



HAL
open science

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

► To cite this version:

Maria Stitz, Cristian Chaparro, Zhigang Lu, V Janett Janett Olzog, Christina E Weinberg, et al.. Satellite-like W Elements: repetitive, transcribed, and putative mobile genetic factors with potential roles for biology and evolution of *Schistosoma mansoni*. *Genome Biology and Evolution*, 2021, 13 (10), pp.evab204. <10.1093/gbe/evab204>. <hal-03333536>

HAL Id: hal-03333536

<https://hal.science/hal-03333536v1>

Submitted on 3 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

1
2
3 **1 Satellite-like W elements: repetitive, transcribed, and putative mobile genetic factors with**
4 **2 potential roles for biology and evolution of *Schistosoma mansoni***
5
6
7
8
9

10 4 Maria Stitz^{1*}, Cristian Chaparro^{2*}, Zhigang Lu¹, V. Janett Olzog³, Christina E. Weinberg³, Jochen Blom⁴,
11 5 Alexander Goesmann⁴, Christoph Grunau^{2#}, Christoph G. Grevelding^{1#}
12
13
14
15
16

17 7 ¹Institute of Parasitology, BFS, Justus Liebig University Giessen, Germany; ²IHPE, University
18 8 Montpellier, CNRS, IFREMER, UPVD, IHPE, F-66000 Perpignan, France; ³Institute for Biochemistry,
19 9 Leipzig University, Germany; ⁴Bioinformatics and Systems Biology, Justus Liebig University Giessen,
20 10 Germany.
21
22
23
24
25

26 12 * Authors equally contributing to the work
27

28 13 # Authors sharing paper correspondence
29
30
31

32 15 Correspondence:

33 16 Prof. Christoph G. Grevelding
34 17 Justus Liebig University
35 18 BFS, Institute of Parasitology
36 19 Schubertstr. 81
37 20 35392 Giessen
38 21 Germany
39 22 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

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

24

25

26

27

28

29

30

59 31 © The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
60 32 This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License
(<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any
medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

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.

68 Introduction

69 Schistosomes are parasitic plathyhelminths and cause schistosomiasis (bilharzia). Listed as neglected
70 tropical disease (NTD) by the WHO, schistosomiasis ranks close to malaria in terms of parasite-induced
71 human morbidity and mortality (Hotez and Kamatha 2009; Colley et al. 2014). The life cycle of the
72 parasite is complex with a fresh-water snail of the genus *Biomphalaria* as intermediate host, producing
73 human-dwelling larvae (cercariae) that develop in the vertebrate host into adults, pair and produce
74 eggs that are excreted with the faeces and hatch snail-infecting miracidia when in contact with water.
75 Schistosomes are evolutionary unique as they are the only digenean parasites that have evolved
76 separate sexes (Moné and Boissier 2004). Heteromorphic sex chromosomes determine the sex of
77 schistosomes, with ZZ in the males and ZW in the females. Earlier studies from us and others has shown
78 that the females-specific region of the W chromosome is essentially composed of repetitive DNA
79 sequences (W-Elements, WE) arranged in large satellite blocks (Lepesant et al. 2012a; Lepesant et al.
80 2012b). WE sequences are abundant on the W chromosome, but may occasionally be transmitted to
81 autosomes, shown for the female W1 repeat sequence (Grevelding 1999). In most species with Y or W
82 sex chromosomes, it is found that: (i) repetitive sequences accumulate on these chromosomes, (ii)
83 large regions are heterochromatic, and (iii) these chromosomes deteriorate or are completely absent
84 in an extreme case. Among the accepted evolutionary theories is that the accumulation of repetitive
85 sequences on one of sex chromosomes has facilitated recombination suppression between the
86 heterochromosomes thus protecting sexually beneficial loci (Ohno 1967). Another hypothesis is that
87 chromosome rearrangements, sequence accumulation, and amplification may have occurred near the
88 sex-determining locus as a result of suppression of recombination (Charlesworth 1991; Marais and N.
89 Galtier 2003). Heterochromatization of the W chromosome in schistosomes has long been known and
90 has even been used as a marker for sex identification at the cercaria stage, where male and female
91 phenotypes are indistinguishable (Grossman et al. 1980; Grossman et al. 1981; Liberatos & Short
92 1983). Our earlier results (Lepesant et al. 2012a) showed that the repetitive sequences located in the
93 heterochromatic region of the W chromosome carry a euchromatic signature in the miracidia stage
94 (presence of H3K9ac and H3K4me3), which is gradually lost during development into adults. In most
95 species, the structural change of the chromatin is a highly organised process, and the observed
96 heterochromatin/euchromatin cycle of W-specific repeats suggests functional importance. Full or
97 partial euchromatization could not only allow transcription of functional WE but may also open space
98 for mobilisation of these elements. The WEs W1 and W2 were originally found in Puerto Rican isolates
99 of *Schistosoma mansoni* (Webster et al. 1989; Drew and Brindley 1995; Spotila et al 1987). Southern
100 blot analysis demonstrated their female-specific occurrence, and *in situ* hybridization localized W1 and
101 W2 in the heterochromatic region of the W-chromosome (Hirai et al. 1989; Spotila et al. 1989).
102 According to their repetitive nature, both elements were classified as satellite-like DNA (satDNA).

1
2
3 103 Subsequent studies in the Liberian strain and further strains of *S. mansoni* demonstrated W1 and W2
4 104 occurrence in both sexes of nearly all investigated strains, including another isolate from Puerto Rico
5 105 (Grevelding 1995; Quack et al. 1998). In contrast to the polymorphic occurrence of W1 and W2,
6 106 another family of repetitive elements, called D9 (Spotila et al. 1989), repeatedly occurred in the
7 107 investigated schistosome strains (Quack et al. 1998). Molecular analyses of clonal populations
8 108 obtained from crossing experiments with characterized parental generations, including males with and
9 109 without W1 and W2, showed the variable emergence of both WE among 10 generations of male
10 110 progeny - even when the parental male exhibited no W1 and W2 (Grevelding 1999). These data
11 111 indicated a meiotic level of (illegitimate) recombination and suggested a Z-chromosomal or autosomal
12 112 existence of W1 and W2 in the male progeny of these crosses. However, diversity in W1/W2
13 113 abundance and presence occurred also among siblings of single offspring generations. This pointed to
14 114 an additional process generating variable copy numbers of WEs in the genome. Indeed, the analysis of
15 115 clonal cercariae, generated by monomiracidial snail infections, demonstrated differences in W1 and
16 116 W2 copy numbers within clonal cercariae obtained from the same snails but at different time points
17 117 post infection (Grevelding 1999). This was unexpected since clonal cercariae were considered to be
18 118 genetically identical (Jourdane and Theron 1980; Jones and Kusel 1989). Additionally, the genetic
19 119 heterogeneity found in individuals of the same clonal cercarial population led to the hypothesis that
20 120 mitotic recombination may contribute to variable WE copy numbers in males, because only asexual
21 121 but no sexual reproduction processes take place in the intermediate host (Grevelding 1999). Using an
22 122 *in vitro* technique that allowed the generation of defined clonal daughter sporocysts originating from
23 123 a single mother finally confirmed that W1 copy numbers can vary among daughter sporocysts
24 124 generated by one defined mother (Bayne and Grevelding 1993). These results clearly pointed to the
25 125 possibility of mitotic recombination or transposition in the intermediate host. The underlying
26 126 mechanisms of WE variability remained obscure.

27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43 127 A microarray study with RNA of female cercariae provided first evidence for W1 and W2 transcripts
44 128 (Fitzpatrick et al 2008). The authors hypothesized that these transcripts may be associated to germ-
45 129 line protection in female schistosomes such as interfering with retrotransposable element activity, as
46 130 previously proposed for repeat elements in *Drosophila* (Pélisson et al. 2007). In 2012, a comprehensive
47 131 sequencing analysis to *de-novo* assemble *S. mansoni* repeat elements based on version 5 (V5) of the
48 132 genome (Berriman et al. 2009) predicted 36 WEFs (Lepesant et al. 2012b).

49
50
51
52
53 133 In addition to their peculiar evolutionary position as the only dioeciously living trematodes, another
54 134 unique feature of schistosome biology is the essentiality of a constant pairing contact for the sexual
55 135 maturation of the female. Pairing induces mitotic activity and differentiation processes in the female
56 136 that finally lead to the development of the female gonads, ovary and vitellarium (Popiel and Basch
57 137 1984; Den Hollander and Erasmus 1985; Kunz 2001; Grevelding 2004). This is a prerequisite for egg

1
2
3 138 production and closely associated with the pathology of schistosomiasis since eggs, which fail to reach
4 139 the gut lumen, migrate via the blood stream to liver and spleen. Here, these eggs lodge in the tissues
5 140 causing granuloma formation, inflammatory processes, and fibrosis (Olveda et al. 2014). The sexual
6 141 maturation status of a female is reversible. Upon separation, egg production stops and females de-
7 142 differentiate to an immature status. Upon rearing, differentiation of the gonads and egg production
8 143 start again (Clough 1981). Although the underlying processes have not been completely understood,
9 144 recent transcriptomics approaches have highlighted a complex scenario of male-female interaction. A
10 145 comparative RNA-Seq analysis of paired and unpaired adult *S. mansoni* and their gonads showed the
11 146 occurrence of >7,000 gene transcripts in the gonads of both sexes, of which 243 (testes) and 3,600
12 147 (ovaries) were pairing-dependently transcribed. High numbers of differentially expressed genes in the
13 148 ovary were expected because of the pairing-induced sexual maturation of females (Erasmus 1973;
14 149 Shaw 1987; Kunz 2001; Neves et al., 2005; Beckman et al., 2010). Among others, evidence was
15 150 obtained for the participation of neuronal processes, guided by GPCRs (G protein-coupled receptors)
16 151 and neuropeptides, in male-associated processes of the male-female interaction (Hahnel et al. 2018;
17 152 Lu et al. 2019). In addition, kinase signaling seem to dominate processes in females leading to gonad
18 153 differentiation and the maintenance of the sexual maturation status (Grevelding et al. 2018).
19 154 In this study, we challenge the classical view that repetitive DNA on the sex chromosomes is simply a
20 155 by-product of heterochromatization and provide further evidence for their functional importance. We
21 156 made use of genome sequencing updates in combination with recently obtained transcriptome data
22 157 sets to re-analyse WEFs, their structures, their chromosomal occurrence, their physical relationships,
23 158 and their developmentally regulated and strain-associated transcriptional activities. From the obtained
24 159 results we propose a hypothesis for their functional roles in schistosome biology and evolution.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 160 42 161 **Results**

43 162 **There are 19 W-Element Families in *S. mansoni***

44 163 To get a new overview of WEFs in *S. mansoni*, we performed a local BLAST (Basic Local Alignment
45 164 Search Tool) search of the published WEF (Lepessant et al. 2012b), which were based on version 5 (V5)
46 165 of the genome, against version 7 (V7; PRJEA36577, Puerto Rican strain;
47 166 <https://parasite.wormbase.org>; Howe et al. 2017). Except for W32, we found all 36 examined WEF
48 167 with a high percent identity on W-scaffolds, which cover DNA sequences, which are likely derived
49 168 from/correspond to the female-specific sex chromosome W but have not been exactly assembled yet.
50 169 We detected differences in DNA-sequence similarity among individual WEs of a single WEF. For
51 170 instance, the ten most significant hits of W6 elements (monomer repeat length: 310 bp) showed no
52 171 mismatch and correspond 100%, whereas the ten most significant hits of W23 elements (monomer
53 172 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.

183

184 **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

185

186 Tab. 1: Based on genome version V7, a number of 19 WEF (WE V7) has been newly defined (first column, No. 1 -
 187 19). Each WEF consists of one or more of those repeats that were previously found as individual WEs in version
 188 5 of the genome (WE V5; Lepesant et al. 2012a). In some cases (No. 9, 16 and 19), representing WEF (W11.2,
 189 W26.2 and W36.2, respectively) WEs were found on two different scaffolds each associated with the W
 190 chromosome and thus split into two subfamilies each. Furthermore, information of the length of WE monomers

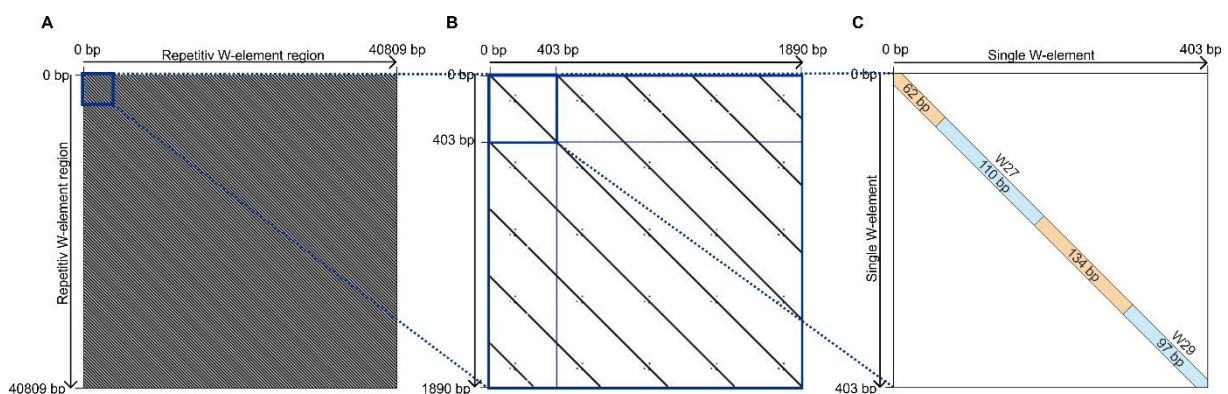
191 within WEFs is given as well as copy numbers (floating points indicate partial sequences), scaffolds with start and
 192 stop positions, and the total lengths of WEFs of the array. In some cases, the individual lengths of a WE monomer
 193 within a unit varied. For instance, a monomer of WEF 2.2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

196

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

210

211 Figure 1. The New Definition of WEF W27.2



212
 213 Figure 1: Showcase for the new definition of WEFs in *S. mansoni*. **A**, Result of the dotplot analysis for WE W29 in
 214 V7. The regular stripe pattern indicates a tandem repeat structure of this element without noticeable inversions
 215 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
 216 for this WE unit. **C**, a further close-up showing that W29 is closely associated to W27, which finally led to the
 217 new designation WEF W27.2 (Table 1).

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

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

225 **Some WEs Are Also Present On Autosomes**

226 Former studies reported on the presence of WEs in male *S. mansoni* (Grevelding 1995; Quack et al.
 227 1998). These findings suggested the presence of WE on Z and/or on autosomes. To find evidence for
 228 one or both possibilities, we took WE monomers of the newly defined WEFs and WE-transcript
 229 sequences detected in males for BLAST analyses against V7 in the NCBI database, which allowed
 230 assigning hits to individual chromosomes. For 4 out of 19 WEFs (W1.2, W7.2, W8.2, W10.2), we found
 231 no autosomal localization. Represented by either full-length or partial WE sequences, 15 WEFs were
 232 distributed among all 7 autosomes (Table 2 and Supplementary Figure 3A).

234 **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)

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

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

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

252 For some of the parental WEFs of those partial, autosomal WEs, we identified sequence similarities to
253 known mobile genetic elements (Supplementary Figure 4). W2.2 showed 65% sequence identity over
254 a stretch of 607 bp to the LTR retrotransposon Saci-1 of *S. mansoni*, which is 5,980 bp in length
255 (DeMarco et al. 2004). Open reading frame (ORF) analysis indicated that this sequence stretch covered
256 the complete protease-coding part of Saci-1. Within WEF W4.2 (1,206 bp) a short but significant
257 sequence homology of 44 bp (89% identity) occurred to the LTR retrotransposon Boudicca (5,858 bp)
258 (Copeland et al. 2003). A 120 bp fragment of W5.2 (1,104 bp) showed 96% identity to the non-LTR
259 retrotransposon Perere-2 (4,544 bp) (DeMarco et al. 2005). This fragment covered part of the reverse
260 transcriptase gene. In W11.2_ZW (1,294 bp), a partial sequence of 150 bp showed 84% identity to the
261 non-LTR retrotransposon Perere-3 (DeMarco et al. 2005). Finally, in W16.2_ZW (1,294 bp) a partial
262 sequence of 317 bp showed 81% identity to the DNA transposon Curupira-1 (4,878 bp) (Jacinto et al.
263 2011). An analysis of potential ORFs indicated partial sequence homologies of W2.2 to *gag* (Group
264 antigen) and *pol* (a reverse transcriptase), and of W5.2 to endonuclease-reverse transcriptase and *pol*
265 (data not shown). Genes like *gag*, *pol*, and endonuclease-reverse transcriptase are parts of mobile
266 genetic elements such as retrotransposons or retroviruses. However, we did not find a complete ORF
267 for a transposase or an endonuclease-reverse transcriptase in one of the investigated WEFs. The
268 exception is W36.2_W001 (335 bp), which showed strong sequence similarity to the SMalpha family
269 of SINE-like retrotransposons as reported in an independent study (Ferbeyre et al. 1998). Furthermore,
270 a focused structural analysis on one element showed direct repeat sequences as part of a monomer
271 of WEF W25.2. In this case, duplicate sequence stretches were identified flanking this WE at the
272 presumptive target site. The duplicated sequences differed from the W25.2 sequence and could have
273 originated from target site duplication (TSD) (Supplementary fig 5; data not shown).

1
2
3 2744
5 275 **WE Transcripts Occur in All Investigated Strains, Life stages, Sexes, and the Gonads of *S. mansoni***

6 276 To get a first overview of WEF transcript occurrence in the different biological samples we performed
7 277 a sample-distance matrix analysis. By pairwise comparisons, this approach provided information about
8 278 WEF transcript amounts based on the summation of all reads of transcribed WEs - independent of the
9 279 composition of each WEF that contributed to the WE transcript pool of each sample. These samples
10 280 included miracidia, cercariae, sporocysts, female and male schistosomula, paired and unpaired males
11 281 and females, as well as testes and ovaries of paired and unpaired females and males, respectively, and
12 282 also samples of other strains, depending on availability of sequencing data. We selected data sources
13 283 with at least two biological replicates for the individual life stages, both sexes, and the gonads
14 284 (Supplementary Table 1) and normalized the data using DESeq2 prior to further analysis. When
15 285 available, we also included replicates.

16 286 The result indicated remarkable differences in WE transcript amounts among the investigated samples.
17 287 Furthermore, we observed clustering of samples with similar or dissimilar levels of the total amount
18 288 of WE transcripts (Supplementary figs. 6 and 7). In the Liberian strain, miracidia (Mir_1, Mir_3, Mir_6)
19 289 showed high deviations of WE transcripts compared to other stages but these levels differed from
20 290 miracidia of the Puerto Rican strain (Mir). We detected similar deviations for cercariae of the Puerto
21 291 Rican strain (Cer1-3) and the Liberian strain (CerM1-CerM3; CerF1-CerF3). Generally, schistosomula
22 292 samples (SomF/M) and samples from males and testes (sM1-sM3; bM1-sM3; sT1-sT3, bT1-bT3)
23 293 seemed to be homogeneous with respect to transcript amounts. Bigger differences occurred between
24 294 unpaired females (sF1-sF3), which revealed an overall higher WE transcript level compared to paired
25 295 females (bF1-bF3). Finally, we found differences also among biological replicates as exemplified by
26 296 samples from females and ovaries (bF1-bF3; bO1-bO3; sO1-sO2), and even among replicates as
27 297 observed for schistosomula (SomM/SomF samples). In summary, this first overview indicated a high
28 298 degree of variability in overall WE transcript amounts between life stages, strains, and sexes, but also
29 299 between biological replicates.

300

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

303

304 Next, we performed differential expression analysis to investigate WEF expression profiles across
305 strains, different life stages (from miracidia to adults), sex-, gonad-, and/or pairing-dependent
306 expression. Again, we discovered remarkable differences (Figure 2 and Supplementary figs. 8 and 9).
307 Strain-dependent expression occurred, among others, for W7.2, with transcripts in schistosomula of
308 the Puerto Rican strain but not in schistosomula of the Guadeloupean strain. We found strain-
309 dependent expression differences also between the Guadeloupean and Liberian strains. For example,
310 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

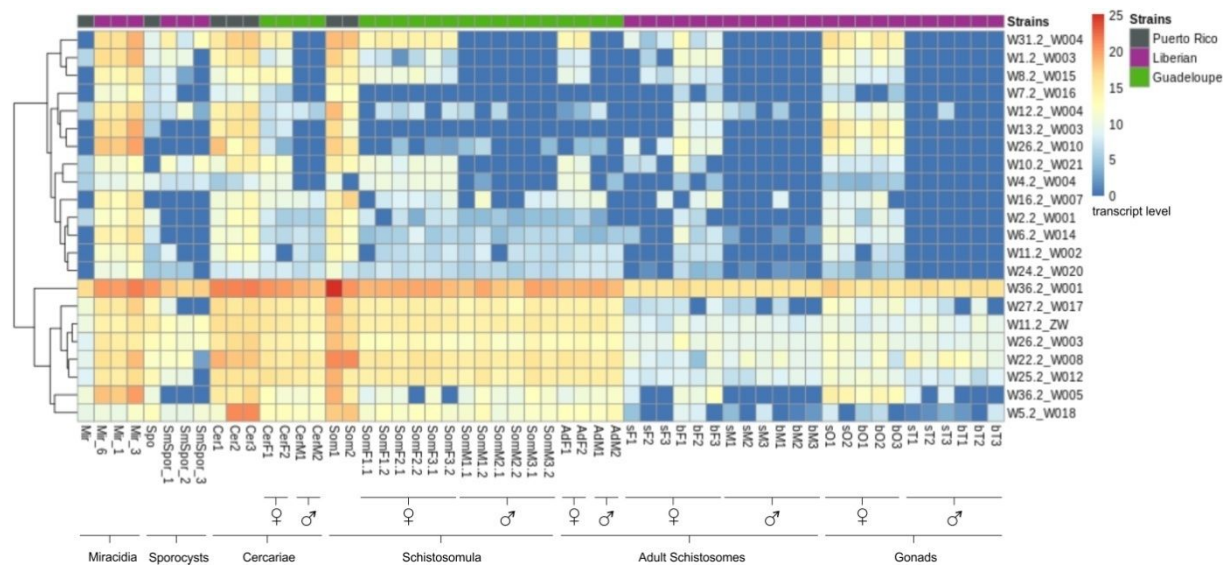


Figure 2: Heatmap generated by Deseq2 analysis showing the hierarchical clustering of transcript profiles of WEs of all 19 WEFs in the following life stages: Mir, miracidia; Cer, cercariae; (Sm)Spo(r), sporocysts; Cer(M/F), cercariae (male/female); Som(F/M), schistosomula (female/male); AdM, adult paired 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

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

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

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

354

355 **Evidence for Non-Coding RNAs in WEFs**

356 Since WEF transcripts occurred, and since we detected nearly no continuous ORFs indicating protein-
357 coding information, we next investigated whether WEFs may contain ncRNAs. These RNAs encode no
358 proteins, but as regulatory RNAs they can directly influence cellular processes (Mattick and Makunin
359 2006; Hombach and Kretz 2016). To search for ncRNAs, we analysed the 19 WEFs using
360 StructRNAFinder, which predicts and annotates RNA families in transcripts or genome sequences
361 (Arias-Carrasco et al. 2018).

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

1
2
3 373 females and ovaries compared to males and testes of the Liberian strain; Supplementary Figure 9). For
4
5 374 mir-891 sequence-containing W8.2, we found sex-, gonad-, strain-, and stage-influenced patterns
6
7 375 (Figure 2). Interestingly, in the Guadeloupean strain a clear sex-biased expression level of W8.2
8
9 376 occurred. We detected W8.2 transcripts in female cercariae, female schistosomula and adult females,
10
11 377 whereas no W8.2 transcripts occurred in the male-sample counterparts of this strain (Supplementary
12
13 378 Figure 8). Among other possibilities, these observations suggest roles for W12.2 and W8.2 in sex-
14
15 379 related processes.

16 380 For sequence parts of WEFs W2.2, W5.2, W11.2_ZW, W22.2, W24.2, and W26.2_W003,
17
18 381 StructRNAFinder predicted similarities to small nucleolar RNAs (snoRNAs) (Supplementary Table 3).
19
20 382 Together with associated proteins, snoRNAs form ribonucleoprotein complexes directing the post-
21
22 383 transcriptional modification of target RNAs (Lui and Lowe 2013). WEFs potentially encoding snoRNAs
23
24 384 showed varying transcript amounts in different strains, life-cycle stages, sexes, and tissues. Among
25
26 385 these is a candidate for SNORD59. It is potentially encoded by WEF W2.2, which contains the snoRNA
27
28 386 C/D family-specific C-box (UGAUGA) and D-box (CUGA) motifs (Galardi et al. 2002) within the SNORD59
29
30 387 region of W2.2 (Supplementary Figure 12). The W2.2 expression level was higher in cercariae and
31
32 388 schistosomula of the Puerto Rican strain compared to the Guadeloupean strain (Supplementary
33
34 389 Figure 8). In the Liberian strain, W2.2 showed preferential expression in miracidia, paired females, and
35
36 390 ovaries (Supplementary Figure 9). Expression levels of WEFs containing snoRNA candidates TB11Cs2H1
37
38 391 (W26.2_W003) and GlsR19 (W36.2_W001) in adults of the Guadeloupean strain exceeded those from
39
40 392 adults of the Liberian strain (Figure 2). In the latter strain, W26.2_W003 and W36.2_W001 expression
41
42 393 dominated in miracidia compared to other samples (Supplementary Figure 9). Within the sequence of
43
44 394 W22.2, different snoRNA candidates were predicted, sR11, snoZ178, and SCARNA7 (Supplementary
45
46 395 Table 3 and Supplementary Figures 10 and 11). W22.2 appeared expressed at higher levels in miracidia
47
48 396 and schistosomula of the Puerto Rican strain compared to the Guadeloupean strain. In the Liberian
49
50 397 strain, W22.2 expression is higher in miracidia than in other samples (Supplementary Figure 9).

51 398 In case of W5.2 and W11.2_ZW, StructRNAFinder predicted sequence parts for both miRNA and
52
53 399 snoRNA, which partly overlapped (Supplementary figs. 10 and 11). W5.2 and W11.2_ZW expression
54
55 400 levels appeared to be higher in the Puerto Rican strain compared to the Guadeloupean and Liberian
56
57 401 strains (Figure 2 and Supplementary Figure 8). However, compared to W11.2_ZW showing constant
58
59 402 expression among all samples of the Liberian strain, W5.2 expression was higher in miracidia and
60
61 403 sporocysts than all other stages and tissues of this strain (Supplementary Figure 9).

62 404 At this stage of the analysis, however, it is unclear whether these predictions of miRNA and snoRNA
63
64 405 correspond to biologically relevant ncRNAs. In addition, we found four WEF that contain sequences
65
66 406 reminiscent of the HHR class of ribozymes. These are catalytic RNAs and defined as ncRNA molecules
67
68 407 that can catalyse chemical reactions (Lilley 2019; see next section).

1
2
3 408
4
5 409
6
7 410
8
9 411
10
11 412
12
13 413
14
15 414
16
17 415
18
19 416
20
21 417
22
23 418
24
25 419
26
27 420
28
29 421
30
31 422
32
33 423
34
35 424
36
37 425
38
39 426
40
41 427
42
43 428
44
45 429
46
47 430
48
49 431
50
51 432
52
53 433
54
55 434
56
57 435
58
59 436
60
454
455

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      atcgcacaaagcaagtggct-----atcaggactcagtggtcgagtgataacgcgat
W36.2_335bp  attgcacaagcaagtggct-----atcaggactcagtggtcgagtgataacgcgat
W25.2_415bp  gttgaaatagacataaacaccatgggatgccgtctcagtggtgatgtgttgagtgtctc
W26.2_400_bp gttgaagatggaaatgaacacgggtggatgccgactcagtggttcttgatgttctcagctt
W27.2_403bp  tttgaagtaaggcattgagagcgttggatggtgaccagtggtctagtgtttaagtgtctc
434          *. * * . . . * .          ** * *.***** .          *          .
435
Smalpha      ggtgtttgaagcgaggg-tactgggttggagtcccagagtgatatacaactctgagatgc
W36.2_335bp  ggcgtttgaagcgaaagttactgggttcgagtgccagaatggacatcaactctgagatgc
W25.2_415bp  tattgagaaagttgttatttctgtgttggaaatctgg-----tgaggcgaaatcgt
W26.2_400_bp gcgcacgaaaatgataagcgcagtttccgaatctgg-----agaggcggtatcgt
W27.2_403bp  acgcgcgagactgataggtgctgggttggaaatctcg-----cgagactgcatcgt
441          ..* . . . . * * * * * * * . .          . * * . . . * .
442
Smalpha      aggtacatccagctgacgagtgccca-aataggacgaaaacgcgcgtcctggattccactgc
W36.2_335bp  aggtacatccagctgacgagtgccca-aataggacgaaaacgcgcgtcctccattccactgc
W25.2_415bp  ggattggaaatgctgaggagtgccacaatagaacaaatgaccgtccagtggtttccagga
W26.2_400_bp ggatgcacactgctgagaaagatattaatgaagttccaaggctgctcag-attgtcgagg
W27.2_403bp  ggatgcgcactgctgagcagtgccacagtagcagaaaacggccatccagtgcttccaggt
446          .*. * .          ***** * .          *. * . . .          * * . . . * . . . *
447
Smalpha      tatccactattcatc-----tttgcttatcatgcttgtgaaatcaaggctatatcgag
W36.2_335bp  tagccactatccatc-----tttgcttaccatgcttgtgcaatttaggctatatcgag
W25.2_415bp  tttccctagtgggtctaccttcaatccactcatgatttcaaata-----taaata-act
W26.2_400_bp  ttaacatgacgggtctctcaacaattgattcatcatcataacca-----ttaacattacg
W27.2_403bp  tttccatcgtgctccagcttcaattcattcatcatctcaacta-----tcaacattact
454          * * . . . .          * . . . . * . * * . .          *          * . * . .
    
```

```

1
2
3 456 Smalpha gcaatacgcacagt-----atgcacata-tgcccaattagagactgacc
4 457 W36.2_335bp gcaatacgcacagt-----atgcacaca-tgcccaattacagactgacc
5 458 W25.2_415bp caaatctccacaaatgccactcctcataataataataataataatcaaatgctcacc
6 459 W26.2_400_bp agattatccacaaa-accacttctcatattagtcacacg-----agctcact
7 460 W27.2_403bp ataatatctacgaa-acgccttctcatcactgtcaacatt-----tgctcaac
8 461 * * . **.. . * * . ** * .
9 462
10 463 Smalpha agttgcagtcct-aacacat-----cgatgggaagattcaaacaacaata
11 464 W36.2_335bp agttgctgtcctaaaaacat-----caatgggaagatccaaacaacaata
12 465 W25.2_415bp aatgactgaattgaagagatatttcgtggagttgtagtgagaagcagtgaccagtggagt
13 466 W26.2_400_bp agtgactgatttcaacaggatatttcctggagttctggtgagaagcagagaccagtgcagt
14 467 W27.2_403bp agtgactggcttgaagaggcatttctgaagttctcgtgagaagcggtgaccagtggatt
15 468 *. * . * * . * * * . . . ** . * * * * * * * * * *
16 469
17 470 Smalpha ctaagtg-----
18 471 W36.2_335bp ctaagta-----aatttcaact-----
19 472 W25.2_415bp tgaagccacgtctgttgtgagatgtcaactgactgaagacaattgtgaacg-gttggtga
20 473 W26.2_400_bp tgaagcggtttgttttgaaaaatat---ctcactgaagacattggtggatgtgtcgctga
21 474 W27.2_403bp ggaagcag-gtgtgatgtgagatggctcactcaatgaagacattggtggatgtgtcggtca
22 475 ***.
23 476
24 477 Smalpha -----
25 478 W36.2_335bp -----tcacccc
26 479 W25.2_415bp acttgggtggattg
27 480 W26.2_400_bp atatcgtggatcg
28 481 W27.2_403bp tttcagtgaatcg
29 482

```

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

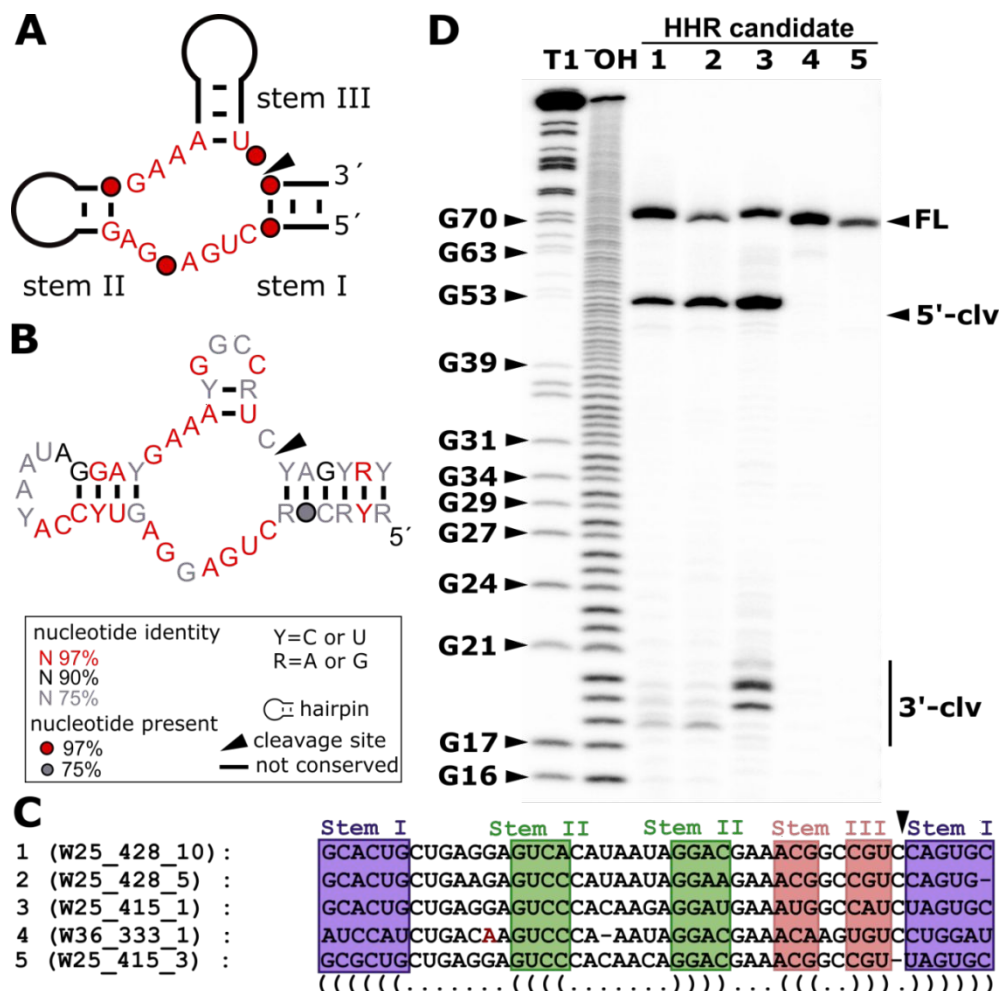
487 Varying sequence similarity for HHRs occurred within these WEF (Figure 4 and Supplementary
488 Figure 13), which we also observed for W-chromosomal and autosomal copies of these WEF showing
489 incomplete or altered HHR consensus sequences with deletions, insertions, and point mutations. As
490 shown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

492 Previous studies already confirmed the activity of the SMalpha-encoded type I HHR, represented by
493 W36.2_W001 (Ferbeyre et al. 1998; Canny et al. 2004). Prior to experimental validation of activities of
494 predicted HHRs of other WEFs, we analysed the appropriate sequences for intact three-stem junctions
495 and perfect conservation in the CUGANGA and GAAA sequence motifs to exclude presumably inactive
496 HHRs (Supplementary Table 4) (Ruffner et al. 1990). To investigate additional HHRs for their self-
497 cleavage activity *in vitro*, we selected predicted candidates from different WEFs (Figure 4). Because of
498 the high similarity between all candidates (Supplementary Figure 13), we selected only five
499 representatives. These candidates descend from WEF copies on autosomes and represent variants of
500 W25.2 and W36.2 (Figure 4C and Supplementary Table 4). Although the HHR core mainly comprised
501 the three-stem junction with its conserved CUGANGA and GAAA sequences (Figure 4A), additional
502 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



515
516 Figure 4: HHR candidates as part of WEs were selected for functional analysis. **A**, Consensus sequence and
517 secondary structure of the HHR type I catalytic core. The stems are labeled I-III, and highly conserved nucleotides
518 are shown in red. **B**, Consensus sequence and secondary structure of HHR candidates found as part of WEs. Only
519 those HHR candidates that contain all conserved sequence and structural features were included into this

1
2
3 520 consensus (Supplementary Table 4 and Supplementary Figure 13). R2R was used to draw the model (Weinberg
4 521 et al. 2011). **C**, Alignment of HHR sequences from selected candidates tested for self-cleavage activity *in vitro*.
5 522 The stems are highlighted, and mutations in highly conserved regions shown in red (i.e. Candidate 4-W36_333).
6 523 At the cleavage site (arrow), candidate 5 W25_415_3 has a deletion and served as control. **D**, Co-transcriptional
7 524 cleavage analysis of selected HHR candidates. Full-length transcript (FL), 5'- (5'-clv), and 3'-cleavage (3'-clv)
8 525 fragments are indicated by arrows. "T1" indicates partial digestion after G nucleotides by RNase T1 and "-OH"
9 526 partial alkaline hydrolysis. Samples were separated by 20% PAGE.
10
11
12
13
14 527

15 528 **WEFs Vary Among Different Schistosome Species**

16 529 Previous studies discussed the existence of WEFs and mobile genetic elements such as
17 530 retrotransposons in different schistosome species. In their original article on W1 elements in
18 531 *S. mansoni*, now WEF W1.2, Webster et al. (1989) reported the absence of W1 from *S. matthei*, *S.*
19 532 *haematobium*, *S. japonicum*, *S. douthitti*, and *Fasciola hepatica* based on Southern blot results in the
20 533 pre-genomic era. Furthermore, the SMalpha family (represented by W36.2), based on PCR data
21 534 originally assumed by Ferbeyre et al. (2000) to be present in *S. haematobium* and *S. douthitti* but
22 535 absent from *S. japonicum*, was later found in this species in high copy numbers and with the HHR core
23 536 motif (Laha et al. 2000). Based on genome data for schistosomes (Howe et al. 2017), we searched for
24 537 WEFs in *S. rodhaini*, which belongs to the same clade as *S. mansoni*, and in *S. haematobium* and *S.*
25 538 *japonicum*, which represent different clades, respectively (Lawton et al. 2011).

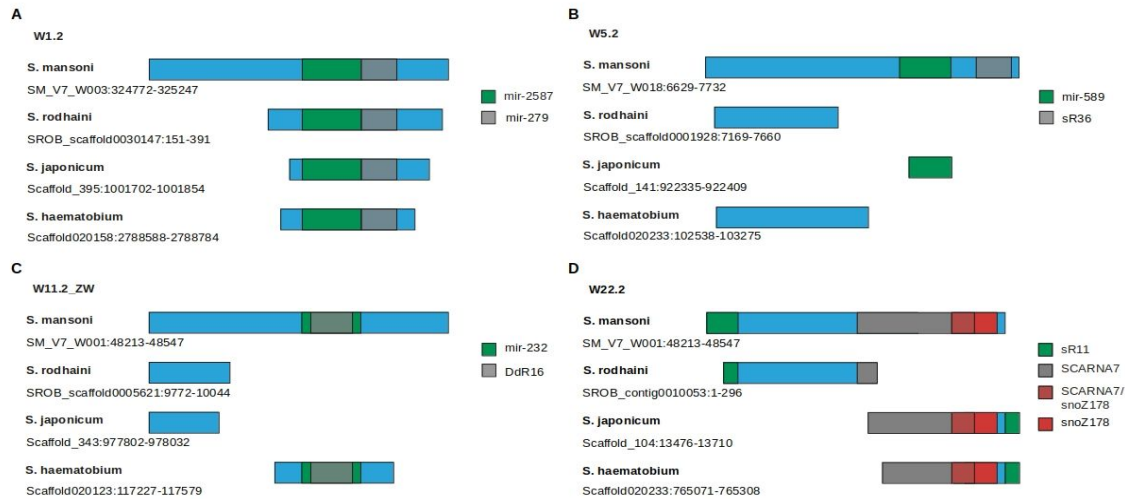
26 539 BLASTn analysis showed the presence of the majority of WEFs in the other species. Compared to
27 540 *S. mansoni*, however, lengths and structures of the elements varied. Most of the WEFs found in the
28 541 other species revealed reduced sizes (Supplementary Table 5). One example is W11.2_ZW, which in
29 542 *S. mansoni* has a length of 1,294 bp, whereas in *S. japonicum* its size is 251 bp, in *S. haematobium*
30 543 375 bp, and in *S. rodhaini* 275 bp, respectively. Partial versions of five WEF, W2.2_709, W4.2, and
31 544 W7.2, occurred in all investigated species except *S. japonicum*. In *S. rodhaini*, a nearly complete version
32 545 of WEF W31.2 exists, which is absent from *S. japonicum* and *S. haematobium*. Remarkably, full-length
33 546 and partial versions of WEF W8.2 and W10.2 exclusively occurred in *S. mansoni*. Only for WEF W36.2,
34 547 size differences of the repeat units were lower among the different species. For W36.2_W005 and
35 548 W36_W001, we found slightly larger unit sizes in *S. rodhaini* (347 bp and 337 bp, respectively)
36 549 compared to *S. mansoni* (332 bp and 335 bp, respectively) (Supplementary Table 5).

37 550 To investigate whether the size variations structurally affected regions with hypothesized regulatory
38 551 functions, we performed alignment analyses using the best BLASTn hits focusing on selected WEFs that
39 552 contain regions potentially coding for regulatory RNAs, W1.2 (mir-279, mir-2587), W5.2 (mir-598,
40 553 sR36), W11.2_ZW (mir-232; DdR16), and W22.2 (SCARNA7, sR11, snoZ178). This analysis indicated a
41 554 mosaic pattern with respect to the presence and integrity of these parts of the WEF sequences that
42 555 potentially code for regulatory RNAs. Whereas the part of W1.2 coding for the miRNAs mir-279 and

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

559

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



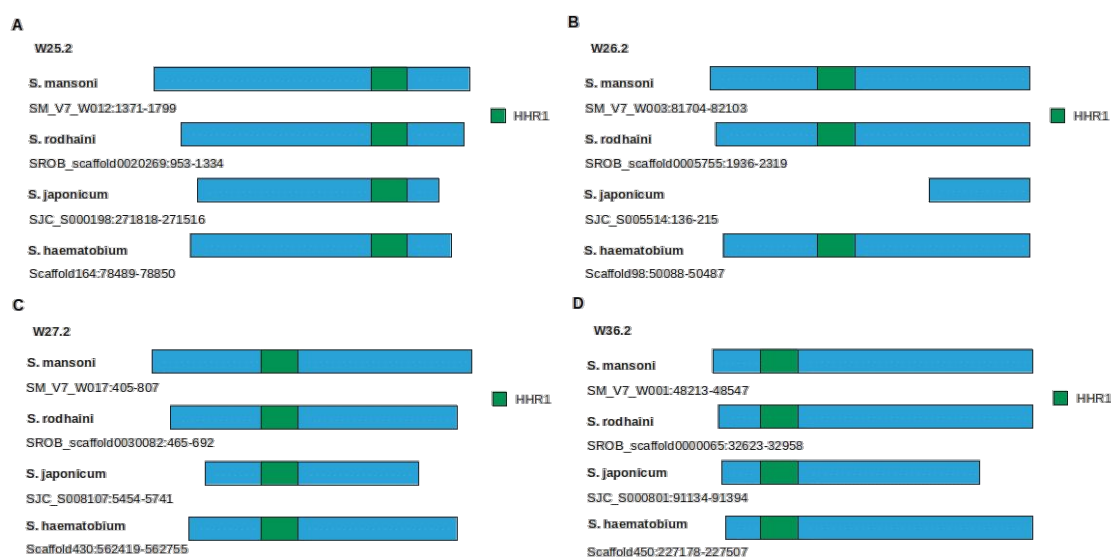
561
 562 Figure 5: Diagram of structural differences of selected WEF in *S. mansoni*, *S. rodhaini*, *S. japonicum*, and
 563 *S. haematobium* that potentially encode regulatory RNAs. **A**, W1.2 (blue) harbours coding sequences for miRNAs
 564 mir-2587 (green) and mir-279 (grey). Compared to *S. mansoni*, best W1.2 hits upon BLASTn analysis showed
 565 shorter orthologs in *S. rodhaini*, *S. haematobium*, and *S. japonicum*. Although shorter, these orthologous
 566 sequences preserved full-length mir-279 and mir-2587. **B**, W5.2 (blue) harbours coding sequences for mir-589
 567 (green) and snoRNA sR36 (grey). We found none of these regulatory RNAs in orthologous sequences of *S. rodhaini*
 568 and *S. haematobium*. In *S. japonicum*, a partial mir-2587 sequence occurs, however, not flanked by W5.2
 569 sequence areas. In this case, we cannot exclude a false positive hit. **C**, W11.2_ZW harbours overlapping coding
 570 sequences for mir-232 (green) and DdR16 (grey), which is part of the mir-232 sequence. This overlapping RNA-
 571 coding part appears to be completely preserved in *S. haematobium*, although the flanking W11.2_ZW sequence
 572 areas are shorter compared to *S. mansoni*. In *S. rodhaini* and *S. japonicum*, we detected smaller W11.2_ZW
 573 elements without RNA-coding parts. **D**, W22.2 harbours sR11 (green) and overlapping coding sequences for
 574 SCARNA7 (grey) and snoZ178 (red), which is part of the SCARNA7 sequence (red/grey area). Only shorter W22.2
 575 elements occur in the other schistosome species. In *S. rodhaini*, a small W22.2 variant exists containing fragments
 576 of the sR11 and SCARNA7 sequence. In *S. japonicum*, a big part of the overlapping RNA-coding region occurs with
 577 a small part of sR11 and a shortened SCARNA7 part, but a completely preserved snoZ178 part. We made a similar
 578 finding for *S. haematobium*, here the SCARNA7 (grey) part was slightly shorter compared to the *S. japonicum*
 579 counterpart. The sequences were obtained from data deposited on WormBase ParaSite (version 15, October
 580 2020; <https://parasite.wormbase.org>), BioProject numbers: PRJEA36577 (*S. mansoni*), *S. rodhaini* (PRJEB526 -

Downloaded from <https://academic.oup.com/gbe/advance-article/doi/10.1093/gbe/evab204/6361599> by guest on 03 September 2021

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

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

589
 590 **Figure 6. Structural Comparison of Hammerhead Ribozyme-Containing WEFs in Different**
 591 **Schistosome Species**



592
 593 Figure 6: Diagram of structural differences of WEF W25.2, W26.2, W27.2, W36.2 in *S. mansoni*, *S. rodhaini*,
 594 *S. japonicum*, and *S. haematobium* that potentially encode type 1 HHR1 (scaffold information is given). The
 595 ribozyme parts of these sequences are highlighted in green. Compared to *S. mansoni*, BLASTn analysis showed
 596 shorter orthologs for all four WEFs in *S. rodhaini*, *S. haematobium*, and *S. japonicum*. Except for W26.2 in *S.*
 597 *japonicum*, the ribozyme motif was highly conserved among the selected species. The sequences were obtained
 598 from data deposited on WormBase ParaSite (version 15, October 2020; <https://parasite.wormbase.org>),
 599 BioProject numbers: PRJEA36577 (*S. mansoni*), *S. rodhaini* (PRJEB526 - Republic of Burundi), PRJNA520774
 600 (HuSjv2, *S. japonicum*), and PRJNA78265 (*S. haematobium*). Scaffold numbers are given.

601 Discussion

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

1
2
3 607 Ohno 1999). This view has fundamentally changed (Biscotti et al. 2015a), and our study is the first one
4
5 608 addressing this issue with structural, functional, and evolutionary aspects for the genome of a
6
7 609 multicellular parasite.

8 610 Using V7 of the *S. mansoni* genome we re-analysed WEFs, which previous studies described as
9
10 611 organized in 36 families of repetitive, non-protein coding DNA elements occurring on the W-
11
12 612 chromosome (Lepesant et al. 2012b). From our results, we deduce 19 WEFs based on the fact that
13
14 613 many of the original 36 WEFs are fused together. These 19 WEFs exhibit a surprising diversity of
15
16 614 features. As expected for repetitive DNA, there is high sequence similarity but also variation among
17
18 615 the repeat units of each single WEF. Besides few partial WE sequences of some WEFs on the WZ
19
20 616 chromosomes, we found for 15 out of 19 WEF copies also on autosomes. Here, copy numbers
21
22 617 considerably varied as well as the lengths of WE units of the respective WEFs. The autosomal
23
24 618 occurrence possibly indicates the potential of element mobility. Indeed, in some WEF sequences we
25
26 619 found residues of features typical for mobile genetic elements such as retrotransposons or
27
28 620 retroviruses. Although we found no evidence of complete ORFs coding for a transposase, in an
29
30 621 exemplary case, we detected direct repeats and TSD, which are often, but not always, found at
31
32 622 integration sites of mobile genetic elements (Jensen et al. 1994; Han et al. 2014). Repetitive elements
33
34 623 such as satDNA have been discussed to originate from transposable elements (TEs). Both share
35
36 624 sequence similarity and organization patterns, which suggests a mutual relationship between satDNA
37
38 625 and TEs. This probably influenced satDNA evolution and its roles on genome architecture and function
39
40 626 (Meštrović et al. 2015; Satović et al. 2016). The mechanism of satDNA formation from TEs is unclear
41
42 627 yet. However, structural elements such as palindromes, direct/inverted repeats, and the ability of
43
44 628 stem-loop formation may be involved as well as illegitimate recombination events and deletions based
45
46 629 on double-strand breaks and excision. These factors might drive rearrangements of TEs and the
47
48 630 production of sequence parts as templates for further amplification to form tandemly organized
49
50 631 satDNA clusters. This may apply to at least some of the WEs described here (W2.2, W4.2, W5.2, W11.2,
51
52 632 and W16.2), which showed partial sequence homology to known mobile genetic elements. SatDNA
53
54 633 monomers might be recognized by transposase-driven mechanisms acting on short DNA sequence
55
56 634 motifs of these DNA sequences, which are recognized by enzymes related to transposases (Meštrović
57
58 635 et al. 2015). As alternative to transposon-like cut and paste mechanisms, reintegration of repeat
59
60 636 elements may happen via circular DNA intermediates generated from tandem repeats, as shown in
637
638 *Arabidopsis* (Woodhouse et al. 2010). Whether this or similar mechanisms may apply to the WEF of *S.*
mansoni has still to be elucidated.

639 Based on evidence for transcriptional activity and dynamic occurrence, recent years have unravelled
640
641 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

1
2
3 642 genome stability. In *S. mansoni*, Lepesant et al. (2012a) reported WEF-associated transcript occurrence
4 643 and chromatin status of repeats to be stage-specific. Open chromatin on the W-chromosome occurred
5 644 in larval stages but not in the sexually dimorphic adult stage. Furthermore, the euchromatic character
6 645 of histone modifications around the WE repeats on W decreased during the life cycle. When
7 646 transcribed, repeat RNA appeared stage-specifically in the larval stages miracidium and cercaria. These
8 647 authors concluded that WEFs may play roles in structural changes of the chromatin and sex-
9 648 chromosome emergence in ZW systems (Lepesant et al. 2012a). Furthermore, heterochromatisation
10 649 was discussed as major factor for the differentiation of sex chromosomes in the schistosomatids and
11 650 also for schistosome speciation (Lawton et al. 2011).

12 651 Here, we provide conclusive evidence for WEF expression throughout schistosome development, from
13 652 the miracidium to the adult stages. Copy numbers considerably varied among the WEFs and depended
14 653 on the life-cycle stage, sex, pairing, and strain. In a Liberian strain of *S. mansoni*, genome instability of
15 654 the W1 element was hypothesized from the finding of gain or loss of W1 elements in the progeny of
16 655 crosses, or even during larval development within the snail host (Grevelding 1999). This pointed to
17 656 recombination processes during meiosis and/or mitosis. Our new results suggest that a mobile
18 657 character of WEF may account for this genome instability, which allows WEs to recombine within a
19 658 single generation. This could also explain the observed strain differences. Accordingly, some of the
20 659 results shown (Figure 2 and Supplementary figs. 6-9), may represent snapshots of the WEF setup within
21 660 the samples at the time they were available for analysis. Therefore, WE occurrence and copy numbers
22 661 may be a matter of change over time and generations, a view that is supported by the observed
23 662 variations among biological and technical replicates. As practical consequence, sex determination of
24 663 clonal cercarial populations based on PCR using WE-specific primers, as previously performed with
25 664 varying success (Gasser et al. 1991; Boissier et al. 2001; Chevalier et al. 2016), may eventually lead to
26 665 inconsistent results.

27 666 Our approach to assign putative roles to WEFs led to the StructRNAFinder-based prediction of different
28 667 classes of regulatory RNAs as parts of WE sequences. Due to their roles in gene regulation, miRNAs
29 668 have attracted attention (Bartel 2018). In mammals, X-chromosomal miRNA expanded by gene
30 669 duplication after the emergence of sex chromosomes, and hundreds of different miRNAs have been
31 670 identified shaping mRNA expression (Meunier et al. 2013). Many are conserved, which may also
32 671 applies to *S. mansoni* (<http://www.mirbase.org/>; Kozomara et al. 2019). As example, mir-181, in our
33 672 study predicted as putative part of W8.2, may represent a miRNA family for which multiple roles in
34 673 immune cell development, hematopoiesis, cell death pathways, cancer, and drug resistance have been
35 674 reported in humans (Weng et al. 2015; Braicu et al. 2019). Species comparison indicated W8.2 in
36 675 *S. mansoni* but not in the closely related species *S. rodhaini* and *S. japonicum* or *S. haematobium*. In
37 676 *Drosophila*, mir-279, potential part of W1.2, influence neuron formation of the olfactory sensory

1
2
3 677 system (Hartl et al. 2011) and eye patterning interfering with EGFR pathways (Duan et al. 2018). Roles
4
5 678 for miRNAs have also been reported for parasites and host-parasite interaction (Zhao and Guo 2019;
6
7 679 Britton et al. 2020). Plant-parasitic nematodes induce feeding site formation in host cells, which
8
9 680 differentially express miRNAs upon infection (Jaubert-Possamai et al. 2019). In schistosomes, the
10
11 681 presence and conservation of genes involved in miRNA pathways and their role in
12
13 682 *B. glabrata*/*S. mansoni* interaction has been discussed (Queiroz et al. 2017; Cardoso et al. 2020).
14
15 683 Furthermore, miRNAs of *S. mansoni* and *S. japonicum* may participate in male-female interaction,
16
17 684 sexual development, and pathological processes in the final host (Zhu et al. 2014; Cai et al. 2016; Yu et
18
19 685 al. 2019). In *S. haematobium*, a genome-wide analysis of small ncRNAs identified mir-785 expression
20
21 686 in both sexes, which corresponds to our findings. Furthermore, homology-based prediction indicated
22
23 687 a voltage-dependent anion-selective channel protein (MS3_05034) as potential target (Stroehlein et
24
25 688 al. 2018). Transcripts of the MS3_05034 ortholog of *S. mansoni*, Smp_091240, occur in males, females
26
27 689 and their gonads (Lu et al. 2016). *S. mansoni* mir-785 is also expressed in adults although with a sex
28
29 690 bias (females > males) including the gonads (ovary >> testes), and influenced by pairing (paired females
30
31 691 >> unpaired females).

32
33 692 Also, snoRNAs are widely distributed among eukaryotes and participate in the modification and
34
35 693 processing of ribosomal and small nuclear RNAs, splicing, rRNA acetylation, mRNA abundance and
36
37 694 translational efficiency (Lui and Lowe 2013; Bratkovič et al. 2020). SNORD59, a putative snoRNA in
38
39 695 W2.2, is a member of the C/D family directing site-specific 2'-O-methylation of substrate RNA such as
40
41 696 18S rRNA (Galardi et al. 2002). The conserved C/D family-specific C-box (UGAUGA) and D-box (CUGA)
42
43 697 motifs are present in W2.2-encoded SNORD59. Perfect matches occur in the 709 bp variants of W2.2,
44
45 698 whereas in the 711 bp variants a G/C mutation has been found at position 3 of the D-box motif
46
47 699 (Supplementary Figure 12). The 2'-O-methylation of the ribose moiety is important for the maturation
48
49 700 of almost all classes of RNAs and involves different snoRNP (ribonucleoprotein) complexes (Kiss-László
50
51 701 et al. 1996; Ojha et al. 2020). They contain nucleolar proteins (Nop), of which orthologs exist in
52
53 702 *S. mansoni* such as Nop 56 (Smp_053470 and Smp_048660; <https://parasite.wormbase.org>).
54
55 703 Smp_053470 and Smp_048660 are most abundantly expressed in sporocysts and ovaries of paired
56
57 704 females (Lu et al. 2016; Lu et al. 2018; <http://schisto.xyz/>, V7), which coincides with SNORD59
58
59 705 transcripts in these stages/tissues (Supplementary Table 3). SnoRNAs can be further processed into
60
706 smaller RNAs with different functionality including miRNAs (Scott and Ono 2011). In mammals, protein-
707 coding genes exist that express both snoRNAs and miRNAs in single introns. The existence of eukaryotic
708 and archean members suggested that snoRNAs - in evolutionary terms - are more ancient compared
709 to miRNA (Dennis and Omer 2005). However, there is also evidence for recently evolved snoRNA and
710 miRNAs (Niwa and Slack 2007). This suggests that both RNA classes evolve dynamically and at fast
711 rates, which may also apply to the predicted *S. mansoni* snoRNA and miRNAs. Furthermore, many of

1
2
3 712 the most recently evolved snoRNA and miRNAs may be derived from TEs. Indeed, previous studies
4 713 showed TSDs at snoRT (human snoRNA-like retrogene) integration sites, which supports their mobile
5 714 character (Weber 2006; Lui and Lowe 2013). Our indicative findings of TSDs and intra-WE inverted
6 715 repeats flanking miRNA and snoRNA subunits of WEFs as well as their mobile character correspond to
7 716 these concepts. To prove whether the predictions of snoRNAs and miRNAs as parts of appropriate
8 717 WEFs have functional relevance will be subject of future studies. In this study, we focused on functional
9 718 evidence of type I HHRs, which are parts of some WEF. HHRs are widely distributed in the animal and
10 719 plant kingdoms, and they can be associated with repeated DNA (Perreault et al. 2011; Cervera et al.
11 720 2014; de la Peña et al. 2017; Lünse et al. 2017). Recent studies on the biology of genomic HHRs discuss
12 721 their potential roles in the propagation of retrotransposons, which are major components of
13 722 eukaryotic genomes including schistosomes (Venancio et al. 2010), and which contribute to genome
14 723 evolution shaping developmental processes and eukaryotic complexity (Mita et al. 2016). Remarkably,
15 724 our species analysis showed a high conservation of the ribozyme-like sequences in W25.2,
16 725 W26.2_W003, W27.2, and W36.2_W001 among the examined species with the exception of the
17 726 absence of the type 1 HHR in W26.2 in *S. japonicum*. To test whether WEF-encoded HHRs are
18 727 catalytically active, we selected predicted ribozyme sequences of autosomal WEs. Alignments showed
19 728 that some WEF-encoded HHR sequences harbour the canonical core sequences CUGANGA and GAAA.
20 729 Other HHRs, such as candidates 4 (W36_333_1) and 5 (W25_415_3), have mutations or deletions in
21 730 highly conserved nucleotides that likely lead to a reduction or loss of the cleavage activity (Ruffner et
22 731 al. 1990). Indeed, all HHR candidates conform to previously described minimal HHRs with an extremely
23 732 short stem III (Forster et al. 1987; Lünse et al. 2017). Although only weak similarity to SMalpha HHR
24 733 existed, it may be tempting to speculate about their functional role as part of SINE-like elements similar
25 734 to the ones already described (Ferbeyre et al. 1998). Furthermore, W25.2 showed higher expression
26 735 in the Puerto Rican and Guadeloupean strains compared to the Liberian strain. Thus, it may contribute
27 736 to strain-specific differences at the post-transcriptional level. Within the Liberian strain, W25.2
28 737 appeared to be expressed at a slightly higher level in the ovary of females compared to male testes or
29 738 the adults. This suggests a gonad-associated function in females, which awaits confirmation in
30 739 subsequent studies. These are needed to substantiate the expounded hypotheses and to assure that
31 740 *S. mansoni* WEs elements are not just selfish DNA (Biscotti et al. 2015b, Thakur et al. 2021).
32 741 Creating variability and genome plasticity are hallmarks of parasites with different molecular principles
33 742 invented during evolution (Lanzer et al. 1995). In schistosomes, repetitive WEs may represent one of
34 743 these principles. Using the intermediate snail-host stage for asexual recombination (Grevelding 1999;
35 744 Bayne and Grevelding 2003) and the final-host stage for sexual recombination, schistosomes have
36 745 exceptional preconditions for rapid evolution driven by adaptation to new environments. Prerequisite
37 746 for asexual recombination of WEFs from the heterochromatin area of the W-chromosome is a biphasic

1
2
3 747 chromatin stage, in which both euchromatic and heterochromatic states can be adopted. Indeed, this
4
5 748 has been confirmed for *S. mansoni* by showing that euchromatic histone modifications around WEs
6
7 749 dominate in the intermediate snail-host but decrease afterwards (Lepesant et al. 2012a). In
8
9 750 *P. falciparum*, bistable chromatin has been discussed as a mechanism regulating variant gene
10
11 751 expression patterns within clonal populations (Llorà-Batlle et al. 2019). In nematodes, mobile genetic
12
13 752 elements contribute to genome plasticity in the absence of sexual reproduction (Castagnone-Sereno
14
15 753 et al. 2014). This fits to the general view of the impact of mobile genetic elements on genome evolution
16
17 754 and adaptation, which is important for organisms frequently facing new environments such as
18
19 755 parasites (Schrader and Schmitz 2019), particularly those with complex life cycles. The evolution of
20
21 756 repetitive DNA was also associated with reproductive isolation, founding new species, genome
22
23 757 integrity, and karyotype evolution (Coghlan 2005; Biscotti et al. 2015a; Lower et al. 2018). Especially
24
25 758 with respect to heterochromatic regions, remarkable diversity of karyotype patterns exist for different
26
27 759 schistosome species leading to the definition of six clades correlating with the different geographical
28
29 760 distribution as well as with the hypothesized Asian origin of schistosomes (Hirai et al. 2000; Lawton et
30
31 761 al. 2011). Our comparison of WEFs in selected schistosome species of different clades provide a first
32
33 762 hint for their potential contribution to karyotype variability, and thus speciation. This may include a
34
35 763 varying repertoire of WEF-encoded regulatory RNAs, which differs between these species.
36
37 764 Repetitive DNA discloses high sequence and copy number variability among and within species but
38
39 765 also in closely related organisms, which points to rapid evolution (Lower et al. 2018). This may involve
40
41 766 coevolution with regulatory RNAs as it was hypothesized for satellite repeats and long noncoding RNAs
42
43 767 (Lee et al. 2019) as well as the dual relationship between Alu repeats (short interspersed nuclear
44
45 768 elements of the human genome) and miRNAs. Duplication events involving Alu elements have
46
47 769 favoured the expansion of miRNA clusters and their expression (Lehnert et al. 2009).
48
49 770 From the data obtained in our study and against the background of recent literature, it is tempting to
50
51 771 speculate that more of the WE “junk-DNA” than expected might be functional and relevant. WEs of all
52
53 772 investigated WEFs exhibit a capricious incidence, , and they are transcribed in a stage-, sex-, pairing-,
54
55 773 gonad, and strain-specific or -preferential pattern. From exemplary findings of features typical for the
56
57 774 activity of mobile genetic elements, we hypothesize that W elements may have a mobile character.
58
59 775 Together with previous findings of intra-clonal recombination events of WEs (Grevelding 1999; Bayne
60
776 and Grevelding 2003), their presumptive role in sex chromosome emergence (Lepesant et al. 2012),
777 their putative capacity to express regulatory RNAs, we propose that W elements might influence the
778 biology of *S. mansoni*. Furthermore, based on the variable occurrence of WEFs in different schistosome
779 strains, isolates, and even species, we hypothesize that the W elements represent one of the sources
780 of heritable variability in the evolution of the family Schistosomatidae.
781

782 **Materials and Methods**

783 **Ethics Approval**

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

789

790 **Mapping and Characterizing WEF (V5) Against Genome Version V7**

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

801

802 **Mapping and Counting of RNA-Seq Reads**

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

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

1
2
3 817 *alignIntronMin 10*. Counts per gene were summarised with FeatureCounts v1.4.5-p1 (Liao et al. 2014)
4 818 based on the exon feature, using the annotation from WormBase Parasite WBPS14
5 819 (<https://parasite.wormbase.org/>).
6
7
8 820

9

821 **Open Reading Frame Analysis**

822 We used the NCBI program “ORFfinder” (<https://www.ncbi.nlm.nih.gov/orffinder/>) to detect
823 translatable sequence areas. For the analysis, we selected a minimum length of 30 nucleotides and
824 only the start codon “ATG”. The amino acid sequences were examined by BLASTp (BLAST+, v.2.6.0;
825 Camacho et al. 2009) against UniProtKB/Swiss-Prot.
826

826

827 **Differential Expression Analysis and Normalization**

828 To determine quantitative changes in WE transcript levels between the RNA-Seq datasets we
829 performed differential expression analysis using raw reads in DESeq2 (Love et al. 2014). Throughout
830 the manuscript, expression is defined as occurrence or levels of transcripts. For normalization of the
831 data sets, we combined WE-specific read counts and read counts of protein-coding genes (Smp
832 identifiers). For generation of the heatmap of the count matrix and the sample-distance matrix, we
833 used R (<https://www.r-project.org>, v3.6.3) package DESeq2 (v1.24.0) and pheatmap (v1.0.12), applying
834 log2 transformation of normalized counts of WEs using the *normTransform* and *rlogTransformation*
835 function, respectively.
836

836

837 **Identification of WEs on Autosomes**

838 We used Blast+ (v.2.6.0; Camacho et al. 2009) to identify highly similar sequences of WEs on
839 autosomes, applying the megablast task and a 0.001 E-value cut-off. We only extracted sequences if
840 the alignment covered more than 80% of the query sequence, and if the overall alignment percentage
841 (OAP; percent identity of the alignment multiplied by the coverage divided by 100) was higher than
842 60. Exceptions of this rule were made in case fragmented autosomal W elements were small; fragment
843 sizes below 25 bp were not considered. Using the sequence alignment tool MAFFT (v7.471;
844 <https://mafft.cbrc.jp/alignment/server/>), we calculated sequence homologies (Kuraku et al. 2013;
845 Katoh et al. 2019). As input, we used the Fasta sequences and chose “Adjust direction according to the
846 first sequence”. Clustal-formatted alignments were produced as results. To find patterns of WE mobile
847 activity, such as potential transposition events, we used MITE Tracker, an open source program that
848 provides positional information on inverted-repeat sequences and target site duplications (TSD)
849 (Crescente et al. 2018).
850

850

851

852 **Functional Prediction**

853 For the prediction of sequences coding functional RNA, we used StructRNAfinder, an integrative tool
854 allowing the identification, functional annotation, and taxonomic allocation of sequences to RNA
855 families by secondary structure inference
(<https://structrnafinder.integrativebioinformatics.me/run.html>) (Arias-Carrasco et al. 2018).
857 StructRNAfinder displays sequence consensus alignments for RNA families, according to Rfam
858 database (RNA families, data base version 14.2, Kalvari et al. 2018; <https://rfam.xfam.org>), but also
859 provides a taxonomic overview for each assigned functional RNA. As input, we generated FASTA files
860 with complete RNAs from the Bowtie2 output by using the tool "bedtools MergeBED" (Galaxy Version
861 2.27.1) (Quinlan and Hall 2010) to merge overlapping and adjacent regions. We chose the option
862 cmsearch and an E-value of 0.01. In the output options, we selected „report only best hit per
863 sequence“.

865 **Determination of Ribozyme Activity During *in vitro* Transcription**

866 We produced transcription templates containing the T7 promoter sequence by extension of partly
867 complementary oligonucleotides (Supplementary Table 6) using the Phusion DNA Polymerase
868 (ThermoScientific™). We purified templates by phenol/chloroform extraction and ethanol
869 precipitation. For *in vitro* transcription, 1 µg purified template, 1x transcription buffer, 3% DMSO, 4
870 mM NTPs, 0.013 U thermostable inorganic pyrophosphatase (NEB), and 25 ng of T7 RNA polymerase
871 (laboratory preparation) were combined in a 30 µl reaction and incubated at 37°C for 2 h. The
872 1x transcription buffer contained 80 mM HEPES-KOH (pH 7.5), 24 mM MgCl₂, 2 mM Spermidine, and
873 40 mM DTT. Labelling of HHR candidates during transcription occurred by the addition of [α^{32} P]-CTP.
874 We stopped reactions by adding RNA loading-dye composed of 2.5 mM Tris-HCl (pH 7.6), 20%
875 formamide, 0.06% bromophenol, and 0.06% xylene cyanol. A cleavage-deficient, elongated variant of
876 W25_415_3 (154 nt) was gel-purified after *in vitro* transcription, dephosphorylated using the antarctic
877 phosphatase (NEB) and labelled at its 5'-end using the T4 polynucleotidyl kinase (NEB) and [γ^{32} P]-ATP.
878 Following another gel-purification, we used the 5'-labeled RNA to create a size standard by alkaline
879 hydrolysis and RNase T1 digestion. For experimental validation and product visualisation, we separated
880 the size standard and internally labelled ribozyme cleavage products by denaturing 20%
881 polyacrylamide gel electrophoresis (PAGE) and detected the bands with an Amersham Typhoon Imager
882 (GE Healthcare).

884 **Data Availability**

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

1
2
3 887 schistosomes (Grevelding 1995; Protasio et al. 2012; Wang et al. 2013; Lu et al. 2016; Picard et al. 2016;
4 888 Lu et al. 2017; <https://www.ebi.ac.uk/ena>). Sample types, strains, replicates, accession numbers and
5
6 889 references are listed in Supplementary Table 1. Further sequence information, as indicated in the text,
7
8 890 was obtained from WormBase ParaSite (version 15, October 2020; <https://parasite.wormbase.org>),
9
10 891 BioProject numbers: PRJEA36577 (*S. mansoni*), *S. rodhaini* (PRJEB526 - Republic of Burundi),
11 892 PRJNA520774 (HuSjv2, *S. japonicum*), and PRJNA78265 (*S. haematobium*).
12

13 893

14 894 **Author Contributions**

15 895 M.S. and C.C. performed the bioinformatics analyses supported by advice and help from Z.L., J.B., and
16 896 A.G. J.O. and C.E.W. planned and performed the biochemical experiments with ribozymes and analysed
17 897 the data. C.G.G., C.G., and C.E.W. wrote the manuscript. C.G.G. and C.G. designed the study and
18 898 provided funding (FUGI). C.E.W. provided funds (DFG) for the ribozyme part.
19

20 899

21 900 **Acknowledgements**

22 901 We thank all the members of the Grevelding laboratory, particularly Simone Häberlein, Thomas Quack
23 902 and Oliver Weth for fruitful discussions. The work was supported by the Wellcome Trust FUGI grant
24 903 [107475/Z/15/Z to C.G.G., C.G.] and by the German Research Foundation [DFG; LU1889/4-1 to C.E.W.].
25 904 C.E.W. is also supported by a fellowship of the Peter and Traudl Engelhorn Foundation. This study is
26 905 set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP [ANR-10-LABX-41 to C.G.].
27 906 With the support of LabEx CeMEB, an ANR "Investissements d'avenir" program [ANR-10-LABX-04-01
28 907 to C.G.] through the Environmental Epigenomics Platform. The funders had no role in study design,
29 908 data collection and analysis, decision to publish, or preparation of the manuscript.
30

31 909

32 910 **Literature Cited**

- 33 911 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol*
34 912 *Biol.* 215:403–410.
- 35 913 Arias-Carrasco R, Vásquez-Morán Y, Nakaya HI, Maracaja-Coutinho V. 2018. StructRNAfinder: an
36 914 automated pipeline and web server for RNA families prediction. *BMC Bioinform.* 19:55. DOI:
37 915 10.1186/s12859-018-2052-2
- 38 916 Bartel DP. 2018. Metazoan microRNAs. *Cell* 173:20–51.
- 39 917
- 40 918 Basch P. 1991. In: Basch P. Editor. *Schistosomes: development, reproduction and host relations*. New
41 919 York, Oxford University Press, p. 1–248.
- 42 920
- 43 921 Bayne CJ, Grevelding CG. 2003. Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of
44 922 genetic heterogeneity among individuals within clones. *J Parasitol.* 89:1056–1060.
- 45 923
- 46 924 Berriman M, et al. 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460:352–358.
47 925

- 1
2
3 926 Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M. 2015a. Transcription of tandemly repetitive
4 927 DNA: functional roles. *Chromosome Res.* 23:463–477.
5 928
6 929 Biscotti MA, Olmo E, Heslop-Harrison JSP. 2015b. Repetitive DNA in eukaryotic genomes.
7 930 *Chromosome Res.* 23:415–420.
8 931
9 932 Boissier J, Durand P, Moné H. 2001. PCR effectiveness for sexing *Schistosoma mansoni* cercariae:
10 933 application for sexing clonal cercarial populations. *Mol Biochem Parasitol.* 112:139–141.
11 934
12 935 Braicu C, et al. 2019. Altered expression of miR-181 affects cell fate and targets drug resistance-
13 936 related mechanisms. *Mol Aspects Med.* 70:90–105.
14 937
15 938 Bratkovič T, Božič J, Rogelj B. 2020. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res.*
16 939 48: 1627–1651.
17 940
18 941 Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA
19 942 sequences have been incorporated into the genomes of higher organisms. *Science* 161:529–540.
20 943
21 944 Britton C, Laing R, Devaney E. 2020. Small RNAs in parasitic nematodes - forms and functions.
22 945 *Parasitology* 147:855–864.
23 946
24 947 Cai P, et al. 2016. Comprehensive transcriptome analysis of sex-biased expressed genes reveals
25 948 discrete biological and physiological features of male and female *Schistosoma japonicum*. *PLoS*
26 949 *Negl Trop Dis.* 10: e0004684. DOI: 10.1371/journal.pntd.0004684
27 950
28 951 Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI:
29 952 10.1186/1471-2105-10-421
30 953
31 954 Canny MD, et al. 2004. Fast cleavage kinetics of a natural hammerhead ribozyme. *J Am Chem Soc.*
32 955 126:1084810849.
33 956
34 957 Cardoso TCdS, et al. 2020. Computational prediction and characterisation of miRNAs and their
35 958 pathway genes in human schistosomiasis caused by *Schistosoma haematobium*. *Mem Inst*
36 959 *Oswaldo Cruz* 115:e190378.
37 960
38 961 Castagnone-Sereno P, Danchin EGJ. 2014. Parasitic success without sex – the nematode experience. *J*
39 962 *Evol Biol.* 27:1323–1333.
40 963
41 964 Cervera A, La Peña M de. 2014. Eukaryotic penelope-like retroelements encode hammerhead
42 965 ribozyme motifs. *Mol Biol Evol.* 31:2941–2947.
43 966
44 967 Chamberlin M, Ring J. 1973. Characterization of T7-specific ribonucleic acid polymerase. 1. General
45 968 properties of the enzymatic reaction and the template specificity of the enzyme. *J Biol Chem.*
46 969 248:2235–2244.
47 970
48 971 Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 251(4997):1030-1033.
49 972
50 973 Cheever AW, Macedonia JG, Mosimann JE, Cheever EA. 1994. Kinetics of egg production and egg
51 974 excretion by *Schistosoma mansoni* and *S. japonicum* in mice infected with a single pair of worms.
52 975 *Am J Trop Med Hyg.* 50:281–295.
53 976
54 977 Chevalier FD, Le Clec'h W, Alves de Mattos AC, LoVerde PT, Anderson TJC. 2016. Real-time PCR for
55 978 sexing *Schistosoma mansoni* cercariae. *Mol Biochem Parasitol.* 205:35–38.

- 1
2
3 979
4 980 Clough ER. 1981. Morphology and reproductive organs and oogenesis in bisexual and unisexual
5 981 transplants of mature *Schistosoma mansoni* females. *J Parasitol.* 67:535–539.
6 982
7 983 Coghlan A. 2005. *Nematode genome evolution*. WormBook: 1–15.
8 984
9 985 Colley DG, Bustinduy AL, Secor WE, King CH. 2014. Human schistosomiasis. *Lancet* 383:2253–2264.
10 986
11 987 Comings DE. 1972. The structure and function of chromatin. *Adv Hum Genet.* 3:237–431.
12 988
13 989 Copeland CS, et al. 2003. Boudicca, a retrovirus-like long terminal repeat retrotransposon from the
14 990 genome of the human blood fluke *Schistosoma mansoni*. *J Virol.* 77:6153–6166.
15 991
16 992 Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to
17 993 identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics*
18 994 19(1):348. DOI: 10.1186/s12859-018-2376-y
19 995
20 996 de la Peña M and Garcia-Robles I. 2010. Ubiquitous presence of the hammerhead ribozyme motif
21 997 along the tree of life. *RNA* 16:1943–1950.
22 998
23 999 de la Peña M, García-Robles I, Cervera A. 2017. The hammerhead ribozyme: a long history for a short
24 1000 RNA. *Molecules* 22(1):78.
25 1001
26 1002 DeMarco R, et al. 2004. Saci-1, -2, and -3 and Perere, four novel retrotransposons with high
27 1003 transcriptional activities from the human parasite *Schistosoma mansoni*. *J Virol.* 78:2967–2978.
28 1004
29 1005 DeMarco R, Machado AA, Bisson-Filho AW, Verjovski-Almeida S. 2005. Identification of 18 new
30 1006 transcribed retrotransposons in *Schistosoma mansoni*. *Biochem Biophys Res Commun.* 333:230–
31 1007 240.
32 1008
33 1009 Den Hollander JE, Erasmus DA. 1985. *Schistosoma mansoni*: male stimulation and DNA synthesis by
34 1010 the female. *Parasitology* 91:449–457.
35 1011
36 1012 Dennis PP, Omer A. 2005. Small non-coding RNAs in Archaea. *Curr Opin Microbiol.* 8:685–694.
37 1013
38 1014 Dobin A, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
39 1015
40 1016 Drew AC, Brindley PJ. 1995. Female-specific sequences isolated from *Schistosoma mansoni* by
41 1017 representational difference analysis. *Mol Biochem Parasitol.* 71:173–181.
42 1018
43 1019 Duan H, et al. 2018. The mir-279/996 cluster represses receptor tyrosine kinase signaling to
44 1020 determine cell fates in the *Drosophila* eye. *Development* 145(7):dev159053.
45 1021
46 1022 Erasmus DA. 1973. A comparative study of the reproductive system of mature, immature and
47 1023 "unisexual" female *Schistosoma mansoni*. *Parasitology* 67:165–183.
48 1024
49 1025 Ferbeyre G, Smith JM, Cedergren R. 1998. Schistosome satellite DNA encodes active hammerhead
50 1026 ribozymes. *Mol Cell Biol.* 18(7):3880–3888.
51 1027
52 1028 Fitzpatrick JM, et al. 2008. Use of genomic DNA as an indirect reference for identifying gender-
53 1029 associated transcripts in morphologically identical, but chromosomally distinct, *Schistosoma*
54 1030 *mansoni* cercariae. *PLoS Negl Trop Dis.* 2:e323. DOI: 10.1371/journal.pntd.0000323
55 1031

- 1
2
3 1032 Forster AC, Symons RH. 1987. Self-cleavage of plus and minus RNAs of a virusoid and a structural
4 1033 model for the active sites. *Cell* 49 211–220.
5 1034
6 1035 Galardi S, et al. 2002. Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation
7 1036 of target RNA in vitro. *Mol Cell Biol.* 22:6663–6668.
8 1037
9 1038 Gasser RB, Morahan G, Mitchell GF. 1991. Sexing single larval stages of *Schistosoma mansoni* by
10 1039 polymerase chain reaction. *Mol Biochem Parasitol.* 47:255–258.
11 1040
12 1041 Gibbs AJ, McIntyre GA. 1970. The diagram, a method for comparing sequences. Its use with amino
13 1042 acid and nucleotide sequences. *Eur J Biochem.* 16(1):1–11.
14 1043
15 1044 Grevelding CG. 1995. The female-specific W1 sequence of the Puerto Rican strain of *Schistosoma*
16 1045 *mansoni* occurs in both genders of a Liberian strain. *Mol Biochem Parasitol.* 71:269–272.
17 1046
18 1047 Grevelding CG. 1999. Genomic instability in *Schistosoma mansoni*. *Mol Biochem Parasitol.* 101:207–
19 1048 216.
20 1049
21 1050 Grevelding CG. 2004. *Schistosoma*. *Curr Biol.* 14:R545.
22 1051
23 1052 Grevelding CG, Langner S, Dissous C. 2018. Kinases: molecular stage directors for schistosome
24 1053 development and differentiation. *Trends Parasitol.* 34:246–260.
25 1054
26 1055 Hahnel S, et al. 2018. Tissue-specific transcriptome analyses provide new insights into GPCR signalling
27 1056 in adult *Schistosoma mansoni*. *PLoS Pathog.* 14:e1006718. DOI: 10.1371/journal.ppat.1006718
28 1057
29 1058 Hammann C, Luptak A, Perreault J, La Peña M de. 2012. The ubiquitous hammerhead ribozyme. *RNA*
30 1059 18:871–885.
31 1060
32 1061 Han M-J, Xu H-E, Zhang H-H, Feschotte C, Zhang Z. 2014. Spy: a new group of eukaryotic DNA
33 1062 transposons without target site duplications. *Genome Biol Evol.* 6:1748–1757.
34 1063
35 1064 Hartl M, et al. 2011. A new Prospero and microRNA-279 pathway restricts CO2 receptor neuron
36 1065 formation. *J Neurosci.* 31:15660–15673.
37 1066
38 1067 Hirai H, Spotila LD, LoVerde PT. 1989. *Schistosoma mansoni*: chromosomal localization of DNA repeat
39 1068 elements by in situ hybridization using biotinylated DNA probes. *Exp Parasitol.* 69:175–188.
40 1069
41 1070 Hirai H, et al. 2000. Chromosomal differentiation of the *Schistosoma japonicum* complex. *Int J*
42 1071 *Parasitol.* 30:441–452.
43 1072 Hombach S, Kretz M. 2016. Non-coding RNAs: classification, biology and functioning. *Adv Exp Med*
44 1073 *Biol.* 937:3–17.
45 1074
46 1075 Hotez PJ, Kamath A. 2009. Neglected tropical diseases in sub-saharan Africa: review of their
47 1076 prevalence, distribution, and disease burden. *PLoS Negl Trop Dis.* 3:e412. DOI:
48 1077 10.1371/journal.pntd.0000412
49 1078
50 1079 Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. WormBase ParaSite - a comprehensive
51 1080 resource for helminth genomics. *Mol Biochem Parasitol.* 215:2–10.
52 1081
53 1082 Jacinto DS, et al. 2011. Curupira-1 and Curupira-2, two novel Mutator-like DNA transposons from the
54 1083 genomes of human parasites *Schistosoma mansoni* and *Schistosoma japonicum*. *Parasitology*
55 1084 138:1124–1133.

- 1
2
3 1085
4 1086 Jaubert-Possamai S, Nouredine Y, Favery B. 2019. MicroRNAs, new players in the plant-nematode
5 1087 interaction. *Front Plant Sci.* 10:1180.
6 1088
7
8 1089 Jensen S, Gassama MP, Heidmann T. 1994. Retrotransposition of the *Drosophila* LINE I element can
9 1090 induce deletion in the target DNA: a simple model also accounting for the variability of the
10 1091 normally observed target site duplications. *Biochem Biophys Res Commun.* 202:111–119.
11 1092
12 1093 Jimenez RM, Delwart E, Lupták A. 2011. Structure-based search reveals hammerhead ribozymes in
13 1094 the human microbiome. *J Biol Chem.* 286:7737–7743.
14 1095
15 1096 Jimenez RM, Polanco JA, Lupták A. 2015. Chemistry and biology of self-cleaving ribozymes. *Trends*
16 1097 *Biochem Sci.* 40:648–661.
17 1098
18 1099 Jones JT, Kusel JR. 1989. Intra-specific variation in *Schistosoma mansoni*. *Parasitol Today (Regul Ed)*
20 1100 5:37–39.
21 1101
22 1102 Jourdane J, Theron A. 1980. *Schistosoma mansoni*: cloning by microsurgical transplantation of
23 1103 sporocysts. *Exp Parasitol.* 50:349–357.
24 1104
25 1105 Jourdane J, Theron A, Combes C. 1980. Demonstration of several sporocysts generations as a normal
26 1106 pattern of reproduction of *Schistosoma mansoni*. *Acta Trop.* 37:177–182.
27 1107
28 1108 Kalvari I, et al. 2018. Non-coding rna analysis using the rfam database. *Curr Protoc Bioinformatics*
30 1109 62:e51.
31 1110
32 1111 Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment,
33 1112 interactive sequence choice and visualization. *Brief Bioinform.* 20:1160–1166.
34 1113
35 1114 Khvorova A, Lescoute A, Westhof E, Jayasena SD. 2003. Sequence elements outside the hammerhead
36 1115 ribozyme catalytic core enable intracellular activity. *Nat Struct Biol.* 10:708–712.
37 1116
38 1117 Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose
40 1118 methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85:1077–1088.
41 1119
42 1120 Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function.
43 1121 *Nucleic Acids Res.* 47:D155-D162.
44 1122
45 1123 Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on
46 1124 genome scale. *Bioinformatics* 23:1026–1028.
47 1125
48 1125 Kunz W. 2001. Schistosome male-female interaction: induction of germ-cell differentiation. *Trends*
49 1126 *Parasitol.* 17:227–231.
50 1127
51 1128 Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of
52 1129 metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity.
53 1130 *Nucleic Acids Res.* 41:W22-8.
54 1131
55 1132 La Peña M de, García-Robles I. 2010. Ubiquitous presence of the hammerhead ribozyme motif along
56 1133 the tree of life. *RNA* 16:1943–1950.
57 1134
58 1135 La Peña M de, García-Robles I, Cervera A. 2017. The hammerhead ribozyme: a long history for a short
59 1136 RNA. *Molecules* 22.
60 1137

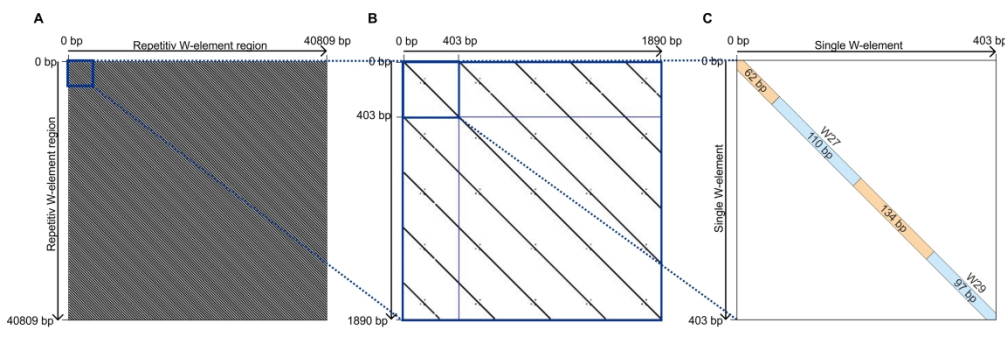
- 1
2
3 1138 Laha T, McManus DP, Loukas A, Brindley PJ. 2000. Salpha elements, short interspersed element-like
4 1139 retroposons bearing a hammerhead ribozyme motif from the genome of the oriental blood fluke
5 1140 *Schistosoma japonicum*. *Biochim Biophys Acta*. 1492:477–482.
6 1141
7 1142 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
8 1143
9 1144 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short
10 1145 DNA sequences to the human genome. *Genome Biol*. 10:R25.
11 1146
12 1147 Lanzer M, Fischer K, Le Blancq SM. 1995. Parasitism and chromosome dynamics in protozoan
13 1148 parasites: is there a connection? *Mol Biochem Parasitol*. 70(1-2):1-8.
14 1149
15 1150 Lawton SP, Hirai H, Ironside JE, Johnston DA, Rollinson D. 2011. Genomes and geography: genomic
16 1151 insights into the evolution and phylogeography of the genus *Schistosoma*. *Parasit Vectors* 4:131.
17 1152
18 1153 Lee H, Zhang Z, Krause HM. 2019. Long noncoding RNAs and repetitive elements: junk or intimate
19 1154 evolutionary partners? *Trends Genet*. 35:892–902.
20 1155
21 1156 Lehnert S, et al. 2009. Evidence for co-evolution between human microRNAs and Alu-repeats. *PLoS*
22 1157 *ONE* 4:e4456. DOI: 10.1371/journal.pone.0004456
23 1158
24 1159 Lepesant JM, et al. 2012a. Chromatin structural changes around satellite repeats on the female sex
25 1160 chromosome in *Schistosoma mansoni* and their possible role in sex chromosome emergence.
26 1161 *Genome Biol*. 13:R14.
27 1162
28 1163 Lepesant JM, et al. 2012b. Combination of de novo assembly of massive sequencing reads with
29 1164 classical repeat prediction improves identification of repetitive sequences in *Schistosoma*
30 1165 *mansoni*. *Exp Parasitol*. 130:470–474.
31 1166
32 1167 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning
33 1168 sequence reads to genomic features. *Bioinformatics* 30:923–930.
34 1169
35 1170 Lilley DMJ. 2019. Classification of the nucleolytic ribozymes based upon catalytic mechanism.
36 1171 *F1000Res* 8. DOI: 10.12688/f1000research.19324.1
37 1172
38 1173 Llorà-Batlle O, Tintó-Font E, Cortés A. 2019. Transcriptional variation in malaria parasites: why and
39 1174 how. *Brief Funct Genomics* 18:329–341.
40 1175
41 1176 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq
42 1177 data with DESeq2. *Genome Biol*. 15(12):550.
43 1178
44 1179 Lower SS, McGurk MP, Clark AG, Barbash DA. 2018. Satellite DNA evolution: old ideas, new
45 1180 approaches. *Curr Opin Genet Dev*. 49:70–78.
46 1181
47 1182 Lu Z, Spaenig S, Weth O, Grevelding CG. 2019. Males, the wrongly neglected partners of the
48 1183 biologically unprecedented male-female interaction of schistosomes. *Front Genet*. 10:796.
49 1184
50 1185 Lu Z, et al. 2016. Schistosome sex matters: a deep view into gonad-specific and pairing-dependent
51 1186 transcriptomes reveals a complex gender interplay. *Sci Rep*. 6:31150.
52 1187
53 1188 Lu Z, et al. 2017. A gene expression atlas of adult *Schistosoma mansoni* and their gonads. *Sci Data*
54 1189 4:170118. DOI: 10.1038/sdata.2017.118

- 1
2
3 1190 Lu Z, Zhang Y, Berriman M. 2018. A web portal for gene expression across all life stages of
4 1191 *Schistosoma mansoni*. bioRxiv:PPR13955. DOI: <https://doi.org/10.1101/308213>
5 1192
6 1193 Lui L, Lowe T. 2013. Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays*
7 1194 *Biochem.* 54:53–77.
8 1195
9 1196 Lünse CE, Weinberg Z, Breaker RR. 2017. Numerous small hammerhead ribozyme variants associated
10 1197 with Penelope-like retrotransposons cleave RNA as dimers. *RNA Biol.* 14:1499–1507.
11 1198
12 1199 Marais G, Galtier N. 2003. Sex chromosomes: how X-Y recombination stops. *Curr Biol.* 13(16):R641–
13 1200 R643.
14 1201
15 1202 Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for
16 1203 catalysis. *Cell* 126:309–320.
17 1204
18 1205 Mattick JS, Makunin IV. 2006. Non-coding RNA. *Hum Mol Genet.* 15:R17-29.
19 1206
20 1207 Meštrović N, et al. 2015. Structural and functional liaisons between transposable elements and
21 1208 satellite DNAs. *Chromosome Res.* 23:583–596.
22 1209
23 1210 Meunier J, et al. 2013. Birth and expression evolution of mammalian microRNA genes. *Genome Res.*
24 1211 23:34–45.
25 1212
26 1213 Moore DV, Sandground JH. 1956. The relative egg producing capacity of *Schistosoma mansoni* and
27 1214 *Schistosoma japonicum*. *Am J Trop Med Hyg.* 5:831–840.
28 1215
29 1215 Moné H, Boissier J. 2004. Sexual biology of schistosomes. *Adv Parasitol.* 75:89-189.
30
31
32 1216 Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*
33 1217 29(22):2933-2935.
34 1218
35 1219 Neves RH, et al. 2005. A new description of the reproductive system of *Schistosoma mansoni*
36 1220 (Trematoda: Schistosomatidae) analyzed by confocal laser scanning microscopy. *Parasitol Res.*
37 1221 95:43–49.
38 1222
39 1223 Niwa R, Slack FJ. 2007. The evolution of animal microRNA function. *Curr Opin Genet Dev.* 17:145–
40 1224 150.
41 1225
42 1226 Ohno S. 1967. *Sex chromosomes and sex-linked genes*. Springer-Verlag, Berlin Heidelberg.
43 1227
44 1228 Ohno S. 1999. Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. *Semin*
45 1229 *Cell Dev Biol.* 10:517–522.
46 1230
47 1231 Ojha S, Malla S, Lyons SM. 2020. snoRNPs: functions in ribosome biogenesis. *Biomolecules* 10(5):783.
48 1232
49 1233 Olveda DU, et al. 2014. The chronic enteropathogenic disease schistosomiasis. *Int J Infect Dis.*
50 1234 28:193–203.
51 1235
52 1236 Pélisson A, Sarot E, Payen-Groschêne G, Bucheton A. 2007. A novel repeat-associated small
53 1237 interfering RNA-mediated silencing pathway downregulates complementary sense gypsy
54 1238 transcripts in somatic cells of the *Drosophila* ovary. *J Virol.* 81:1951–1960.
55 1239
56
57
58
59
60

- 1
2
3 1240 Perreault J, et al. 2011. Identification of hammerhead ribozymes in all domains of life reveals novel
4 1241 structural variations. PLoS Comput Biol. 7:e1002031. DOI: 10.1371/journal.pcbi.1002031
5 1242
6 1243 Picard MAL, et al. 2016. Sex-biased transcriptome of *Schistosoma mansoni*: host-parasite interaction,
7 1244 genetic determinants and epigenetic regulators are associated with sexual differentiation. PLoS
8 1245 Negl Trop Dis. 10:e0004930. DOI: 10.1371/journal.pntd.0004930
9 1246
10 1247 Popiel I, Basch PF. 1984. Reproductive development of female *Schistosoma mansoni* (Digenea:
11 1248 Schistosomatidae) following bisexual pairing of worms and worm segments. J Exp Zool. 232:141–
12 1249 150.
13 1250
14 1251 Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G. 1986. Autolytic processing of dimeric
15 1252 plant virus satellite RNA. Science 231:1577–1580.
16 1253
17 1254 Protasio AV, et al. 2012. A systematically improved high quality genome and transcriptome of the
18 1255 human blood fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 6:e1455. DOI:
19 1256 10.1371/journal.pntd.0001455
20 1257
21 1258 Quack T, Doenhoff M, Kunz W, Grevelding CG. 1998. *Schistosoma mansoni*: the varying occurrence of
22 1259 repetitive elements in different strains shows sex-specific polymorphisms. Exp Parasitol. 89:222–
23 1260 227.
24 1261
25 1262 Queiroz FR, et al. 2017. Differential expression of small RNA pathway genes associated with the
26 1263 *Biomphalaria glabrata/Schistosoma mansoni* interaction. PLoS ONE 12:e0181483. DOI:
27 1264 10.1371/journal.pone.0181483
28 1265
29 1266 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
30 1267 Bioinformatics 26:841842.
31 1268
32 1269 Ruffner DE, Stormo GD, Uhlenbeck OC. 1990. Sequence requirements of the hammerhead RNA self-
33 1270 cleavage reaction. Biochemistry 29:10695–10702.
34 1271
35 1272 Satović E, Vojvoda Zeljko T, Luchetti A, Mantovani B, Plohl M. 2016. Adjacent sequences disclose
36 1273 potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network
37 1274 with transposable elements. BMC Genomics 17:997. DOI: 10.1186/s12864-016-3347-1
38 1275
39 1276 Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. Mol Ecol.
40 1277 28:1537–1549.
41 1278
42 1279 Scott MS, Ono M. 2011. From snoRNA to miRNA: Dual function regulatory non-coding RNAs.
43 1280 Biochimie 93:1987–1992.
44 1281 Seehafer C, Kalweit A, Steger G, Gräf S, Hamann C. 2011. From alpaca to zebrafish: hammerhead
45 1282 ribozymes wherever you look. RNA 17:21–26.
46 1283
47 1284 Shaw MK. 1987. *Schistosoma mansoni*: vitelline gland development in females from single sex
48 1285 infections. J Helminthol. 61:253–259.
49 1286
50 1287 Spotila LD, Hirai H, Rekosh DM, Lo Verde PT. 1989. A retroposon-like short repetitive DNA element in
51 1288 the genome of the human blood fluke, *Schistosoma mansoni*. Chromosoma 97:421–428.
52 1289
53 1290 Spotila LD, Rekosh DM, Boucher JM, LoVerde PT. 1987. A cloned DNA probe identifies the sex of
54 1291 *Schistosoma mansoni* cercariae. Mol Biochem Parasitol. 26:17–20.
55 1292

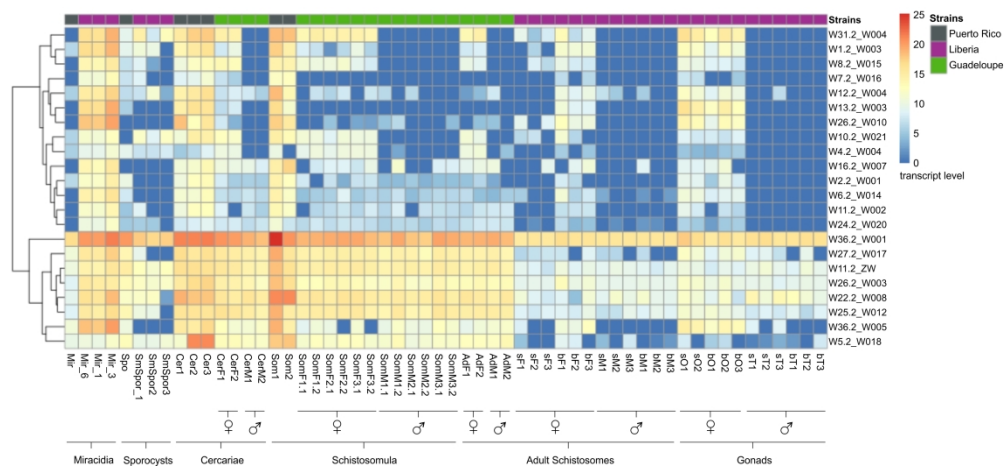
- 1
2
3 1293 Stroehlein AJ, et al. 2018. The small RNA complement of adult *Schistosoma haematobium*. PLoS Negl
4 1294 Trop Dis. 12:e0006535. DOI: 10.1371/journal.pntd.0006535
5 1295
6 1296 Thakur J, Packiaraj J, Henikoff S. 2021. Sequence, chromatin and evolution of satellite DNA. Int J Mol
7 1297 Sci. 22(9):4309. DOI: 10.3390/ijms22094309.
8 1298
9 1299 Uhlenbeck OC. 2003. Less isn't always more. *RNA* 9:1415–1417.
10 1300
11 1301 Venancio TM, Wilson RA, Verjovski-Almeida S, DeMarco R. 2010. Bursts of transposition from non-
12 1302 long terminal repeat retrotransposon families of the RTE clade in *Schistosoma mansoni*. Int J
13 1303 Parasitol. 40:743–749.
14 1304
15 1305 Wang B, Collins JJ, Newmark PA. 2013. Functional genomic characterization of neoblast-like stem
16 1306 cells in larval *Schistosoma mansoni*. *Elife* 2:e00768. DOI: 10.7554/eLife.00768
17 1307
18 1308 Weber MJ. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet. 2:e205.
19 1309 DOI: 10.1371/journal.pgen.0020205
20 1310
21 1311 Webster P, Mansour TE, Bieber D. 1989. Isolation of a female-specific, highly repeated *Schistosoma*
22 1312 *mansoni* DNA probe and its use in an assay of cercarial sex. Mol Biochem Parasitol. 36:217–222.
23 1313
24 1314 Weinberg CE, Weinberg Z, Hammann C. 2019. Novel ribozymes: discovery, catalytic mechanisms, and
25 1315 the quest to understand biological function. *Nucleic Acids Res.* 47:9480–9494.
26 1316
27 1317 Weinberg Z, Breaker RR. 2011. R2R--software to speed the depiction of aesthetic consensus RNA
28 1318 secondary structures. *BMC Bioinformatics* 12:3. DOI: 10.1186/1471-2105-12-3
29 1319
30 1320 Weng H, Lal K, Yang FF, Chen J. 2015. The pathological role and prognostic impact of miR-181 in
31 1321 acute myeloid leukemia. *Cancer Genet.* 208:225–229.
32 1322
33 1323 Woodhouse MR, Pedersen B, Freeling M. 2010. Transposed genes in *Arabidopsis* are often associated
34 1324 with flanking repeats. PLoS Genet. 6(5):e1000949. DOI: 10.1371/journal.pgen.1000949
35 1325
36 1326 Yu J, Yu Y, Li Q, Chen M, Shen H, Zhang R, Song M, Hu W. 2019. Comprehensive analysis of miRNA
37 1327 profiles reveals the role of *Schistosoma japonicum* miRNAs at different developmental stages.
38 1328 *Vet Res.* 50:23.
39 1329
40 1330 Zhao J-H, Guo H-S. 2019. Trans-kingdom RNA interactions drive the evolutionary arms race between
41 1331 hosts and pathogens. *Curr Opin Genet Dev.* 58–59:62–69.
42 1332
43 1333 Zhu L, Liu J, Cheng G. 2014. Role of microRNAs in schistosomes and schistosomiasis. *Front Cell Infect*
44 1334 *Microbiol.* 4:165. DOI: 10.3389/fcimb.2014.00165
45 1335
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



The New Definition of WEF W27.2
316x100mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/gbe/advance-article/doi/10.1093/gbe/evab204/6361599 by guest on 03 September 2021

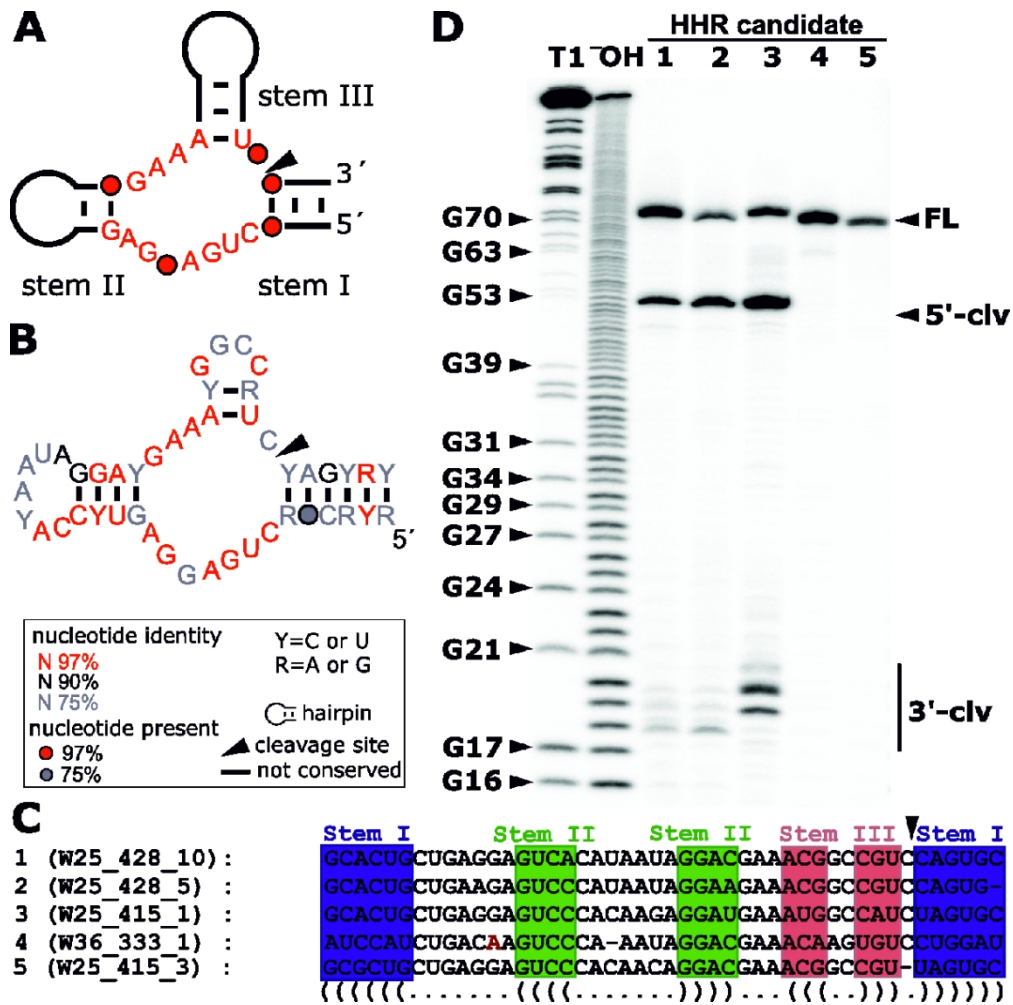


WE Transcript Profiles Across Different *S. mansoni* Strains, Life Stages, Sexes, and Gonads

353x162mm (300 x 300 DPI)

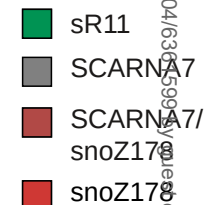
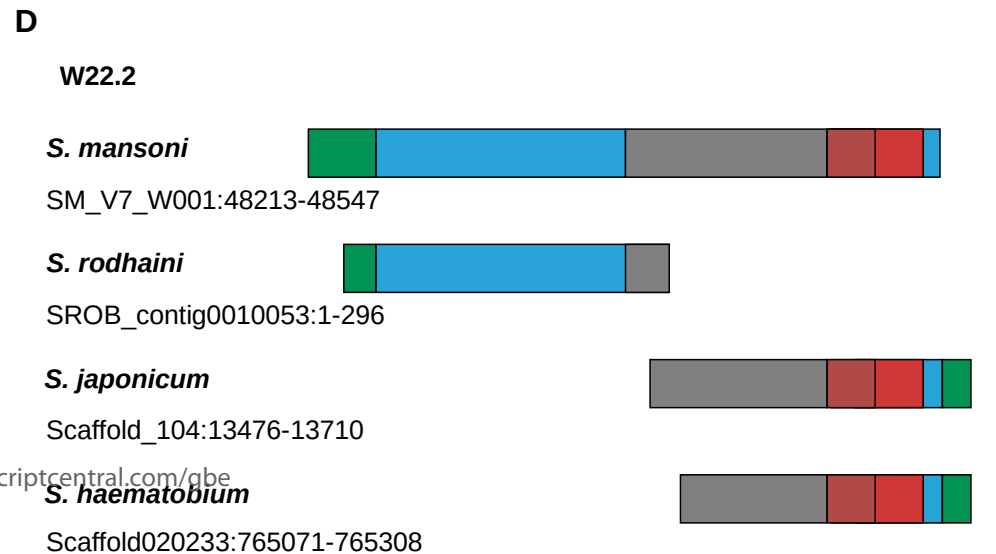
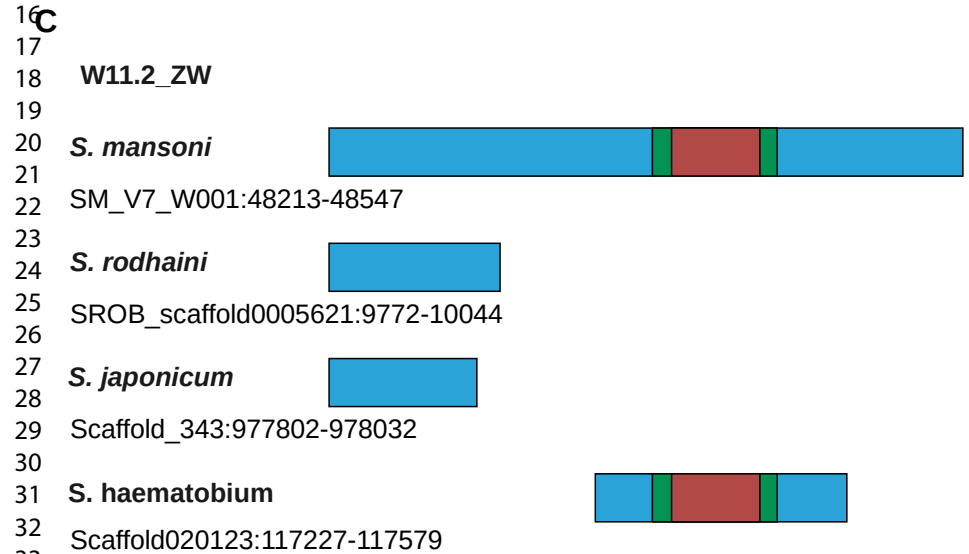
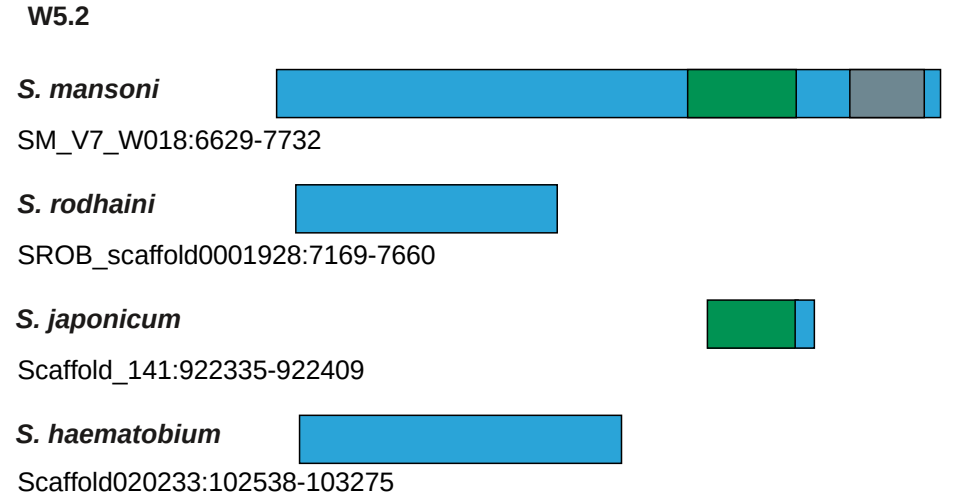
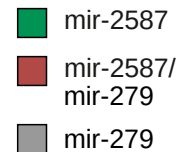
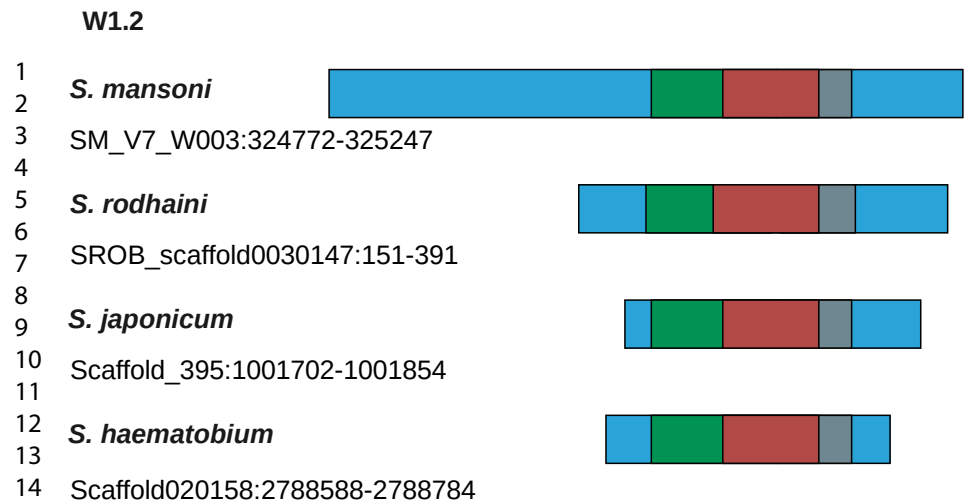
1
2 Smalphi atcgcaagcaagtggct-----atcaggactcagtgccgagtgataacgcgat
3 W36.2_335bp attgcacaagcaagtggct-----atcaggactcagtggtcgagtgataacgcgat
4 W25.2_415bp gttggaatagacataaacaccatgggatgccgtctcagtggtgtatgtgttgagtgctc
5 W26.2_400_bp gttgaagatggaatgaacacggttggatgccgactcagtggttcttgatgttctcagctt
6 W27.2_403bp tttgaagtaaggcattgagagcgttggatgggtgaccagtggtctagtgtttaagtgtc
7 * . * * . . . * ** * * . * * * * . . . *
8
9 Smalphi ggtgtttgaagcgaggg-tactgggttggagtcccagagtgatcaactctgagatgc
10 W36.2_335bp ggcgtttgaagcgaaagtactgggttcgagtgccagaatggacatcaactctgagatgc
11 W25.2_415bp tattgagaaagtgttattcctgtgttggaaatctgg-----tgaggcgaaatcgt
12 W26.2_400_bp ggcgcagaaaatgataagcgcagtttctcgaatctgg-----agaggcggtatcgt
13 W27.2_403bp acgcgcgagactgataggtgctgggttggaaatctcg-----cgagactgcatcgt
14 ..* * * ** ** * * * . . . * .
15
16 Smalphi aggtacatccagctgacgagtgcca-aataggacgaaacgcgcgctcctggattccactgc
17 W36.2_335bp aggtacatccagctgacgagtgcca-aataggacgaaacgcgcgctcctccattccactgc
18 W25.2_415bp ggattggaatgctgaggagtcgcacaatagaaccaaagaccgctccagtggttccagga
19 W26.2_400_bp ggatgcacactgctgagaaagatattaatgaagtccaaggtgctcag-attgtcgagg
20 W27.2_403bp ggatgcgactgctgagcagtgccacagtagcagcgaacggccatccagtgcttccaggt
21 . * . * . . . * * * * * * * * . . . * .
22
23 Smalphi tatccactattcatc-----tttgcttatcatgcttgtgaatcaaggctatatcgag
24 W36.2_335bp tagccactatccatc-----tttgcttaccatgcttgtcaatttaggctatatcgag
25 W25.2_415bp tttccctagtggtctaccttcaatccactcatgatttcaata-----taaata-cact
26 W26.2_400_bp ttaacatgacggtctctcaacaattgattcatcatcataacca-----tcaacattacg
27 W27.2_403bp tttccatcgtgctccagcttcaattcattcatcatctcaacta-----tcaacattact
28 * * * * * . . . * * * . . .
29
30 Smalphi gcaatacgcacagt-----atgcacata-tgccaattagagactgacc
31 W36.2_335bp gcaatacgcacagt-----atgcacaca-tgccaattagagactgacc
32 W25.2_415bp caaatctccacaaatgccactcctcataataataataataataatcaaatgctcacc
33 W26.2_400_bp agattatccacaaa-accacttctcatattagtcaacacg-----agctcact
34 W27.2_403bp ataatatctacgaa-acgccttctcatcactgtcaacatt-----tgctcaac
35 * * . . * * * * * * * . . .
36
37 Smalphi agttgcagtcct-aacacat-----cgatgggaagattcaaacaacaata
38 W36.2_335bp agttgctgtcctaaaaacat-----caatgggaagatcaaacaacaata
39 W25.2_415bp aatgactgaattgaagagatatttctgaggagttgtagtgagaagcagtgaccagtgagg
40 W26.2_400_bp agtgactgatttcaacaggtatttctgaggagttctggtgagaagcagagaccagtgag
41 W27.2_403bp agtgactggcttgaagaggcatttctgaggagttctggtgagaagcagtgaccagtgagg
42 * . * * * . * * * * * * * * * . * * * . *
43
44 Smalphi ctaagt-----
45 W36.2_335bp ctaagta-----aatttcaact-----
46 W25.2_415bp tgaagccacgtctgttgtgagatgtcaactgactgaagacaattgtgaacg-gttggtga
47 W26.2_400_bp tgaagcggtttgttttgaataat---ctcactgaagacattggtggatgtgtcgctga
48 W27.2_403bp ggaagcag-gtgtgatgtgagatggtcactcaatgaagacattggtggatgtgtcggtca
49 *** .
50
51 Smalphi -----
52 W36.2_335bp -----tcacccc
53 W25.2_415bp acttgggtgattg
54 W26.2_400_bp atatcgtggatcg
55 W27.2_403bp tttcagtgatcg

Downloaded from https://academic.oup.com/gbe/advance-article/doi/10.1093/gbe/evab204/6361599 by guest on 03 September 2021



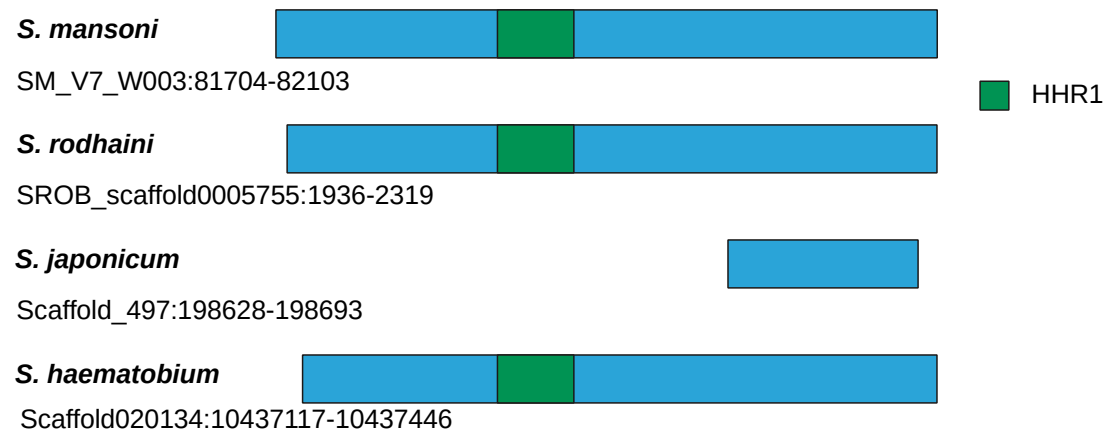
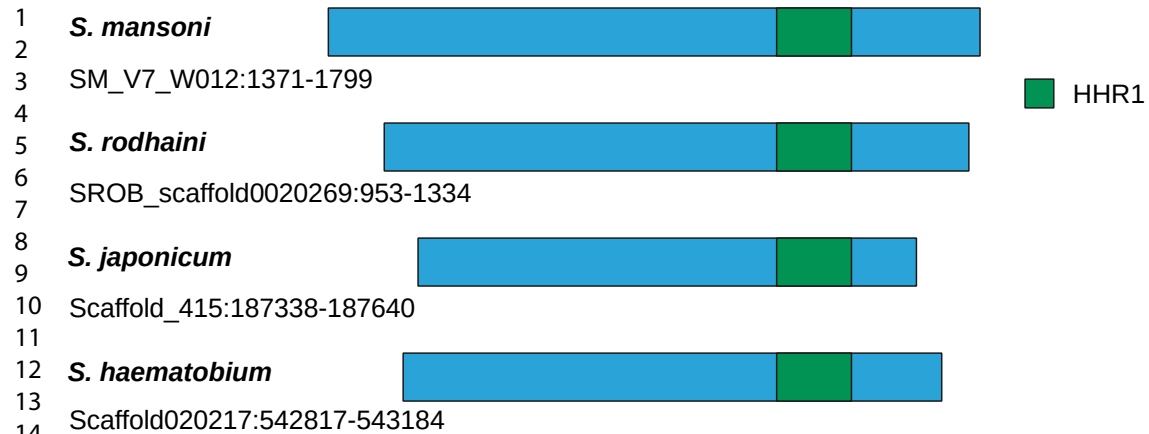
Active Hammerhead Ribozymes are Part of Autosomal WEs

Downloaded from https://academic.oup.com/gbe/advance-article/doi/10.1093/gbe/evab204/6361599 by guest on 03 September 2021



W25.2

W26.2

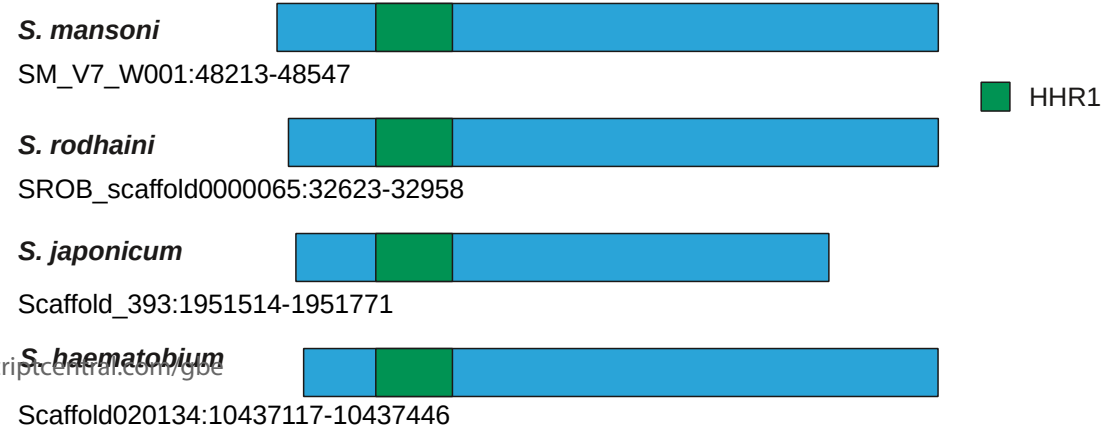
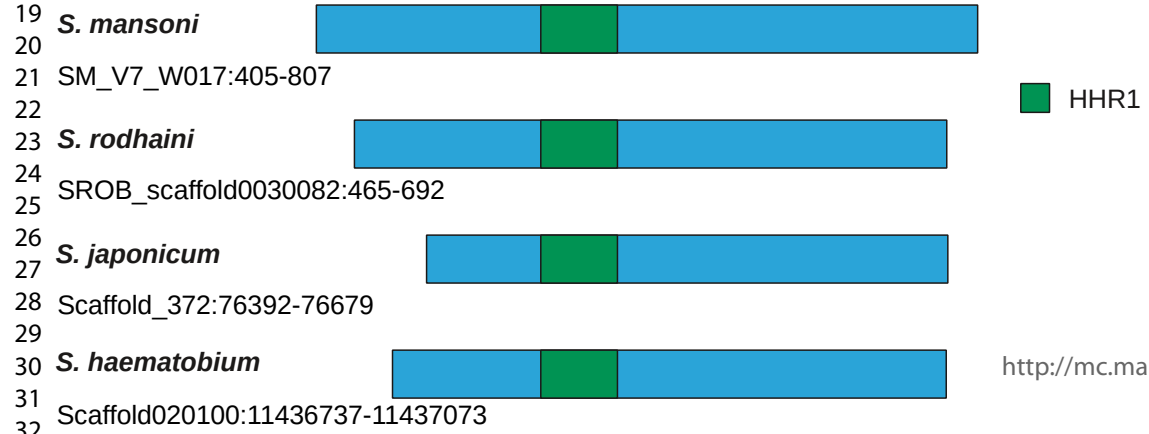


C

W27.2

D

W36.2



<http://mc.manuscriptcentral.com/gbe>

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)