonicum*, we detected smaller W11.2_ZW elements without RNA-coding parts. **D**, W22.2 harbours sR11 (green) and overlapping coding sequences for SCARNA7 (grey) and snoZ178 (red), which is part of the SCARNA7 sequence (red/grey area). Only shorter W22.2 elements occur in the other schistosome species. In *S. rodhaini*, a small W22.2 variant exists containing fragments of the sR11 and SCARNA7 sequence. In *S. japonicum*, a big part of the overlapping RNA-coding region occurs with a small part of sR11 and a shortened SCARNA7 part, but a completely preserved snoZ178 part. We made a similar finding for *S. haematobium*, here the SCARNA7 (grey) part was slightly shorter compared to the *S. japonicum* counterpart. The sequences were obtained from data deposited on WormBase ParaSite (version 15, October 2020; https://parasite.wormbase.org), BioProject numbers: PRJEA36577 (*S. mansoni*), *S. rodhaini* (PRJEB526 -

Republic of Burundi), PRJNA520774 (HuSjv2, *S. japonicum)*, and PRJNA78265 (*S. haematobium*). Scaffold numbers are given.

For the HHR-containing WEF, species comparison showed a different picture. Looking again for best hits, we detected orthologs of three of four WEFs (W25.2, W27.2, W36.2) in all examined schistosome species, with *S. mansoni* showing the longest size variants. In nearly all cases, the integrity of the HHR motifs was fully preserved. The exception was W26.2, of which only a small part without HHR motif appeared to be conserved in *S. japonicum* (Figure 6). When we searched for the HHR sequence of W26.2 in the *S. japonicum* genome by BLASTn, we found no hit (data not shown).

Figure 6. Structural Comparison of Hammerhead Ribozyme-Containing WEFs in Different Schistosome Species

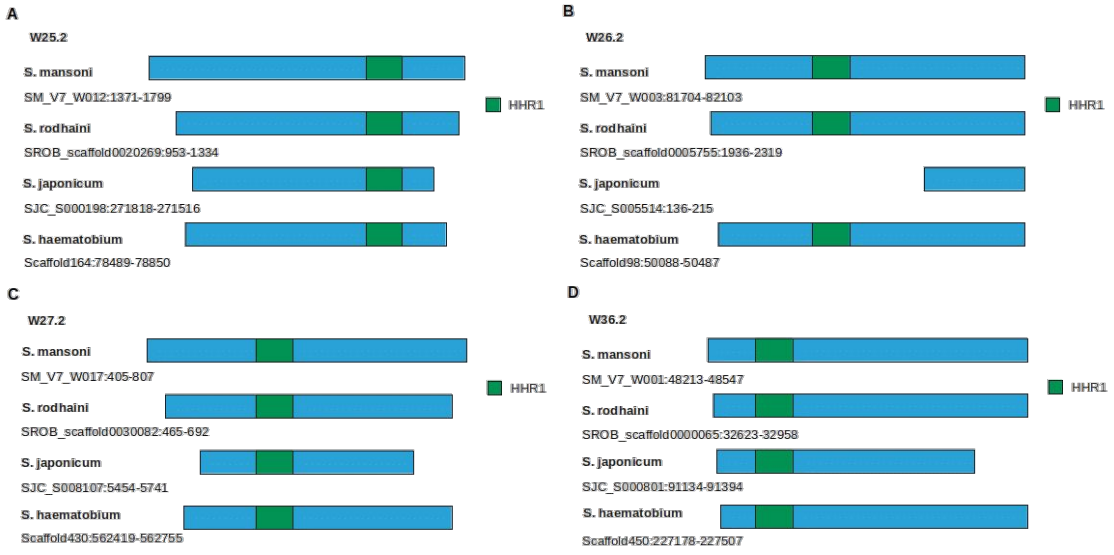


Figure 6: Diagram of structural differences of WEF W25.2, W26.2, W27.2, W36.2 in *S. mansoni*, *S. rodhaini*, *S. japonicum*, and *S. haematobium* that potentially encode type 1 HHR1 (scaffold information is given). The ribozyme parts of these sequences are highlighted in green. Compared to *S. mansoni*, BLASTn analysis showed shorter orthologs for all four WEFs in *S. rodhaini*, *S. haematobium*, and *S. japonicum*. Except for W26.2 in *S. japonicum*, the ribozyme motif was highly conserved among the selected species. The sequences were obtained from data deposited on WormBase ParaSite (version 15, October 2020; https://parasite.wormbase.org), BioProject numbers: PRJEA36577 (*S. mansoni*), *S. rodhaini* (PRJEB526 - Republic of Burundi), PRJNA520774 (HuSjv2, *S. japonicum)*, and PRJNA78265 (*S. haematobium*). Scaffold numbers are given.

Discussion

The days of “junk DNA“ are over. When the senior authors of this article studied genetics at their respective universities, the common doctrine was that the non-protein coding part of eukaryotic genomes consists of interspersed, “useless” sequences, often organized in repetitive elements such as satDNA. The latter might have accumulated during evolution, e.g. as a consequence of gene duplication events to separate and individualise gene function (Britten and Kohne 1968; Comings 1972;

Ohno 1999). This view has fundamentally changed (Biscotti et al. 2015a), and our study is the first one addressing this issue with structural, functional, and evolutionary aspects for the genome of a multicellular parasite.

Using V7 of the *S. mansoni* genome we re-analysed WEFs, which previous studies described as organized in 36 families of repetitive, non-protein coding DNA elements occurring on the W-chromosome (Lepesant et al. 2012b). From our results, we deduce 19 WEFs based on the fact that many of the original 36 WEFs are fused together. These 19 WEFs exhibit a surprising diversity of features. As expected for repetitive DNA, there is high sequence similarity but also variation among the repeat units of each single WEF. Besides few partial WE sequences of some WEFs on the WZ chromosomes, we found for 15 out of 19 WEF copies also on autosomes. Here, copy numbers considerably varied as well as the lengths of WE units of the respective WEFs. The autosomal occurrence possibly indicates the potential of element mobility. Indeed, in some WEF sequences we found residues of features typical for mobile genetic elements such as retrotransposons or retroviruses. Although we found no evidence of complete ORFs coding for a transposase, in an exemplary case, we detected direct repeats and TSD, which are often, but not always, found at integration sites of mobile genetic elements (Jensen et al. 1994; Han et al. 2014). Repetitive elements such as satDNA have been discussed to originate from transposable elements (TEs). Both share sequence similarity and organization patterns, which suggests a mutual relationship between satDNA and TEs. This probably influenced satDNA evolution and its roles on genome architecture and function (Meštrović et al. 2015; Satović et al. 2016). The mechanism of satDNA formation from TEs is unclear yet. However, structural elements such as palindromes, direct/inverted repeats, and the ability of stem-loop formation may be involved as well as illegitimate recombination events and deletions based on double-strand breaks and excision. These factors might drive rearrangements of TEs and the production of sequence parts as templates for further amplification to form tandemly organized satDNA clusters. This may apply to at least some of the WEs described here (W2.2, W4.2, W5.2, W11.2, and W16.2), which showed partial sequence homology to known mobile genetic elements. SatDNA monomers might be recognized by transposase-driven mechanisms acting on short DNA sequence motifs of these DNA sequences, which are recognized by enzymes related to transposases (Meštrović et al. 2015). As alternative to transposon-like cut and paste mechanisms, reintegration of repeat elements may happen via circular DNA intermediates generated from tandem repeats, as shown in *Arabidopsis* (Woodhouse et al. 2010). Whether this or similar mechanisms may apply to the WEF of *S. mansoni* has still to be elucidated.

Based on evidence for transcriptional activity and dynamic occurrence, recent years have unravelled novel roles for satDNA (Biscotti et al. 2015a). Among these is the establishment of heterochromatic states at centromeres and telomeres, which is indispensable for preserving chromosome integrity and

genome stability. In *S. mansoni*, Lepesant et al. (2012a) reported WEF-associated transcript occurrence and chromatin status of repeats to be stage-specific. Open chromatin on the W-chromosome occurred in larval stages but not in the sexually dimorphic adult stage. Furthermore, the euchromatic character of histone modifications around the WE repeats on W decreased during the life cycle. When transcribed, repeat RNA appeared stage-specifically in the larval stages miracidium and cercaria. These authors concluded that WEFs may play roles in structural changes of the chromatin and sex-chromosome emergence in ZW systems (Lepesant et al. 2012a). Furthermore, heterochromatisation was discussed as major factor for the differentiation of sex chromosomes in the schistosomatids and also for schistosome speciation (Lawton et al. 2011).

Here, we provide conclusive evidence for WEF expression throughout schistosome development, from the miracidium to the adult stages. Copy numbers considerably varied among the WEFs and depended on the life-cycle stage, sex, pairing, and strain. In a Liberian strain of *S. mansoni*, genome instability of the W1 element was hypothesized from the finding of gain or loss of W1 elements in the progeny of crosses, or even during larval development within the snail host (Grevelding 1999). This pointed to recombination processes during meiosis and/or mitosis. Our new results suggest that a mobile character of WEF may account for this genome instability, which allows WEs to recombine within a single generation. This could also explain the observed strain differences. Accordingly, some of the results shown (Figure 2 and Supplementary figs. 6-9), may represent snapshots of the WEF setup within the samples at the time they were available for analysis. Therefore, WE occurrence and copy numbers may be a matter of change over time and generations, a view that is supported by the observed variations among biological and technical replicates. As practical consequence, sex determination of clonal cercarial populations based on PCR using WE-specific primers, as previously performed with varying success (Gasser et al. 1991; Boissier et al. 2001; Chevalier et al. 2016), may eventually lead to inconsistent results.

Our approach to assign putative roles to WEFs led to the StructRNAFinder-based prediction of different classes of regulatory RNAs as parts of WE sequences. Due to their roles in gene regulation, miRNAs have attracted attention (Bartel 2018). In mammals, X-chromosomal miRNA expanded by gene duplication after the emergence of sex chromosomes, and hundreds of different miRNAs have been identified shaping mRNA expression (Meunier et al. 2013). Many are conserved, which may also applies to *S. mansoni* (http://www.mirbase.org/; Kozomara et al. 2019). As example, mir-181, in our study predicted as putative part of W8.2, may represent a miRNA family for which multiple roles in immune cell development, hematopoiesis, cell death pathways, cancer, and drug resistance have been reported in humans (Weng et al. 2015; Braicu et al. 2019). Species comparison indicated W8.2 in *S. mansoni* but not in the closely related species *S. rodhaini* and *S. japonicum* or *S. haematobium*. In *Drosophila*, mir-279, potential part of W1.2, influence neuron formation of the olfactory sensory

system (Hartl et al. 2011) and eye patterning interfering with EGFR pathways (Duan et al. 2018). Roles for miRNAs have also been reported for parasites and host-parasite interaction (Zhao and Guo 2019; Britton et al. 2020). Plant-parasitic nematodes induce feeding site formation in host cells, which differentially express miRNAs upon infection (Jaubert-Possamai et al. 2019). In schistosomes, the presence and conservation of genes involved in miRNA pathways and their role in *B. glabrata/S. mansoni* interaction has been discussed (Queiroz et al. 2017; Cardoso et al. 2020). Furthermore, miRNAs of *S. mansoni* and *S. japonicum* may participate in male-female interaction, sexual development, and pathological processes in the final host (Zhu et al. 2014; Cai et al. 2016; Yu et al. 2019). In *S. haematobium*, a genome-wide analysis of small ncRNAs identified mir-785 expression in both sexes, which corresponds to our findings. Furthermore, homology-based prediction indicated a voltage-dependent anion-selective channel protein (MS3_05034) as potential target (Stroehlein et al. 2018). Transcripts of the MS3_05034 ortholog of *S. mansoni*, Smp_091240, occur in males, females and their gonads (Lu et al. 2016). *S. mansoni* mir-785 is also expressed in adults although with a sex bias (females > males) including the gonads (ovary >> testes), and influenced by pairing (paired females >> unpaired females).

Also, snoRNAs are widely distributed among eukaryotes and participate in the modification and processing of ribosomal and small nuclear RNAs, splicing, rRNA acetylation, mRNA abundance and translational efficiency (Lui and Lowe 2013; Bratkovič et al. 2020). SNORD59, a putative snoRNA in W2.2, is a member of the C/D family directing site-specific 2'-O-methylation of substrate RNA such as 18S rRNA (Galardi et al. 2002). The conserved C/D family-specific C-box (UGAUGA) and D-box (CUGA) motifs are present in W2.2-encoded SNORD59. Perfect matches occur in the 709 bp variants of W2.2, whereas in the 711 bp variants a G/C mutation has been found at position 3 of the D-box motif (Supplementary Figure 12). The 2'-O-methylation of the ribose moiety is important for the maturation of almost all classes of RNAs and involves different snoRNP (ribonucleoprotein) complexes (Kiss-László et al. 1996; Ojha et al. 2020). They contain nucleolar proteins (Nop), of which orthologs exist in *S. mansoni* such as Nop 56 (Smp_053470 and Smp_048660; https://parasite.wormbase.org). Smp_053470 and Smp_048660 are most abundantly expressed in sporocysts and ovaries of paired females (Lu et al. 2016; Lu et al. 2018; http://schisto.xyz/, V7), which coincides with SNORD59 transcripts in these stages/tissues (Supplementary Table 3). SnoRNAs can be further processed into smaller RNAs with different functionality including miRNAs (Scott and Ono 2011). In mammals, protein-coding genes exist that express both snoRNAs and miRNAs in single introns. The existence of eukaryotic and archean members suggested that snoRNAs - in evolutionary terms - are more ancient compared to miRNA (Dennis and Omer 2005). However, there is also evidence for recently evolved snoRNA and miRNAs (Niwa and Slack 2007). This suggests that both RNA classes evolve dynamically and at fast rates, which may also apply to the predicted *S. mansoni* snoRNA and miRNAs. Furthermore, many of

the most recently evolved snoRNA and miRNAs may be derived from TEs. Indeed, previous studies showed TSDs at snoRT (human snoRNA-like retrogene) integration sites, which supports their mobile character (Weber 2006; Lui and Lowe 2013). Our indicative findings of TSDs and intra-WE inverted repeats flanking miRNA and snoRNA subunits of WEFs as well as their mobile character correspond to these concepts. To prove whether the predictions of snoRNAs and miRNAs as parts of appropriate WEFs have functional relevance will be subject of future studies. In this study, we focused on functional evidence of type I HHRs, which are parts of some WEF. HHRs are widely distributed in the animal and plant kingdoms, and they can be associated with repeated DNA (Perreault et al. 2011; Cervera et al. 2014; de la Peña et al. 2017; Lünse et al. 2017). Recent studies on the biology of genomic HHRs discuss their potential roles in the propagation of retrotransposons, which are major components of eukaryotic genomes including schistosomes (Venancio et al. 2010), and which contribute to genome evolution shaping developmental processes and eukaryotic complexity (Mita et al. 2016). Remarkably, our species analysis showed a high conservation of the ribozyme-like sequences in W25.2, W26.2_W003, W27.2, and W36.2_W001 among the examined species with the exception of the absence of the type 1 HHR in W26.2 in *S. japonicum*. To test whether WEF-encoded HHRs are catalytically active, we selected predicted ribozyme sequences of autosomal WEs. Alignments showed that some WEF-encoded HHR sequences harbour the canonical core sequences CUGANGA and GAAA. Other HHRs, such as candidates 4 (W36_333_1) and 5 (W25_415_3), have mutations or deletions in highly conserved nucleotides that likely lead to a reduction or loss of the cleavage activity (Ruffner et al. 1990). Indeed, all HHR candidates conform to previously described minimal HHRs with an extremely short stem III (Forster et al. 1987; Lünse et al. 2017). Although only weak similarity to SMalpha HHR existed, it may be tempting to speculate about their functional role as part of SINE-like elements similar to the ones already described (Ferbeyre et al. 1998). Furthermore, W25.2 showed higher expression in the Puerto Rican and Guadeloupean strains compared to the Liberian strain. Thus, it may contribute to strain-specific differences at the post-transcriptional level. Within the Liberian strain, W25.2 appeared to be expressed at a slightly higher level in the ovary of females compared to male testes or the adults. This suggests a gonad-associated function in females, which awaits confirmation in subsequent studies. These are needed to substantiate the expounded hypotheses and to assure that *S. mansoni* WEs elements are not just selfish DNA (Biscotti et al. 2015b, Thakur et al. 2021).

Creating variability and genome plasticity are hallmarks of parasites with different molecular principles invented during evolution (Lanzer et al. 1995). In schistosomes, repetitive WEs may represent one of these principles. Using the intermediate snail-host stage for asexual recombination (Grevelding 1999; Bayne and Grevelding 2003) and the final-host stage for sexual recombination, schistosomes have exceptional preconditions for rapid evolution driven by adaptation to new environments. Prerequisite for asexual recombination of WEFs from the heterochromatin area of the W-chromosome is a biphasic

chromatin stage, in which both euchromatic and heterochromatic states can be adopted. Indeed, this has been confirmed for *S. mansoni* by showing that euchromatic histone modifications around WEs dominate in the intermediate snail-host but decrease afterwards (Lepesant et al. 2012a). In *P. falciparum*, bistable chromatin has been discussed as a mechanism regulating variant gene expression patterns within clonal populations (Llorà-Batlle et al. 2019). In nematodes, mobile genetic elements contribute to genome plasticity in the absence of sexual reproduction (Castagnone-Sereno et al. 2014). This fits to the general view of the impact of mobile genetic elements on genome evolution and adaptation, which is important for organisms frequently facing new environments such as parasites (Schrader and Schmitz 2019), particularly those with complex life cycles. The evolution of repetitive DNA was also associated with reproductive isolation, founding new species, genome integrity, and karyotype evolution (Coghlan 2005; Biscotti et al. 2015a; Lower et al. 2018). Especially with respect to heterochromatic regions, remarkable diversity of karyotype patterns exist for different schistosome species leading to the definition of six clades correlating with the different geographical distribution as well as with the hypothesized Asian origin of schistosomes (Hirai et al. 2000; Lawton et al. 2011). Our comparison of WEFs in selected schistosome species of different clades provide a first hint for their potential contribution to karyotype variability, and thus speciation. This may include a varying repertoire of WEF-encoded regulatory RNAs, which differs between these species.

Repetitive DNA discloses high sequence and copy number variability among and within species but also in closely related organisms, which points to rapid evolution (Lower et al. 2018). This may involve coevolution with regulatory RNAs as it was hypothesized for satellite repeats and long noncoding RNAs (Lee et al. 2019) as well as the dual relationship between Alu repeats (short interspersed nuclear elements of the human genome) and miRNAs. Duplication events involving Alu elements have favoured the expansion of miRNA clusters and their expression (Lehnert et al. 2009).

From the data obtained in our study and against the background of recent literature, it is tempting to speculate that more of the WE “junk-DNA” than expected might be functional and relevant. WEs of all investigated WEFs exhibit a capricious incidence, , and they are transcribed in a stage-, sex-, pairing-, gonad, and strain-specific or -preferential pattern. From exemplary findings of features typical for the activity of mobile genetic elements, we hypothesize that W elements may have a mobile character. Together with previous findings of intra-clonal recombination events of WEs (Grevelding 1999; Bayne and Grevelding 2003), their presumptive role in sex chromosome emergence (Lepesant et al. 2012), their putative capacity to express regulatory RNAs, we propose that W elements might influence the biology of *S. mansoni*. Furthermore, based on the variable occurrence of WEFs in different schistosome strains, isolates, and even species, we hypothesize that the W elements represent one of the sources of heritable variability in the evolution of the family Schistosomatidae.

Materials and Methods

Ethics Approval

Experiments with hamsters to obtain *S. mansoni* material as basis for RNA-seq studies leading to the bioinformatics data analysis were done in accordance with the European Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes (ETS No 123; revised Appendix A) and had been approved by the Regional Council (Regierungspraesidium) Giessen (V54-19 c 20/15 c GI 18/10).

Mapping and Characterizing WEF (V5) Against Genome Version V7

WE repeat sequences, originally described as being organized in 36 families (Lepesant et al. 2012b), were identified in the current genome version V7 of *S. mansoni* by local alignment searches via BLASTn (BLAST+, v. 2.6.0) (Camacho et al. 2009) using a coverage cut-off of 66% and an identity cut-off of 80%. In addition, all WEFs were also detected with Gepard (GEnome PAir - Rapid Dotter, v. 1.40) through the representation of the repetitive patterns in dotplot graphics (Krumsiek et al. 2007). Dotplot analysis (Gibbs et al. 1970) was performed and visualized as two-dimensional matrices with sequences being compared along vertical and horizontal axes. In case of identity, individual cells within the matrix are shaded black, thus matching sequence segments appear as diagonal lines across the matrix. We also used Gepard to provide a graphical overview of existing patterns that are typical for mobile genetic elements such as transposons.

Mapping and Counting of RNA-Seq Reads

First, we used the tool Trim Galore (https://www.bioinformatics.babraham.ac.uk/index.html), a wrapper script to automate quality and adapter trimming as well as quality control, to remove the adapters required for Illumina sequencing from the RNA-Seq reads. In order to detect WE transcripts, RNA-Seq reads of different samples were aligned to a Multifasta file with all WE sequences using Bowtie2 (version 2.3.4.3; Reporting Option: all alignments) (Langmead et al. 2009; Langmead and Salzberg 2012). Using bedtools intersect (Quinlan and Hall 2010; version 2.27.1+galaxy1), RNA-Seq reads were screened for overlaps with WEF sequences to produce the raw read counts. To this end, we used the mapped reads of the Bowtie2 (bam file) analysis and a bed file containing length information of all WEF (Table 1).

Additionally, we included RNA-Seq reads of all protein-coding genes in *S. mansoni* to increase the library sizes for between-sample normalization. Such read counts were obtained as described before (Lu et al. 2018). Briefly, the quality of raw RNA-seq reads was assessed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and reads were aligned to *S. mansoni* V7 genome (WormBase Parasite WBPS14) using STAR 2.7.2a (Dobin et al. 2013) with the option --

alignIntronMin 10. Counts per gene were summarised with FeatureCounts v1.4.5-p1 (Liao et al. 2014) based on the exon feature, using the annotation from WormBase Parasite WBPS14 (https://parasite.wormbase.org/).

Open Reading Frame Analysis

We used the NCBI program "ORFfinder" (https://www.ncbi.nlm.nih.gov/orffinder/) to detect translatable sequence areas. For the analysis, we selected a minimum length of 30 nucleotides and only the start codon "ATG". The amino acid sequences were examined by BLASTp (BLAST+, v.2.6.0; Camacho et al. 2009) against UniProtKB/Swiss-Prot.

Differential Expression Analysis and Normalization

To determine quantitative changes in WE transcript levels between the RNA-Seq datasets we performed differential expression analysis using raw reads in DESeq2 (Love et al. 2014). Throughout the manuscript, expression is defined as occurrence or levels of transcripts. For normalization of the data sets, we combined WE-specific read counts and read counts of protein-coding genes (Smp identifiers). For generation of the heatmap of the count matrix and the sample-distance matrix, we used R (https://www.r-project.org, v3.6.3) package DESeq2 (v1.24.0) and pheatmap (v1.0.12), applying log2 transformation of normalized counts of WEs using the *normTransform* and *rlogTransformation* function, respectively.

Identification of WEs on Autosomes

We used Blast+ (v.2.6.0; Camacho et al. 2009) to identify highly similar sequences of WEs on autosomes, applying the megablast task and a 0.001 E-value cut-off. We only extracted sequences if the alignment covered more than 80% of the query sequence, and if the overall alignment percentage (OAP; percent identity of the alignment multiplied by the coverage divided by 100) was higher than 60. Exceptions of this rule were made in case fragmented autosomal W elements were small; fragment sizes below 25 bp were not considered. Using the sequence alignment tool MAFFT (v7.471; https://mafft.cbrc.jp/alignment/server/), we calculated sequence homologies (Kuraku et al. 2013; Katoh et al. 2019). As input, we used the Fasta sequences and chose "Adjust direction according to the first sequence". Clustal-formatted alignments were produced as results. To find patterns of WE mobile activity, such as potential transposition events, we used MITE Tracker, an open source program that provides positional information on inverted-repeat sequences and target site duplications (TSD) (Crescente et al. 2018).

Functional Prediction

For the prediction of sequences coding functional RNA, we used StructRNAfinder, an integrative tool allowing the identification, functional annotation, and taxonomic allocation of sequences to RNA families by secondary structure inference (https://structrnafinder.integrativebioinformatics.me/run.html) (Arias-Carrasco et al. 2018). StructRNAfinder displays sequence consensus alignments for RNA families, according to Rfam database (RNA families, data base version 14.2, Kalvari et al. 2018; https://rfam.xfam.org), but also provides a taxonomic overview for each assigned functional RNA. As input, we generated FASTA files with complete RNAs from the Bowtie2 output by using the tool "bedtools MergeBED" (Galaxy Version 2.27.1) (Quinlan and Hall 2010) to merge overlapping and adjacent regions. We chose the option cmsearch and an E-value of 0.01. In the output options, we selected „report only best hit per sequence“.

Determination of Ribozyme Activity During *in vitro* Transcription

We produced transcription templates containing the T7 promoter sequence by extension of partly complementary oligonucleotides (Supplementary Table 6) using the Phusion DNA Polymerase (ThermoScientific™). We purified templates by phenol/chloroform extraction and ethanol precipitation. For *in vitro* transcription, 1 µg purified template, 1x transcription buffer, 3% DMSO, 4 mM NTPs, 0.013 U thermostable inorganic pyrophosphatase (NEB), and 25 ng of T7 RNA polymerase (laboratory preparation) were combined in a 30 µl reaction and incubated at 37°C for 2 h. The 1x transcription buffer contained 80 mM HEPES-KOH (pH 7.5), 24 mM $MgCl_2$, 2 mM Spermidine, and 40 mM DTT. Labelling of HHR candidates during transcription occurred by the addition of [α^{32}P]-CTP. We stopped reactions by adding RNA loading-dye composed of 2.5 mM Tris-HCl (pH 7.6), 20% formamide, 0.06% bromophenol, and 0.06% xylene cyanol. A cleavage-deficient, elongated variant of W25_415_3 (154 nt) was gel-purified after *in vitro* transcription, dephosphorylated using the antarctic phosphatase (NEB**)** and labelled at its 5'-end using the T4 polynucleotidyl kinase (NEB) and [γ^{32}P]-ATP. Following another gel-purification, we used the 5'-labeled RNA to create a size standard by alkaline hydrolysis and RNase T1 digestion. For experimental validation and product visualisation, we separated the size standard and internally labelled ribozyme cleavage products by denaturing 20% polyacrylamide gel electrophoresis (PAGE) and detected the bands with an Amersham Typhoon Imager (GE Healthcare).

Data Availability

All sequencing data used for analysis in this study are available at ENA (http://www.ebi.ac.uk/ena). These data originated from published and yet unpublished studies covering adult and larval

schistosomes (Grevelding 1995; Protasio et al. 2012; Wang et al. 2013; Lu et al. 2016; Picard et al. 2016; Lu et al. 2017; https://www.ebi.ac.uk/ena). Sample types, strains, replicates, accession numbers and references are listed in Supplementary Table 1. Further sequence information, as indicated in the text, was obtained from WormBase ParaSite (version 15, October 2020; https://parasite.wormbase.org), BioProject numbers: PRJEA36577 (*S. mansoni*), *S. rodhaini* (PRJEB526 - Republic of Burundi), PRJNA520774 (HuSjv2, *S. japonicum)*, and PRJNA78265 (*S. haematobium*).

Author Contributions

M.S. and C.C. performed the bioinformatics analyses supported by advice and help from Z.L., J.B., and A.G. J.O. and C.E.W. planned and performed the biochemical experiments with ribozymes and analysed the data. C.G.G., C.G., and C.E.W. wrote the manuscript. C.G.G. and C.G. designed the study and provided funding (FUGI). C.E.W. provided funds (DFG) for the ribozyme part.

Acknowledgements

We thank all the members of the Grevelding laboratory, particularly Simone Häberlein, Thomas Quack and Oliver Weth for fruitful discussions. The work was supported by the Wellcome Trust FUGI grant [107475/Z/15/Z to C.G.G., C.G.] and by the German Research Foundation [DFG; LU1889/4-1 to C.E.W.]. C.E.W. is also supported by a fellowship of the Peter and Traudl Engelhorn Foundation. This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP [ANR-10-LABX-41 to C.G.]. With the support of LabEx CeMEB, an ANR “Investissements d’avenir” program [ANR-10-LABX-04-01 to C.G.] through the Environmental Epigenomics Platform. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Arias-Carrasco R, Vásquez-Morán Y, Nakaya HI, Maracaja-Coutinho V. 2018. StructRNAfinder: an automated pipeline and web server for RNA families prediction. BMC Bioinform. 19:55. DOI: 10.1186/s12859-018-2052-2

Bartel DP. 2018. Metazoan microRNAs. Cell 173:20–51.

Basch P. 1991. In: Basch P. Editor. *Schistosomes: development, reproduction and host relations.* New York, Oxford University Press, p. 1–248.

Bayne CJ, Grevelding CG. 2003. Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. J Parasitol. 89:1056–1060.

Berriman M, et al. 2009. The genome of the blood fluke *Schistosoma mansoni*. Nature 460:352–358.

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M. 2015a. Transcription of tandemly repetitive DNA: functional roles. Chromosome Res. 23:463–477.

Biscotti MA, Olmo E, Heslop-Harrison JSP. 2015b. Repetitive DNA in eukaryotic genomes. Chromosome Res. 23:415–420.

Boissier J, Durand P, Moné H. 2001. PCR effectiveness for sexing *Schistosoma mansoni* cercariae: application for sexing clonal cercarial populations. Mol Biochem Parasitol. 112:139–141.

Braicu C, et al. 2019. Altered expression of miR-181 affects cell fate and targets drug resistance-related mechanisms. Mol Aspects Med. 70:90–105.

Bratkovič T, Božič J, Rogelj B. 2020. Functional diversity of small nucleolar RNAs. Nucleic Acids Res. 48: 1627–1651.

Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161:529–540.

Britton C, Laing R, Devaney E. 2020. Small RNAs in parasitic nematodes - forms and functions. Parasitology 147:855–864.

Cai P, et al. 2016. Comprehensive transcriptome analysis of sex-biased expressed genes reveals discrete biological and physiological features of male and female *Schistosoma japonicum*. PLoS Negl Trop Dis. 10: e0004684. DOI: 10.1371/journal.pntd.0004684

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. DOI: 10.1186/1471-2105-10-421

Canny MD, et al. 2004. Fast cleavage kinetics of a natural hammerhead ribozyme. J Am Chem Soc. 126:1084810849.

Cardoso TCdS, et al. 2020. Computational prediction and characterisation of miRNAs and their pathway genes in human schistosomiasis caused by *Schistosoma haematobium*. Mem Inst Oswaldo Cruz 115:e190378.

Castagnone-Sereno P, Danchin EGJ. 2014. Parasitic success without sex – the nematode experience. J Evol Biol. 27:1323–1333.

Cervera A, La Peña M de. 2014. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. Mol Biol Evol. 31:2941–2947.

Chamberlin M, Ring J. 1973. Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template specificity of the enzyme. J Biol Chem. 248:2235–2244.

Charlesworth B. 1991. The evolution of sex chromosomes. Science 251(4997):1030-1033.

Cheever AW, Macedonia JG, Mosimann JE, Cheever EA. 1994. Kinetics of egg production and egg excretion by *Schistosoma mansoni* and *S. japonicum* in mice infected with a single pair of worms. Am J Trop Med Hyg. 50:281–295.

Chevalier FD, Le Clec'h W, Alves de Mattos AC, LoVerde PT, Anderson TJC. 2016. Real-time PCR for sexing *Schistosoma mansoni* cercariae. Mol Biochem Parasitol. 205:35–38.

Clough ER. 1981. Morphology and reproductive organs and oogenesis in bisexual and unisexual transplants of mature *Schistosoma mansoni* females. J Parasitol. 67:535–539.

Coghlan A. 2005. *Nematode genome evolution*. WormBook: 1–15.

Colley DG, Bustinduy AL, Secor WE, King CH. 2014. Human schistosomiasis. Lancet 383:2253–2264.

Comings DE. 1972. The structure and function of chromatin. Adv Hum Genet. 3:237–431.

Copeland CS, et al. 2003. Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. J Virol. 77:6153–6166.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinformatics 19(1):348. DOI: 10.1186/s12859-018-2376-y

de la Peña M and Garcıa-Robles I. 2010. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. RNA 16:1943–1950.

de la Peña M, García-Robles I, Cervera A. 2017. The hammerhead ribozyme: a long history for a short RNA. Molecules 22(1):78.

DeMarco R, et al. 2004. Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. J Virol. 78:2967–2978.

DeMarco R, Machado AA, Bisson-Filho AW, Verjovski-Almeida S. 2005. Identification of 18 new transcribed retrotransposons in *Schistosoma mansoni*. Biochem Biophys Res Commun. 333:230–240.

Den Hollander JE, Erasmus DA. 1985. *Schistosoma mansoni*: male stimulation and DNA synthesis by the female. Parasitology 91:449–457.

Dennis PP, Omer A. 2005. Small non-coding RNAs in Archaea. Curr Opin Microbiol. 8:685–694.

Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Drew AC, Brindley PJ. 1995. Female-specific sequences isolated from *Schistosoma mansoni* by representational difference analysis. Mol Biochem Parasitol. 71:173–181.

Duan H, et al. 2018. The mir-279/996 cluster represses receptor tyrosine kinase signaling to determine cell fates in the *Drosophila* eye. Development 145(7):dev159053.

Erasmus DA. 1973. A comparative study of the reproductive system of mature, immature and "unisexual" female *Schistosoma mansoni*. Parasitology 67:165–183.

Ferbeyre G, Smith JM, Cedergren R. 1998. Schistosome satellite DNA encodes active hammerhead ribozymes. Mol Cell Biol. 18(7):3880-3888.

Fitzpatrick JM, et al. 2008. Use of genomic DNA as an indirect reference for identifying gender-associated transcripts in morphologically identical, but chromosomally distinct, *Schistosoma mansoni* cercariae. PLoS Negl Trop Dis. 2:e323. DOI: 10.1371/journal.pntd.0000323

Forster AC, Symons RH. 1987. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. Cell 49 211–220.

Galardi S, et al. 2002. Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA in vitro. Mol Cell Biol. 22:6663–6668.

Gasser RB, Morahan G, Mitchell GF. 1991. Sexing single larval stages of *Schistosoma mansoni* by polymerase chain reaction. Mol Biochem Parasitol. 47:255–258.

Gibbs AJ, McIntyre GA. 1970. The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. Eur J Biochem. 16(1):1–11.

Grevelding CG. 1995. The female-specific W1 sequence of the Puerto Rican strain of *Schistosoma mansoni* occurs in both genders of a Liberian strain. Mol Biochem Parasitol. 71:269–272.

Grevelding CG. 1999. Genomic instability in *Schistosoma mansoni*. Mol Biochem Parasitol. 101:207–216.

Grevelding CG. 2004. *Schistosoma*. Curr Biol. 14:R545.

Grevelding CG, Langner S, Dissous C. 2018. Kinases: molecular stage directors for schistosome development and differentiation. Trends Parasitol. 34:246–260.

Hahnel S, et al. 2018. Tissue-specific transcriptome analyses provide new insights into GPCR signalling in adult *Schistosoma mansoni*. PLoS Pathog. 14:e1006718. DOI: 10.1371/journal.ppat.1006718

Hammann C, Luptak A, Perreault J, La Peña M de. 2012. The ubiquitous hammerhead ribozyme. RNA 18:871–885.

Han M-J, Xu H-E, Zhang H-H, Feschotte C, Zhang Z. 2014. Spy: a new group of eukaryotic DNA transposons without target site duplications. Genome Biol Evol. 6:1748–1757.

Hartl M, et al. 2011. A new Prospero and microRNA-279 pathway restricts CO2 receptor neuron formation. J Neurosci. 31:15660–15673.

Hirai H, Spotila LD, LoVerde PT. 1989. *Schistosoma mansoni*: chromosomal localization of DNA repeat elements by in situ hybridization using biotinylated DNA probes. Exp Parasitol. 69:175–188.

Hirai H, et al. 2000. Chromosomal differentiation of the *Schistosoma japonicum* complex. Int J Parasitol. 30:441–452.

Hombach S, Kretz M. 2016. Non-coding RNAs: classification, biology and functioning. Adv Exp Med Biol. 937:3–17.

Hotez PJ, Kamath A. 2009. Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. PLoS Negl Trop Dis. 3:e412. DOI: 10.1371/journal.pntd.0000412

Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. WormBase ParaSite - a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 215:2–10.

Jacinto DS, et al. 2011. Curupira-1 and Curupira-2, two novel Mutator-like DNA transposons from the genomes of human parasites *Schistosoma mansoni* and *Schistosoma japonicum*. Parasitology 138:1124–1133.

Jaubert-Possamai S, Noureddine Y, Favery B. 2019. MicroRNAs, new players in the plant-nematode interaction. Front Plant Sci. 10:1180.

Jensen S, Gassama MP, Heidmann T. 1994. Retrotransposition of the *Drosophila* LINE I element can induce deletion in the target DNA: a simple model also accounting for the variability of the normally observed target site duplications. Biochem Biophys Res Commun. 202:111–119.

Jimenez RM, Delwart E, Lupták A. 2011. Structure-based search reveals hammerhead ribozymes in the human microbiome. J Biol Chem. 286:7737–7743.

Jimenez RM, Polanco JA, Lupták A. 2015. Chemistry and biology of self-cleaving ribozymes. Trends Biochem Sci. 40:648–661.

Jones JT, Kusel JR. 1989. Intra-specific variation in *Schistosoma mansoni*. Parasitol Today (Regul Ed) 5:37–39.

Jourdane J, Theron A. 1980. *Schistosoma mansoni*: cloning by microsurgical transplantation of sporocysts. Exp Parasitol. 50:349–357.

Jourdane J, Theron A, Combes C. 1980. Demonstration of several sporocysts generations as a normal pattern of reproduction of *Schistosoma mansoni*. Acta Trop. 37:177–182.

Kalvari I, et al. 2018. Non-coding rna analysis using the rfam database. Curr Protoc Bioinformatics 62:e51.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 20:1160–1166.

Khvorova A, Lescoute A, Westhof E, Jayasena SD. 2003. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol. 10:708–712.

Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell 85:1077–1088.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. Nucleic Acids Res. 47:D155-D162.

Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23:1026–1028.

Kunz W. 2001. Schistosome male-female interaction: induction of germ-cell differentiation. Trends Parasitol. 17:227–231.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41:W22-8.

La Peña M de, García-Robles I. 2010. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. RNA 16:1943–1950.

La Peña M de, García-Robles I, Cervera A. 2017. The hammerhead ribozyme: a long history for a short RNA. Molecules 22.

Laha T, McManus DP, Loukas A, Brindley PJ. 2000. Sjalpha elements, short interspersed element-like retroposons bearing a hammerhead ribozyme motif from the genome of the oriental blood fluke *Schistosoma japonicum*. Biochim Biophys Acta. 1492:477–482.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

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.

Lanzer M, Fischer K, Le Blancq SM. 1995. Parasitism and chromosome dynamics in protozoan parasites: is there a connection? Mol Biochem Parasitol. 70(1-2):1-8.

Lawton SP, Hirai H, Ironside JE, Johnston DA, Rollinson D. 2011. Genomes and geography: genomic insights into the evolution and phylogeography of the genus *Schistosoma*. Parasit Vectors 4:131.

Lee H, Zhang Z, Krause HM. 2019. Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners? Trends Genet. 35:892–902.

Lehnert S, et al. 2009. Evidence for co-evolution between human microRNAs and Alu-repeats. PLoS ONE 4:e4456. DOI: 10.1371/journal.pone.0004456

Lepesant JMJ, et al. 2012a. Chromatin structural changes around satellite repeats on the female sex chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence. Genome Biol. 13:R14.

Lepesant JMJ, et al. 2012b. Combination of de novo assembly of massive sequencing reads with classical repeat prediction improves identification of repetitive sequences in *Schistosoma mansoni*. Exp Parasitol. 130:470–474.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

Lilley DMJ. 2019. Classification of the nucleolytic ribozymes based upon catalytic mechanism. F1000Res 8. DOI: 10.12688/f1000research.19324.1

Llorà-Batlle O, Tintó-Font E, Cortés A. 2019. Transcriptional variation in malaria parasites: why and how. Brief Funct Genomics 18:329–341.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550.

Lower SS, McGurk MP, Clark AG, Barbash DA. 2018. Satellite DNA evolution: old ideas, new approaches. Curr Opin Genet Dev. 49:70–78.

Lu Z, Spaenig S, Weth O, Grevelding CG. 2019. Males, the wrongly neglected partners of the biologically unprecedented male-female interaction of schistosomes. Front Genet. 10:796.

Lu Z, et al. 2016. Schistosome sex matters: a deep view into gonad-specific and pairing-dependent transcriptomes reveals a complex gender interplay. Sci Rep. 6:31150.

Lu Z, et al. 2017. A gene expression atlas of adult *Schistosoma mansoni* and their gonads. Sci Data 4:170118. DOI: 10.1038/sdata.2017.118

Lu Z, Zhang Y, Berriman M. 2018. A web portal for gene expression across all life stages of *Schistosoma mansoni*. bioRxiv:PPR13955. DOI: https://doi.org/10.1101/308213

Lui L, Lowe T. 2013. Small nucleolar RNAs and RNA-guided post-transcriptional modification. Essays Biochem. 54:53–77.

Lünse CE, Weinberg Z, Breaker RR. 2017. Numerous small hammerhead ribozyme variants associated with Penelope-like retrotransposons cleave RNA as dimers. RNA Biol. 14:1499–1507.

Marais G, Galtier N. 2003. Sex chromosomes: how X-Y recombination stops. Curr Biol. 13(16):R641–R643.

Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 126:309–320.

Mattick JS, Makunin IV. 2006. Non-coding RNA. Hum Mol Genet. 15:R17-29.

Meštrović N, et al. 2015. Structural and functional liaisons between transposable elements and satellite DNAs. Chromosome Res. 23:583–596.

Meunier J, et al. 2013. Birth and expression evolution of mammalian microRNA genes. Genome Res. 23:34–45.

Moore DV, Sandground JH. 1956. The relative egg producing capacity of *Schistosoma mansoni* and *Schistosoma japonicum*. Am J Trop Med Hyg. 5:831–840.

Moné H, Boissier J. 2004. Sexual biology of schistosomes. Adv Parasitol. 75:89-189.

Nawrocki EP, Eddy, SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933-2935.

Neves RH, et al. 2005. A new description of the reproductive system of *Schistosoma mansoni* (Trematoda: Schistosomatidae) analyzed by confocal laser scanning microscopy. Parasitol Res. 95:43–49.

Niwa R, Slack FJ. 2007. The evolution of animal microRNA function. Curr Opin Genet Dev. 17:145–150.

Ohno S. 1967. *Sex chromosomes and sex-linked genes*. Springer-Verlag, Berlin Heidelberg.

Ohno S. 1999. Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. Semin Cell Dev Biol. 10:517–522.

Ojha S, Malla S, Lyons SM. 2020. snoRNPs: functions in ribosome biogenesis. Biomolecules 10(5):783.

Olveda DU, et al. 2014. The chronic enteropathogenic disease schistosomiasis. Int J Infect Dis. 28:193–203.

Pélisson A, Sarot E, Payen-Groschêne G, Bucheton A. 2007. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. J Virol. 81:1951–1960.

Perreault J, et al. 2011. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. PLoS Comput Biol. 7:e1002031. DOI: 10.1371/journal.pcbi.1002031

Picard MAL, et al. 2016. Sex-biased transcriptome of *Schistosoma mansoni*: host-parasite interaction, genetic determinants and epigenetic regulators are associated with sexual differentiation. PLoS Negl Trop Dis. 10:e0004930. DOI: 10.1371/journal.pntd.0004930

Popiel I, Basch PF. 1984. Reproductive development of female *Schistosoma mansoni* (Digenea: Schistosomatidae) following bisexual pairing of worms and worm segments. J Exp Zool. 232:141–150.

Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G. 1986. Autolytic processing of dimeric plant virus satellite RNA. Science 231:1577–1580.

Protasio AV, et al. 2012. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 6:e1455. DOI: 10.1371/journal.pntd.0001455

Quack T, Doenhoff M, Kunz W, Grevelding CG. 1998. *Schistosoma mansoni*: the varying occurrence of repetitive elements in different strains shows sex-specific polymorphisms. Exp Parasitol. 89:222–227.

Queiroz FR, et al. 2017. Differential expression of small RNA pathway genes associated with the *Biomphalaria glabrata*/*Schistosoma mansoni* interaction. PLoS ONE 12:e0181483. DOI: 10.1371/journal.pone.0181483

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841842.

Ruffner DE, Stormo GD, Uhlenbeck OC. 1990. Sequence requirements of the hammerhead RNA self-cleavage reaction. Biochemistry 29:10695–10702.

Satović E, Vojvoda Zeljko T, Luchetti A, Mantovani B, Plohl M. 2016. Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. BMC Genomics 17:997. DOI: 10.1186/s12864-016-3347-1

Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. Mol Ecol. 28:1537–1549.

Scott MS, Ono M. 2011. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. Biochimie 93:1987–1992.

Seehafer C, Kalweit A, Steger G, Gräf S, Hammann C. 2011. From alpaca to zebrafish: hammerhead ribozymes wherever you look. RNA 17:21–26.

Shaw MK. 1987. *Schistosoma mansoni*: vitelline gland development in females from single sex infections. J Helminthol. 61:253–259.

Spotila LD, Hirai H, Rekosh DM, Lo Verde PT. 1989. A retroposon-like short repetitive DNA element in the genome of the human blood fluke, *Schistosoma mansoni*. Chromosoma 97:421–428.

Spotila LD, Rekosh DM, Boucher JM, LoVerde PT. 1987. A cloned DNA probe identifies the sex of *Schistosoma mansoni* cercariae. Mol Biochem Parasitol. 26:17–20.

Stroehlein AJ, et al. 2018. The small RNA complement of adult *Schistosoma haematobium*. PLoS Negl Trop Dis. 12:e0006535. DOI: 10.1371/journal.pntd.0006535

Thakur J, Packiaraj J, Henikoff S. 2021. Sequence, chromatin and evolution of satellite DNA. Int J Mol Sci. 22(9):4309. DOI: 10.3390/ijms22094309.

Uhlenbeck OC. 2003. Less isn't always more. *RNA* 9:1415–1417.

Venancio TM, Wilson RA, Verjovski-Almeida S, DeMarco R. 2010. Bursts of transposition from non-long terminal repeat retrotransposon families of the RTE clade in *Schistosoma mansoni*. Int J Parasitol. 40:743–749.

Wang B, Collins JJ, Newmark PA. 2013. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. Elife 2:e00768. DOI: 10.7554/eLife.00768

Weber MJ. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet. 2:e205. DOI: 10.1371/journal.pgen.0020205

Webster P, Mansour TE, Bieber D. 1989. Isolation of a female-specific, highly repeated *Schistosoma mansoni* DNA probe and its use in an assay of cercarial sex. Mol Biochem Parasitol. 36:217–222.

Weinberg CE, Weinberg Z, Hammann C. 2019. Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. Nucleic Acids Res. 47:9480–9494.

Weinberg Z, Breaker RR. 2011. R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinformatics 12:3. DOI: 10.1186/1471-2105-12-3

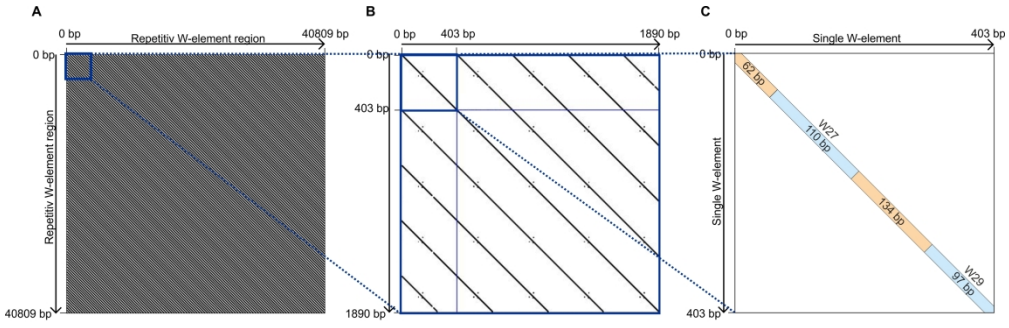
Weng H, Lal K, Yang FF, Chen J. 2015. The pathological role and prognostic impact of miR-181 in acute myeloid leukemia. Cancer Genet. 208:225–229.

Woodhouse MR, Pedersen B, Freeling M. 2010. Transposed genes in *Arabidopsis* are often associated with flanking repeats. PLoS Genet. 6(5):e1000949. DOI: 10.1371/journal.pgen.1000949

Yu J, Yu Y, Li Q, Chen M, Shen H, Zhang R, Song M, Hu W. 2019. Comprehensive analysis of miRNA profiles reveals the role of *Schistosoma japonicum* miRNAs at different developmental stages. Vet Res. 50:23.

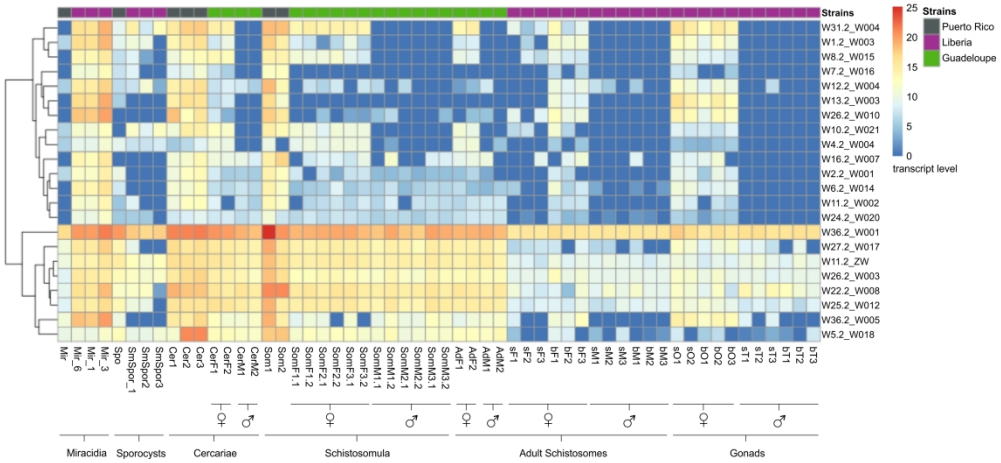
Zhao J-H, Guo H-S. 2019. Trans-kingdom RNA interactions drive the evolutionary arms race between hosts and pathogens. Curr Opin Genet Dev. 58–59:62–69.

Zhu L, Liu J, Cheng G. 2014. Role of microRNAs in schistosomes and schistosomiasis. Front Cell Infect Microbiol. 4:165. DOI: 10.3389/fcimb.2014.00165



The New Definition of WEF W27.2

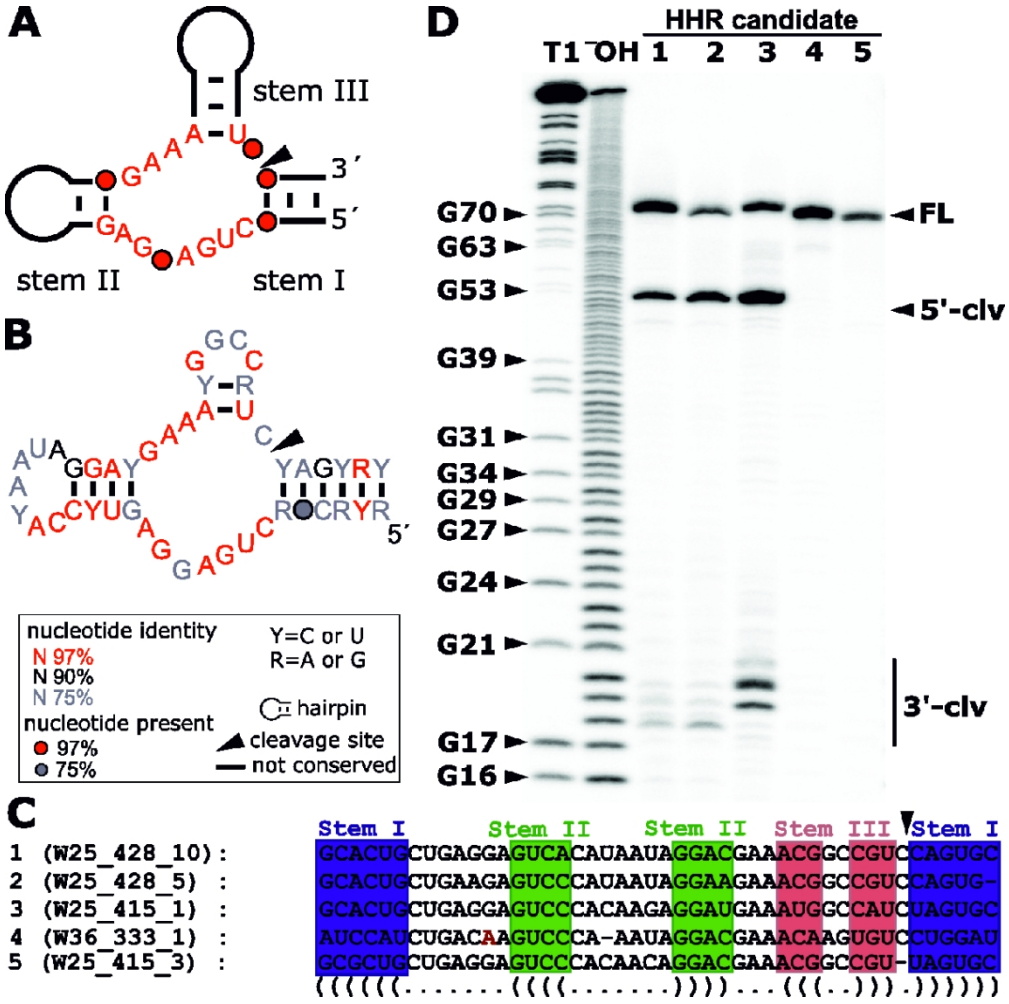
316x100mm (300 x 300 DPI)



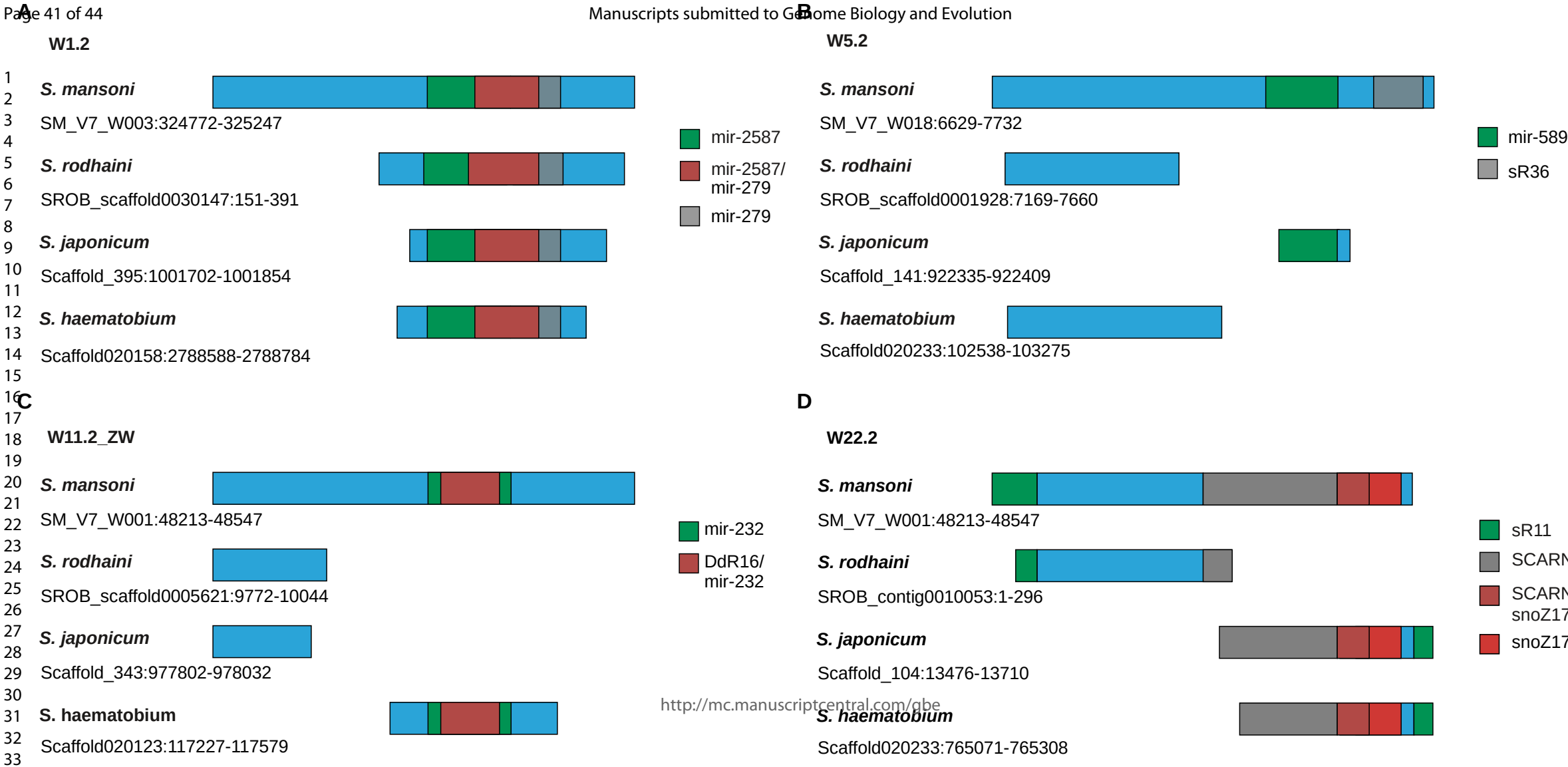
WE Transcript Profiles Across Different S. mansoni Strains, Life Stages, Sexes, and Gonads

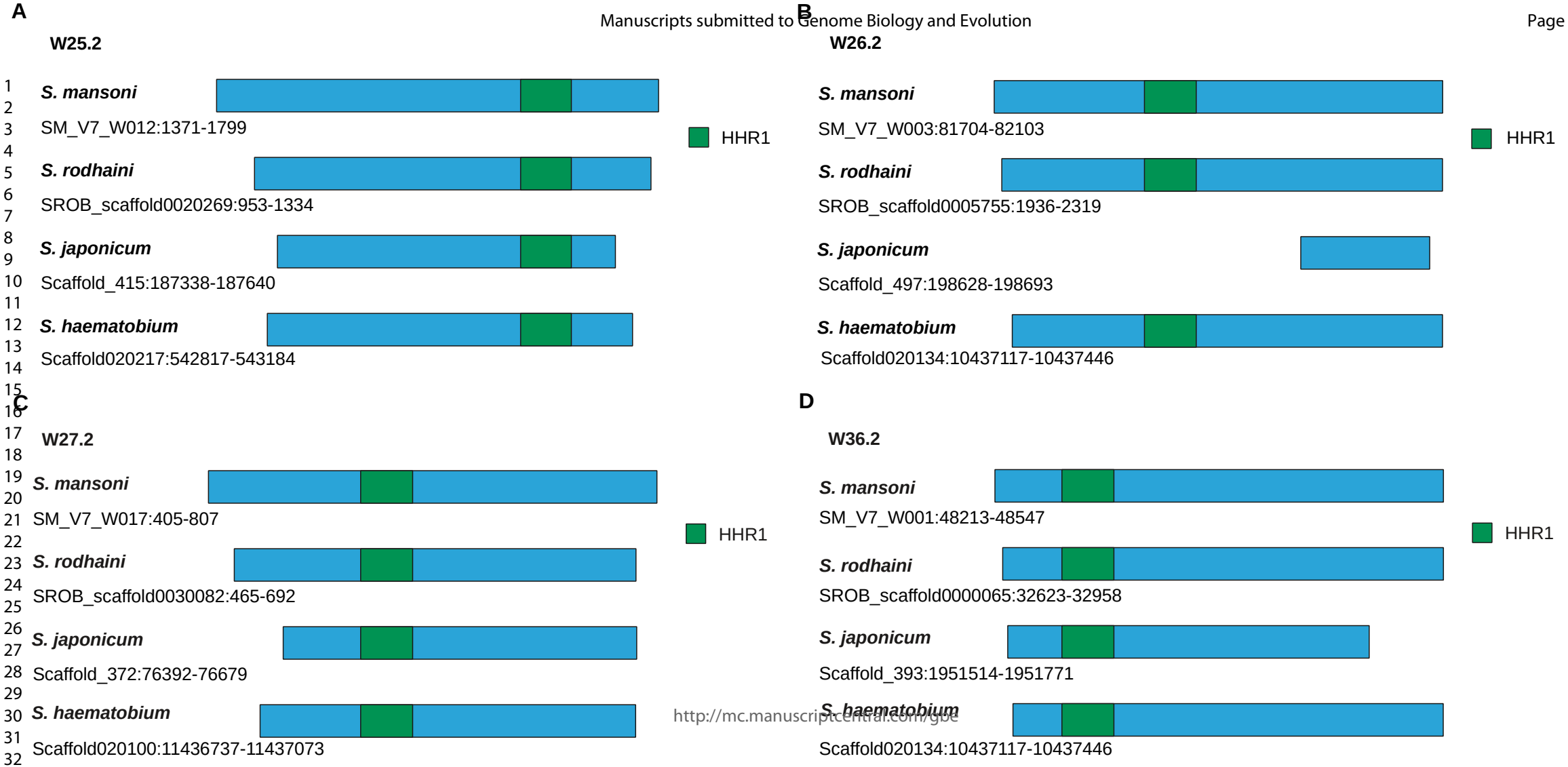
353x162mm (300 x 300 DPI)

```
Smalpha          atcgcacaagcaagtggct--------atcaggactcagtggccgagtggataacgcgat
W36.2_335bp      attgcacaagcaagtggct--------atcaggactcagtggtcgagtggataacgcgat
W25.2_415bp      gttggaaatagacataaacaccatgggatgccgtctcagtggtgtatgtgttgagtgctc
W26.2_400_bp     gttgaagatggaaatgaacacggttggatgccgactcagtgttcttgatgttctcagctt
W27.2_403bp      tttgaagtaaggcattgagagcgttggatggtgacccagtggtctagtgtttaagtgctc
                 *.* *   . . .* .          **   * *.***** .       *       .

Smalpha          ggtgtttgaagcgaggg-tactgggttggagtcccagagtggatatcaactctgagatgc
W36.2_335bp      ggcgtttgaagcgaaagttactgggttcgagtgccagaatggacatcaactctgagatgc
W25.2_415bp      tattgagaaagttgttattcctgtgttggaatctgg----------tgaggcgaaatcgt
W26.2_400_bp     gcgcacgaaaatgataagcgcagttttcgaatctgg----------agaggcggtatcgt
W27.2_403bp      acgcgcgagactgataggtgctgggttggaatctcg----------cgagactgcatcgt
                 ..* . .  . . * *  ** **.* . .          .*  * . . .*.

Smalpha          aggtacatccagctgacgagtccca-aataggacgaaacgcgcgtcctggattccactgc
W36.2_335bp      aggtacatccagctgacgagtccca-aataggacgaaacgcgcgtcctccattccactgc
W25.2_415bp      ggattggaaatgctgaggagtcgcacaatagaaccaaatgaccgtccagtgtttccagga
W26.2_400_bp     ggatgcacactgctgagaaagatattaatgaagttccaaggctgctcag-attgtcgagg
W27.2_403bp      ggatgcgcactgctgagcagtcccacagtacgacgaaacggccatccagtgcttccaggt
                 .*.*  .    *****  *.     *.*. ...   * *  ....*   ..* .   *

Smalpha          tatccactattcatc--------tttgcttatcatgcttgtgaatcaaggctatatcgag
W36.2_335bp      tagccactatccatc--------tttgcctaccatgcttgtcaatttaggctatatcgag
W25.2_415bp      tttccctagtggtctaccttcaatccactcatgatttcaaata------taaata-cact
W26.2_400_bp     ttaacatgacggtctctcaacaattgattcatcatcataacca------ttaacattacg
W27.2_403bp      tttccatcgtgctccagcttcaattcattcatcatctcaacta------tcaacattact
                 *   * . ..   ..       *. ....*. **  . .  *        *.* ..

Smalpha          gcaatacgcacagt-----------------atgcacata-tgccaattagagactgacc
W36.2_335bp      gcaatacgcacagt-----------------atgcacaca-tgccaattacagactgacc
W25.2_415bp      caaatctccacaaatgccactcctcataataataataataataataatcaaatgctcacc
W26.2_400_bp     agattatccacaaa-accacttctcatattagtcaacacg------------agctcact
W27.2_403bp      ataatatctacgaa-acgccttctcatcactgtcaacatt------------tgctcaac
                 * * . .**..                 .*    *.            .** * .

Smalpha          agttgcagtcct-aacacat--------------cgatgggaagattcaaacaaacaata
W36.2_335bp      agttgctgtcctaaaaacat--------------caatgggaagatccaaacaaacaata
W25.2_415bp      aatgactgaattgaagagatatttcgtggagttgtagtgagaagcagtgaccagtggagt
W26.2_400_bp     agtgactgatttcaacaggtatttcctggagttctggtgagaagcagagaccagtgcagt
W27.2_403bp      agtgactggcttgaagaggcatttcctgaagttctcgtgagaagcggtgaccagtggatt
                 *.* .* *  .* ** * ..             . .**.****    .* **.   *

Smalpha          ctaagtg-----------------------------------------------------
W36.2_335bp      ctaagta------------aatttcaact-------------------------------
W25.2_415bp      tgaagccacgtctgttgtgagatgtcaactgactgaagacaattgtgaacg-gttggtga
W26.2_400_bp     tgaagcggtttgttttgaaaaatat---ctcactgaagacattggtggatgtgtcgctga
W27.2_403bp      ggaagcag-gtgtgatgtgagatggtcactcaatgaagacattggtggatgtgtcggtca
                 ***.

Smalpha          -------------
W36.2_335bp      ------tcacccc
W25.2_415bp      acttggtggattg
W26.2_400_bp     atatcgtggatcg
W27.2_403bp      tttcagtgaatcg
```



Active Hammerhead Ribozymes are Part of Autosomal WEs





No.	WE (V7)	WE (V5)	WE monomer lengths	Number of copies	Scaffold	Start	Stop	Total length of the array
1	W1.2	W1; W23	475	270	SM_V7_W003	320,915	513,415	192,500
2	W2.2	W2; W3	709-711	up to 34.1	SM_V7_W001	422,658	479,629	56,971
3	W4.2	W4; W30	1,206	11.7	SM_V7_W004	1	14,111	14,110
4	W5.2	W5	1,104	41.2	SM_V7_W018	1	45,515	45,514
5	W6.2	W6; W18; W35	715-718	up to 45.7	SM_V7_W014	1	53,884	53,883
6	W7.2	W7	980	29	SM_V7_W016	1	28,424	28,423
7	W8.2	W8	538	60.6	SM_V7_W015	1	32,211	32,210
8	W10.2	W10	671	59.2	SM_V7_W021	1	39,689	39,688
9	W11.2_W002	W11; W14; W28	903	33.8	SM_V7_W002	712,429	755,874	43,445
	W11.2_ZW	W11; W14; W28	1,294	29.5	SM_V7_ZW	1,1804,837	11,863,496	58,659
10	W12.2	W12	475-499	up to 46.1	SM_V7_W004	206,491	279,311	72,820
11	W13.2	W13; W17; W20; W33	524-646	up to 23.8	SM_V7_W003	229,593	276,593	47,000
12	W16.2	W16; W21	317	204.2	SM_V7_W007	57,292	122,025	64,733
13	W22.2	W22	604	106.6	SM_V7_W008	1	64,369	64,368
14	W24.2	W24	636	3.6	SM_V7_W020	1	2,259	2,258
15	W25.2	W25	415-428	up to 115.4	SM_V7_W012	1	49,306	49,305
16	W26.2_W003	W26	399-402	up to 146.7	SM_V7_W003	1	161,189	161,188
	W26.2_W010	W26	400	152.6	SM_V7_W010	1	61,065	61,064
17	W27.2	W27; W29	403	101.3	SM_V7_W017	1	40,810	40,809
18	W31.2	W31; W34	260	205.5	SM_V7_W004	41,351	114,609	73,258
19	W36.2_W001	W36	333-335	173.4	SM_V7_W001	1	104,272	104,271
	W36.2_W005	W36	332	up to 450.5	SM_V7_W005	1	149,646	149,645

WE	Length	Presence on autosomes	Minimal – maximal length	Copy numbers on autosomes	WE on autosomes
W2.2	709-711	all	37 – 705	68	parts
W4.2	1,206	all	44 – 390	1,101	parts
W5.2	1,104	all	26 – 742	3,454	parts
W6.2	715-718	all	36 – 161	5,382	parts
W7.2	980	2,4,6,7	80 – 156	5	parts
W11.2	903-1,294	all	30 – 514	18,610	parts
W12.2	475-499	all	26 – 102	172	parts
W13.2	524-646	1,2,4,6	38 – 131	19	parts
W16.2	317	all	26 – 314	203	parts
W22.2	604	all	26 – 232	1,897	parts
W24.2	636	all	28 – 450	2,302	parts
W25.2	415-428	all	25 – 428	39,648	parts and full-length (408)
W26.2	399-402	all	25 – 399	25,162	parts and full-length (321)
W27.2	403	all	27 – 403	55,303	parts and full-length (174)
W36.2	333-335	all	25 – 335	42,670	parts and full-length (858)