



**HAL**  
open science

## Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination

Marwan Elkrewi, Mikhail A Moldovan, Marion a L Picard, Beatriz Vicoso

### ► To cite this version:

Marwan Elkrewi, Mikhail A Moldovan, Marion a L Picard, Beatriz Vicoso. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. *Molecular Biology and Evolution*, 2021, 10.1093/molbev/msab178 . hal-03269538

**HAL Id: hal-03269538**

<https://hal.sorbonne-universite.fr/hal-03269538v1>

Submitted on 24 Jun 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.

1  
2 **Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution**  
3 **and suggest candidate for sex determination**

4 Marwan Elkrewi 1<sup>†</sup>, Mikhail A. Moldovan 1.2<sup>†</sup>, Marion A.L. Picard 1.3\*, Beatriz Vicoso 1\*

5 1. Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg, 3400,  
6 Austria. / 2. Skolkovo Institute of Science and Technology, Moscow, Russia. / 3. Sorbonne  
7 Université, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire  
8 Océanologique, Banyuls-sur-Mer, France.

9 **\*Corresponding authors:** E-mails: [beatriz.vicoso@ist.ac.at](mailto:beatriz.vicoso@ist.ac.at); [marion.picard@obs-banyuls.fr](mailto:marion.picard@obs-banyuls.fr)

10 *Author notes:* <sup>†</sup>These authors contributed equally to this work and should be considered co-  
11 first authors. \*These authors contributed equally to this work and should be considered co-  
12 last authors.

13 **Abstract**

14 Schistosomes, the human parasites responsible for snail fever, are female-heterogametic.  
15 Different parts of their ZW sex chromosomes have stopped recombining in distinct lineages,  
16 creating “evolutionary strata” of various ages. While the Z-chromosome is well characterized  
17 at the genomic and molecular level, the W-chromosome has remained largely unstudied from  
18 an evolutionary perspective, as only a few W-linked genes have been detected outside of the  
19 model species *Schistosoma mansoni*. Here, we characterize the gene content and evolution of  
20 the W-chromosomes of *S. mansoni* and of the divergent species *S. japonicum*. We use a  
21 combined RNA/DNA k-mer based pipeline to assemble around one hundred candidate W-  
22 specific transcripts in each of the species. About half of them map to known protein coding  
23 genes, the majority homologous to *S. mansoni* Z-linked genes. We perform an extended  
24 analysis of the evolutionary strata present in the two species (including characterizing a  
25 previously undetected young stratum in *S. japonicum*) to infer patterns of sequence and  
26 expression evolution of W-linked genes at different time points after recombination was lost.  
27 W-linked genes show evidence of degeneration, including high rates of protein evolution and  
28 reduced expression. Most are found in young lineage-specific strata, with only a few high  
29 expression ancestral W-genes remaining, consistent with the progressive erosion of non-  
30 recombining regions. Among these, the splicing factor U2AF2 stands out as a promising

31 candidate for primary sex determination, opening new avenues for understanding the  
32 molecular basis of the reproductive biology of this group.

### 33 **1. Introduction**

34 Separate sexes are frequently determined by a specialized pair of sex chromosomes  
35 (Charlesworth et al. 2005) called X and Y in species where males are heterogametic (e.g. in  
36 mammals), and Z and W when females comprise the heterogametic sex (e.g. in birds) (Bull  
37 1983). Sex chromosomes arise from chromosomes containing sex-determining genes when  
38 parts of the sex-specific chromosome lose the ability to recombine (Ohno 1967; Nei 1969).  
39 Inefficient selection on this newly non-recombining Y or W chromosomal region results in  
40 the accumulation of repetitive sequences and deleterious mutations, eventually leading to  
41 extensive gene loss (Bachtrog 2013). Loss of recombination can progressively spread to  
42 larger sections of the chromosome, yielding “evolutionary strata” that have started  
43 degenerating at different time points. Ancient Y/W chromosomes, such as those of mammals  
44 and birds, are typically gene-poor and heterochromatic. Some genes, however, such as those  
45 responsible for sex-determination or sexual differentiation, as well as dosage-sensitive genes,  
46 can be preserved for large periods of time (Lahn et al. 2001; Charlesworth et al. 2005). Thus,  
47 identifying genes on these chromosomes can be an important step towards understanding the  
48 mechanisms underlying differences between the sexes (Lahn et al. 2001; Bellott et al. 2014).

49 Worms of the trematode genus *Schistosoma* cause schistosomiasis -- one of the  
50 deadliest neglected tropical diseases, affecting hundreds of thousands of people in tropical  
51 regions (Centers for Disease Control and Prevention 2011). Many of the clinical symptoms of  
52 schistosomiasis, as well as the spreading of the parasite itself, are due to the massive egg  
53 production during the life-long mating between male and female worms (Kunz 2001;  
54 LoVerde et al. 2004), fueling a long-standing interest in their reproductive biology.  
55 Schistosomes are the only trematodes with separate sexes, which are determined by a pair of  
56 ZW sex chromosomes (Grossman et al. 1981; Lawton et al. 2011). Although homologous  
57 chromosomes correspond to the ZW pair in different schistosome lineages, there are  
58 substantial differences in the gene content of the Z-specific region of Asian (*Schistosoma*  
59 *japonicum*) and African (*S. mansoni* and *S. haematobium*) lineages, suggesting that  
60 recombination was lost between much of the sex chromosomes independently in the two  
61 clades (Picard et al. 2018).

62 Much less is known about the evolution and current gene content of the W-  
63 chromosome, and it is still unclear whether this chromosome plays a role in sex  
64 determination or differentiation. The latest version of the *S. mansoni* genome contains several  
65 large W-scaffolds which harbor 32 W-linked genes (Howe et al. 2016; Howe et al. 2017;  
66 <https://parasite.wormbase.org/index.html>), a much smaller number than the many hundreds  
67 annotated on the Z chromosome (Protasio et al. 2012; Picard et al. 2018), consistent with  
68 widespread genetic degeneration. However, assembling W-derived sequences is difficult due  
69 to their heterochromatic and repetitive nature, and some W-genes may remain  
70 uncharacterized. Furthermore, only three W-linked genes have been identified in the Asian *S.*  
71 *japonicum* (Liu et al. 2020) and none in other species for which draft genomes are available.  
72 This is an important gap, as any gene involved in sex-determination is likely to be shared  
73 between different lineages, providing an important strategy for pinpointing promising  
74 candidates. Finally, the evolutionary history of the ZW pair in this clade, in which loss of  
75 recombination has occurred at different times in the two major lineages, offers a window into  
76 the temporal dynamics of degeneration of these non-recombining W-chromosomes. In  
77 particular, a large section of the ZW pair, the “S0” stratum, is thought to have stopped  
78 recombining in the ancestor of all schistosomes (Picard et al. 2018). Two younger strata were  
79 formed independently in *S. mansoni* (“S1man”) and in *S. japonicum* (“S1jap”). In which of  
80 these strata W-linked genes are located, and whether they differ in their rates of evolution or  
81 patterns of expression depending on how long they have been non-recombining, is still  
82 unknown.

83 Several studies have demonstrated that W-derived transcripts can be efficiently  
84 recovered by combining male and female DNA and RNA sequencing data (Cortez et al.  
85 2014; Moghadam et al. 2012; Mahajan and Bachtrog 2017; Li et al. 2018). We perform the  
86 first such systematic characterization and comparison of W-derived transcripts in the  
87 divergent species *S. mansoni* and *S. japonicum*. We combine genomic and RNA-seq data  
88 spanning much of the parasite life cycle to detect dozens of candidate W-derived transcripts  
89 in each species, and characterize both their evolutionary history and patterns of expression  
90 throughout development. We discuss the relevance of these results to schistosome sexual  
91 differentiation, and to the evolution of ZW chromosomes in this group.

## 92 **2. Results**

93 **2.1 Newly identified W-linked genes in *S. mansoni* and *S. japonicum* map primarily to**  
94 **the Z**

95 We applied a k-mer based pipeline to assemble female-specific transcripts (Methods  
96 and Sup. Figure 1). Similar approaches, in which male and female genomic reads are broken  
97 into shorter segments (k-mers), and k-mers found in only one sex are used to identify Y/W  
98 sequences, have been successfully applied to various organisms (Palmer et al. 2019). Our  
99 pipeline extends these by calling female-specific k-mers only if they are found in both DNA  
100 (one library per sex in each species) and RNA data (for each sex: one RNA-seq sample  
101 obtained by merging reads derived from three developmental stages for *S. japonicum*, and  
102 two samples merged from four stages for *S. mansoni*, see Methods and Sup. Table 1) and  
103 using them to output putative W-derived RNA-seq reads directly. Briefly, for each species,  
104 we selected k-mers that were found in all female DNA and RNA samples but in none of the  
105 male samples. RNA-seq reads containing these female-specific k-mers were extracted and  
106 assembled into putative W-transcripts. Male and female DNA reads were further mapped to  
107 putative W-transcripts longer than 200 base pairs (bp), and only transcripts with a high  
108 number of reads mapping perfectly in females but not in males were kept in our set of  
109 candidates (Sup. Figures 2 and 3; specific steps to improve the assembly in *S. japonicum* are  
110 described in the methods).

111 We used BLAT (Kent 2002) to map our candidates to the gene models (CDS) of the  
112 *S. mansoni* genome (v7, Sup. Table 2) in order to assess the efficacy of our pipeline (as the *S.*  
113 *japonicum* genome is not assembled at the chromosome level): of the 86 *S. mansoni*  
114 candidate W transcripts (Sup. Dataset 1), 37 mapped to known protein coding genes (Table  
115 1.A). The majority of these (24) mapped primarily to annotated W-linked genes in the current  
116 assembly (v7, obtained from female and male DNA), confirming the validity of our approach  
117 for detecting female-specific protein-coding sequences. Another 9 mapped to Z-linked genes,  
118 and likely represent true W-genes which are missing from the current assembly. Finally, 4  
119 mapped to genes in other chromosomal locations; these may represent W-linked genes that  
120 do not have a Z-homolog, or false positives. For the rest of our analyses, we combined our *S.*  
121 *mansoni* W-candidates with the annotated W-genes in this species (when a gene was found in  
122 both sets, only the longest transcript was kept) (Table 1.B, Sup. Dataset 2), yielding a  
123 “combined” set of candidates of 90 transcripts, 42 of them protein-coding. A similar number  
124 of *S. japonicum* W-candidates mapped to *S. mansoni* coding sequences (48 out of 94), all of

125 which did so to known Z-linked genes, again confirming the efficacy of our pipeline for  
126 detecting W-linked protein coding genes (Table 1.D, Sup. Dataset 3).

127 In order to investigate the evolutionary history of these W-derived coding sequences,  
128 we further extracted their closest homologs in the genome. For *S. mansoni*, we remapped the  
129 candidate W-transcripts to the CDS set, after excluding annotated W-genes. We retrieved a  
130 close homolog for 40 out of the 42 protein-coding transcripts; 34 of them mapped to a  
131 homolog on the Z-chromosome, as expected if W and Z-linked genes share a close ancestry  
132 (Table 1.C). In the case of *S. japonicum*, we wanted to avoid possible ZW hybrid assemblies  
133 that may be present in the published genome. We therefore extracted the BLAT best-hit of  
134 each protein-coding W-transcript from a male-derived transcriptome. The final list of protein  
135 coding W-candidates in the two species, as well as their respective homologs, is provided in  
136 Sup. Dataset 4. For the rest of the analysis, we focused on ZW homolog pairs (the 34 *S.*  
137 *mansoni* W-candidates with a Z-linked homolog, and the 48 *S. japonicum* W-candidates that  
138 mapped to Z-linked genes in *S. mansoni*, along with their homologs retrieved from the male  
139 transcriptome).

140 **Table 1. Number of candidate W-derived transcripts, and the genomic location of their**  
141 **closest *S. mansoni* homologs.** “unique genes”, in brackets, refers to the number of *S.*  
142 *mansoni* annotated genes to which a given set of candidates is mapping (several candidates  
143 can map to the same annotated gene). “ZW genes” refers to the ZW linkage group, and can  
144 correspond to either Z-specific genes or pseudoautosomal genes. The sets of ZW homologs  
145 that were used in downstream analyses are in bold.

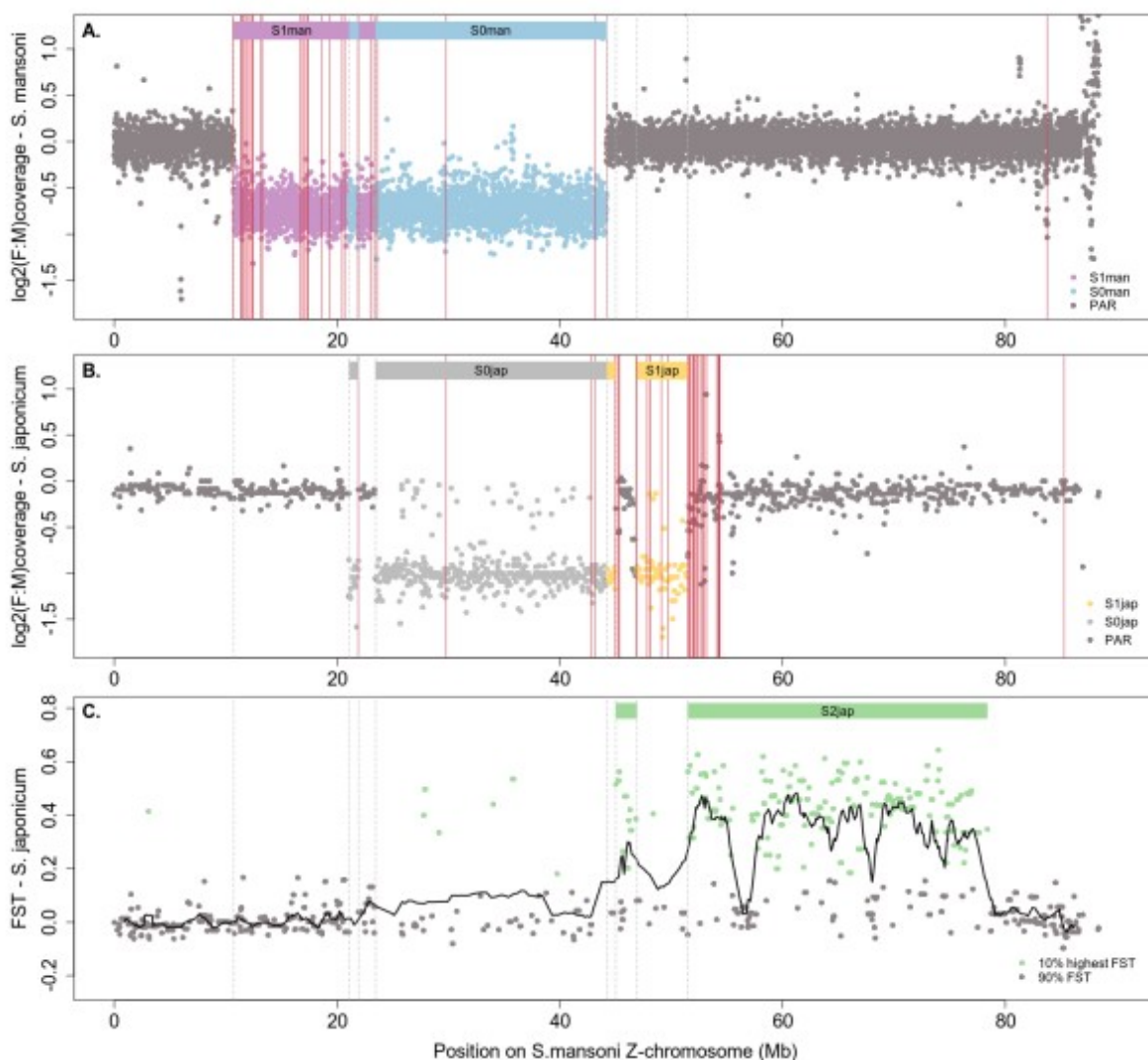
	Number of candidates	Map to <i>S. mansoni</i> CDS (unique genes)	W genes	ZW genes	Other
<b>A.</b> <i>S. mansoni</i> original W-candidates	86	37 (28)	24 (15)	9 (9)	4 (4)
<b>B.</b> <i>S. mansoni</i> combined W-candidates, including annotated genes	90	42 (42)	30 (30)	8 (8)	4 (4)
<b>C.</b> <i>S. mansoni</i> combined set mapped to the [CDS without annotated W-linked genes]	90	40 (39)	NA	<b>34 (33)</b>	6(6)
<b>D.</b> <i>S. japonicum</i> W-candidates	94	48 (37)	0 (0)	<b>48 (37)</b>	0 (0)

## 146 2.2 Most W genes are found in younger evolutionary strata

147 Panels A and B of Figure 1 show the location of Z-linked genes for which at least one

148 homologous W-transcript was detected in *S. mansoni* and *S. japonicum*, along with the  
149 location of the ancestral (S0) and younger lineage-specific strata of the sex chromosomes  
150 (S1man for *S. mansoni* and S1jap for *S. japonicum*, updated from Picard et al. 2018 with the  
151 latest *S. mansoni* assembly v7) (Sup. Dataset 5). Very few S0 genes have a W-homolog (5 in  
152 *S. mansoni* and 4 in *S. japonicum*, or 0.8% and 0.6% of the 635 annotated Z-specific genes  
153 identified in this stratum), as expected from an ancient degenerated non-recombining region.  
154 Three of these are found in both species, consistent with the stratum being ancestral, and  
155 much of the gene loss occurring before the split of the two lineages. The proportion of W-  
156 genes that are preserved is higher for S1man (27 W- versus 299 Z-linked genes, 9.03%, p-  
157 value <0.0001 with a Fisher Exact Probability Test) and S1jap (5 versus 143, 4.2%, p=0.013),  
158 suggesting that degeneration is ongoing in these strata.

159         Interestingly, many putative W-transcripts of *S. japonicum* mapped to a region of the  
160 Z that was previously classified as pseudoautosomal (PAR) (Picard et al. 2018 and Figure 1),  
161 suggesting that this region may have very recently stopped recombining in this species. Such  
162 regions are best detected using population genomic data, as the presence of genetic variants  
163 fixed on the W will lead to detectable levels of genetic differentiation between males and  
164 females (measured as the fixation index  $F_{ST}$ ). Although no such data is available for male and  
165 female *S. japonicum*, we used published genomic data from individual miracidia (Le Clec'h  
166 et al. 2018), which we sexed based on their ratio of Z:autosome genomic coverage, as well as  
167 on the number of reads that mapped perfectly to W-transcripts (Methods and Sup. Figure 4).  
168 Single nucleotide polymorphisms were then called on the resulting 11 males and 8 females,  
169 and mean  $F_{ST}$  between males and females was inferred for each scaffold (Figure 1C and Sup.  
170 Figure 5; scaffold  $F_{ST}$  values are in Sup. Dataset 6). Most of the Z-chromosome to the distal  
171 end of S1jap showed consistently increased female-male  $F_{ST}$ , confirming that this region  
172 corresponds to yet another young stratum of the ZW pair (now referred to as S2jap, a 28.8Mb  
173 region that includes 720 genes). All W-transcripts that mapped to Z-linked scaffolds located  
174 in the *S. japonicum* PAR were consequently assigned to S2jap. Consistent with the  
175 comparable genomic coverage of the male and female samples in this region, a window-  
176 based Copy Number Variation (CNV) analysis inferred less than 1% of gene loss in the S2jap  
177 stratum, a number similar to that observed on the autosomes (Sup. Table 3, Sup. Datasets 7 to  
178 9).



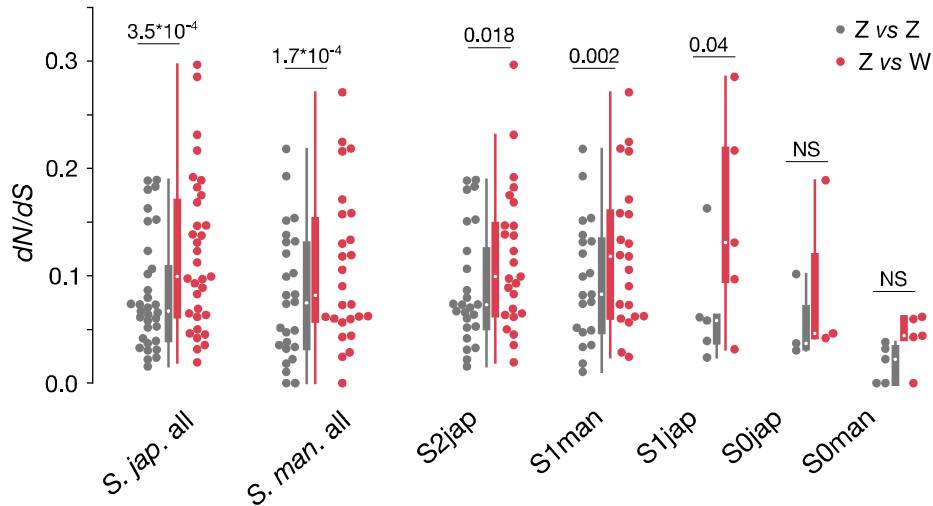
179 **Figure 1: Updated evolutionary strata of schistosome sex chromosomes in *Schistosoma***  
 180 ***mansoni* and *S. japonicum*, and location of pairs of ZW homologous genes.** Panels A and  
 181 B show the log<sub>2</sub> of the female-to-male ratio of coverage along the *S. mansoni* Z-chromosome  
 182 (each dot represents either a 10kb window of the *S. mansoni* genome, or a full scaffold of *S.*  
 183 *japonicum*). Colored rectangles and dots show the various differentiated strata, and the data  
 184 points included in each; the boundaries of the strata are further shown with light dotted grey  
 185 lines. Parts of the Z that were differentiated in both species were assigned to the ancestral  
 186 stratum (S0man in blue, and S0jap in grey), whereas lineage-specific regions of low female  
 187 coverage were assigned to the younger S1man (in purple) and S1jap (in yellow). The  
 188 locations of ZW pairs are shown in dark pink continuous lines. Panel C shows average  
 189 scaffold F<sub>ST</sub> values between males and females of *S. japonicum*: dots represent individual  
 190 scaffolds, and the black line is the mean F<sub>ST</sub> per sliding window of 20 genes. Values above  
 191 the 90% percentile (F<sub>ST</sub>>0.178) of the genome are colored in green, and the corresponding  
 192 putative coordinates of the very young S2jap stratum are denoted by the green rectangles.

### 193 2.3 Patterns of divergence of old and young W-genes



194 The amount of synonymous divergence (estimated as dS, the number of synonymous  
195 substitutions per synonymous site) between Z- and W-homologs is proportional to how long  
196 they have been non-recombining (Kimura 1987; Lahn and Page 1999; Nicolas et al. 2005;  
197 Bergero et al. 2007; Nam and Ellegren 2008; White et al. 2015). Consistent with the putative  
198 strata inferred from the coverage and  $F_{ST}$  analyses, S0 pairs have the highest dS in both  
199 species (median dS of 0.877 and 0.607 in *S. mansoni* and *S. japonicum*, respectively, Sup.  
200 Dataset 4). The median dS of the *S. mansoni* S1 stratum is lower than that of the *S.*  
201 *japonicum* S1 (0.176 versus 0.465), suggesting that at least parts of it may have stopped  
202 recombining more recently. Finally, S2jap ZW homologs have the lowest median dS (0.085),  
203 in agreement with the most recent loss of recombination.

204 W-linked genes typically show an increased ratio of non-synonymous relative to  
205 synonymous divergence (dN/dS) compared to other genomic regions, consistent with the  
206 excessive accumulation of deleterious mutations (Bachtrog et al. 2008; Hough et al. 2014;  
207 Sigeman et al. 2019). To test for increased non-synonymous divergence on the W, we  
208 estimated dN and dS between W-genes and their Z-homologs in the outgroup species (W vs  
209 Z comparisons, e.g. *S. mansoni* W versus *S. japonicum* Z), as well as between the  
210 corresponding Z-genes and their Z-homologs in the outgroup species (Z vs Z) (Figure 2 and  
211 Sup. Dataset 10). As expected under degeneration of the W-chromosome, dN/dS values for  
212 W vs Z comparisons are higher than for Z vs Z comparisons ( $p < 0.001$  in both species, paired  
213 Wilcoxon test). We also tested whether divergence patterns differ between W-genes (and  
214 their Z-homologs) located in older and younger strata. dN/dS values are lower for S0 genes  
215 than for genes in the other strata, for both Z vs Z comparisons (S0 median of 0.02 in *S.*  
216 *mansoni*, compared with 0.08 for S1man,  $p = 0.002$  with a Wilcoxon test; in *S. japonicum*, the  
217 medium dN/dS is 0.04 for S0 versus 0.06 for S1 and 0.07 for S2 genes; the difference is not  
218 significant, but there are only 3 genes in the S0 in this species) and for Z vs W comparisons  
219 (*S. mansoni*: S0 median of 0.04, S1man median of 0.12,  $p$ -value = 0.003, Wilcoxon test; *S.*  
220 *japonicum*: S0 median of 0.05 versus S1jap median of 0.13 and S2jap of 0.10, difference not  
221 significant). This is consistent with genes with important functions being maintained over  
222 long periods of time, as has been observed in mammals and birds (Bellott et al. 2014;  
223 Sigeman et al. 2019; Xu et al. 2019; Bellott and Page 2020).

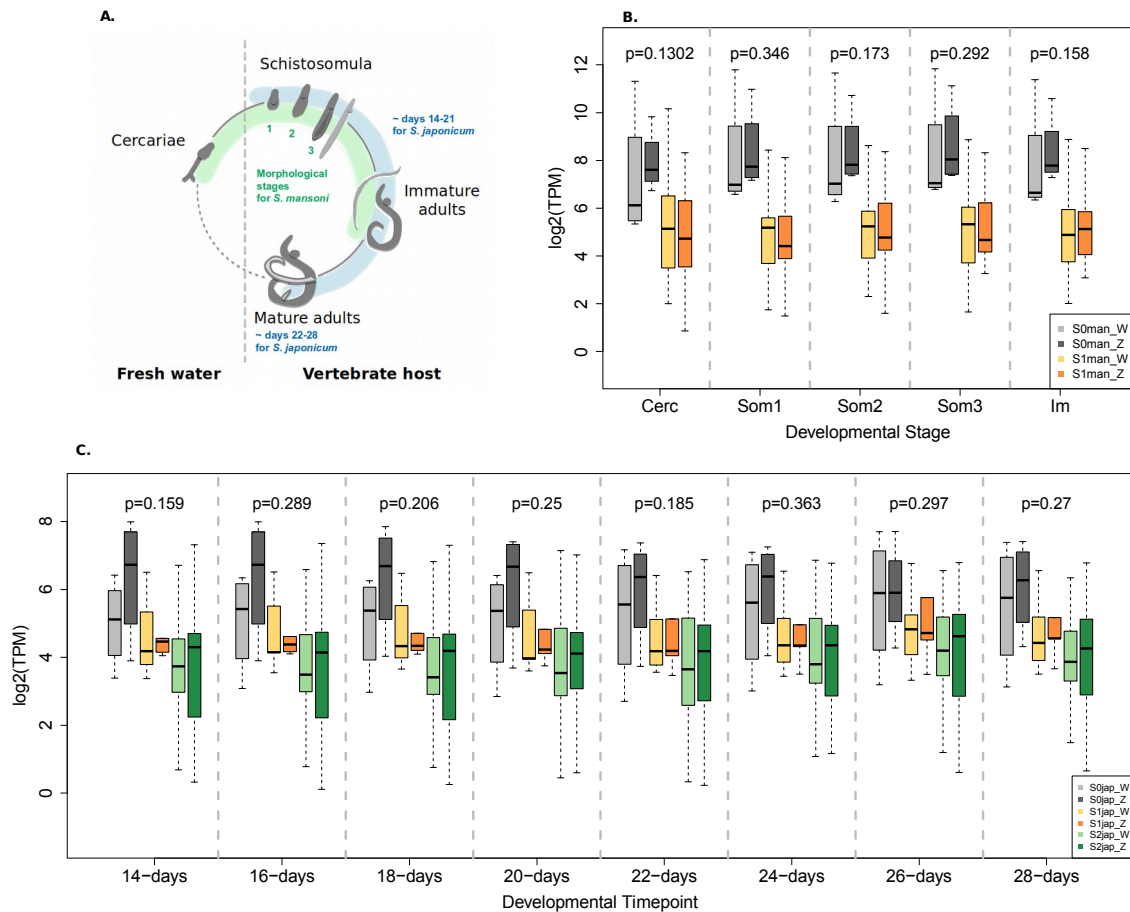


224 **Figure 2: Distribution of dN/dS values between W-genes (Z vs W, in red) or the**  
 225 **corresponding Z-genes (Z vs Z, in grey), and their Z-homologs in the other species.** “*S.*  
 226 *jap.*” refers to W or Z genes in *S. japonicum*, and “*S. man.*” refers to *S. mansoni* genes.  
 227 Boxplots are shown for all ZW homologs (“*S. jap. all*” and “*S. man. all*”), and for ZW  
 228 homologs located in individual strata. The significance of the difference between Z vs Z and  
 229 W vs Z is shown above each boxplot (Wilcoxon tests).

## 230 2.4 Patterns of expression of W-candidates

231 Patterns of expression can be used as an additional measure of functional constraint,  
 232 as essential genes tend to have high and broad expression (Liao et al. 2006; Wang et al. 2015;  
 233 Kabir et al. 2017; Fraïsse et al. 2019). Furthermore, the first evidence of genetic degeneration  
 234 of Y/W-linked genes is often a decrease in expression relative to that of the X/Z (Zhou and  
 235 Bachtrog 2012; Gammerdinger et al. 2014; Hough et al. 2014; Pucholt et al. 2017). We  
 236 therefore estimated gene expression levels (in Transcripts Per Million, TPM) using published  
 237 *S. mansoni* and *S. japonicum* male and female expressions at different developmental time  
 238 points, to investigate differences in the expression patterns of ZW gene pairs between strata,  
 239 and to compare the expression of W-linked genes to their Z-linked counterpart (Sup. Datasets  
 240 11 and 12). Reads were mapped to curated transcriptomes that included our W-candidates  
 241 (see Methods) using Kallisto, an RNA quantification program capable of inferring  
 242 paralog/allele-specific expression (Bray et al. 2016). An overview of gene expression of all  
 243 protein-coding W-candidates is provided in Sup. Figures 6 and 7, and shows that the majority  
 244 of them are expressed in at least some female stages (but not or much less expressed in males,  
 245 confirming that we are for the most part correctly discriminating between Z- and W-derived  
 246 expression of ZW gene pairs).

247 Figure 3 shows the distribution of female expression levels of W-candidates from the  
248 different strata and those of their Z-homologs (when more than one W-candidate mapped to  
249 the same Z-linked gene, only the longest was kept; see Sup. Figure 8 for an additional *S.*  
250 *mansoni* RNA-seq dataset; only transcripts with TPM>1 were considered). The median W:Z  
251 expression ratio is below 1 for all sampled time points in both species (Sup. Figures 9 to 11),  
252 but these differences are not significant (p-values in Figure 3B and 3C for all ZW pairs, and  
253 Sup. Figures 9 to 11 for individual strata). While we may simply lack power, as the number  
254 of genes for which we can perform comparisons is small (especially for the S0), this may also  
255 suggest that there is significant selective pressure against loss of expression of genes  
256 maintained on the W. The median expression of S2jap W transcripts is also below the median  
257 Z expression at every available time point (although this is again not significant for individual  
258 comparisons, Sup. Figure 9), consistent with some loss of expression occurring early in sex  
259 chromosome evolution. Finally, although direct comparisons of the gene expression of W-  
260 candidates (or Z-homologs) between strata are only significant for some developmental  
261 stages (Kruskal-Wallis test, Sup. Tables 4 to 6), S0 W-linked genes and their Z-linked  
262 homologs have the highest median expression level at all time points available for both  
263 species (Figure 3), providing further support for their functional importance (and/or  
264 potentially contributing to the differences in dN/dS observed between strata).



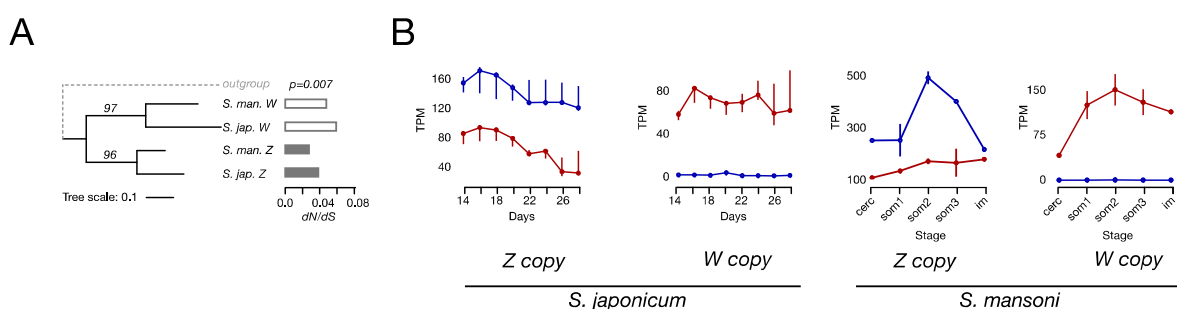
265 **Figure 3: Distribution of S0, S1 and S2 W- and Z- gene expression throughout female**  
 266 **development.** Panel A shows a simplified schematic of the life cycle of schistosomes,  
 267 depicting the larval stage in fresh water and the process of sexual maturation in the vertebrate  
 268 host. The part of development represented in the expression datasets is shown in green for *S.*  
 269 *mansoni*, and blue for *S. japonicum*. An additional dataset with *S. mansoni* mature adults is  
 270 plotted in Sup. Figure 8. Panel B shows gene expression values (TPM) for the *S. mansoni* W-  
 271 candidates and their Z-homologs according to their respective strata (S0 in grey, S1 in  
 272 orange) across five developmental stages: cercariae “Cerc”, three subsequent schistosomulum  
 273 stages “Som1-3” and immature adult schistosomes “Im”. Panel C shows gene expression  
 274 values (TPM) for the *S. japonicum* W-candidates and their Z-homologs according to their  
 275 respective strata (S0 in grey, S1 in orange, S2 in green) across 8 different developmental  
 276 timepoints (in days post-infection). P-values above each boxplot denote the significance of the  
 277 difference in expression between W- and Z-derived transcripts, considering all strata together  
 278 (Wilcoxon test).

### 279 2.5 Shared ancestry suggests candidate for sex-determination

280 While the master switch of sex determination may have changed since the split of the  
 281 Asian and African schistosome lineages (see Discussion), W-linked genes that are shared  
 282 between the two are promising candidates. If the two lineages still share the ancestral sex-  
 283 determining gene, this gene should: 1. show a clear phylogenetic clustering of the W-copies

284 from the two species, consistent with ancestral W-linkage 2. have a low dN/dS value,  
 285 supporting functional conservation 3. show expression in females of the two species during  
 286 sex determination. Figure 4 and Sup. Figure 12 show the phylogeny, branch-specific dN/dS  
 287 and patterns of expression throughout development for the three W-linked genes (and their Z-  
 288 homologs) that were found in both species: a U2 snRNP auxiliary factor large subunit  
 289 (U2AF2, OrthoGroup OG0000710 in Sup. Dataset 4), a Ubiquitin conjugating enzyme  
 290 variant (Uev, OrthoGroup OG0000869) and an Ankyrin repeat and KH domain-containing  
 291 protein 1 (ANKHD1, OrthoGroup OG0000874). Of the three, U2AF2 stands out as fitting all  
 292 three predictions (Figure 4), as it has high expression throughout female development and  
 293 dN/dS below 0.1 in both species. It is also the only W-candidate in either species with the  
 294 term “female sex differentiation” in its functional annotation (Sup. Datasets 13 and 14),  
 295 further strengthening the case for its role in determining sex.

296 The *S. mansoni* W-copy of ANKHD1 is much shorter than the Z-copy, or than the *S.*  
 297 *japonicum* Z- and W-homologs, and it has a higher branch-specific dN/dS than any of its  
 298 homologs, consistent with loss of function. It also shows no expression at any female stage of  
 299 *S. mansoni*, making it an unlikely candidate for sex determination. Phylogenetic clustering of  
 300 Uev homologs occurs by species rather than by chromosome, suggesting that it was not on  
 301 the W-chromosome before the split of the two clades, again arguing against an ancestral sex-  
 302 determining function.



303 **Figure 4: Evolution and expression of the shared S0 gene U2AF2.** Panels A shows the  
 304 gene tree with bootstrap values. Terminal-branch specific dN/dS values along with the Chi-  
 305 squared p-values of the deviations of observed values from the uniform assumption are  
 306 shown as histograms. White bars portray dN/dS of W-specific genes, grey bars show dN/dS  
 307 values of the Z-copies. Panel B shows gene expression values (TPM) of Z- and W-copies of  
 308 U2AF2 on different developmental stages of *S. japonicum* and *S. mansoni*. The spread  
 309 between the lowest and the highest value among the replicates is shown with error-bars,  
 310 medians are shown with dots. Stages in *S. mansoni*: “cerc” means cercariae, “som1-3” are  
 311 three subsequent schistosomula stages and “im” stands for immature adults. Red and blue  
 312 lines show TPM values of females and males, respectively.

### 313 3. Discussion

#### 314 3.1 An efficient k-mer pipeline for detecting W-transcripts at all levels of divergence

315 Y and W chromosomes are notoriously difficult to study, and were largely excluded  
316 from early genome projects. Many bioinformatics approaches have since been developed to  
317 identify Y/W-derived sequences from next-generation sequencing data, typically either based  
318 on differences in male and female DNA/RNA-seq coverage of Y/W-derived transcripts  
319 (Cortez et al. 2014; Zhou et al. 2014; Smeds et al. 2015), or on differences in k-mer  
320 frequencies between male and female samples (Bernardo Carvalho and Clark 2013;  
321 Tomaszekiewicz et al. 2016; Li et al. 2018; Rangavittal et al. 2018; Rangavittal et al. 2019).  
322 Because they require that Y/W-reads do not map to X/Z-derived sequences, coverage-based  
323 approaches are more suitable to identify highly differentiated sex chromosomes (our own  
324 attempt at implementing such an approach in *S. japonicum* yielded only a few candidate W-  
325 transcripts, data not shown). Multiple k-mer based approaches have been used to identify  
326 and/or assemble W and Y specific genomic contigs and transcripts. Early approaches  
327 required a genome or transcriptome assembly obtained from the heterogametic sex (Bernardo  
328 Carvalho and Clark 2013; Tomaszekiewicz et al. 2016; Rangavittal et al. 2018; Rangavittal et  
329 al. 2019). These were also best suited to identify differentiated sex-linked sequences, as Y/W  
330 sequences do not necessarily assemble into separate scaffolds when they are very similar to  
331 X/Z regions. More recently, k-mer based approaches have been used to first extract DNA  
332 reads that contain a large fraction of sex-specific k-mers, which are then assembled separately  
333 from reads derived from the rest of the genome, directly yielding candidate Y/W sequences  
334 (Tomaszekiewicz et al. 2016; Li et al. 2018). Such read-selecting approaches have been  
335 successfully applied to the differentiated sex chromosomes of Gorilla and Human  
336 (Rangavittal et al. 2019), *Bombyx mori* (ZW), *Drosophila melanogaster* (XY), and  
337 *Anopheles gambiae* (XY) (Li et al. 2018) but also to the very young Y chromosome of two  
338 guppy species (Morris et al. 2018). This encouraged us to use a similar strategy to tackle the  
339 unique evolutionary history of the W chromosome in schistosomes. Since we were primarily  
340 interested in finding W-linked genes, only k-mers that were found in both female DNA-seq  
341 and RNA-seq data (but not in male DNA or RNA data) were classified as female-specific,  
342 and used to select and assemble RNA-seq reads directly into a set of W-specific transcripts.  
343 This has several advantages: 1. it reduces the need for extensive genomic data when RNA  
344 samples are available; 2. it reduces the number of sex-specific k-mers that must be dealt with,

345 making the pipeline more efficient; 3. only putative Y/W-derived RNA-seq reads are  
346 assembled into transcripts, avoiding issues of repetitive sequences and hybrid assemblies  
347 between homologous genes when the sex chromosomes are poorly differentiated. Our final  
348 set of candidates included ancestral W-linked genes that were highly differentiated from their  
349 Z-homologs, but also uncovered a new evolutionary stratum of the *S. japonicum* ZW pair that  
350 could not be detected with coverage approaches (Picard et al. 2018), demonstrating the power  
351 of this method for studying sex-specific sequences in species that have varying levels of sex  
352 chromosome differentiation, or for which such information is missing. Finally, our pipeline is  
353 based on the published and efficient k-mer manipulation package BBMAP (Bushnell 2014),  
354 making it easy to implement for any organism for which male and female data is available,  
355 even in the absence of a reference genome.

### 356 **3.2 Temporal dynamics of W degeneration**

357 Schistosome sex chromosomes have various evolutionary strata that differ between  
358 the closely related *S. mansoni* and *S. japonicum*, allowing us to probe the evolution of W-  
359 genes at different timepoints after the loss of recombination. In particular, with the inclusion  
360 of the very young S2jap, a very broad timeline of sex chromosome evolution is represented in  
361 this group. Similar analyses have been performed in species (or species groups) that have XY  
362 systems/strata of varying ages (Bachtrog 2013; Hough et al. 2014; Schultheiß et al. 2015;  
363 Crowson et al. 2017), but are mostly lacking in species with ZW chromosomes, for which  
364 information on early and late sex chromosome evolution typically come from different  
365 lineages (but see (Sigeman et al. 2019) for an investigation of the multiple neo-sex-  
366 chromosomes of lark birds). Our results illustrate in one clade the insights on sex  
367 chromosome evolution gained from these various organisms (Kaiser et al. 2011; Bachtrog  
368 2013; Hough et al. 2014; Bellott et al. 2017; Crowson et al. 2017). Soon after recombination  
369 is lost on the W, genes that are under weak purifying selection start accumulating non-  
370 synonymous mutations, and their expression decreases relative to that of their Z-homologs  
371 (Kaiser et al. 2011; Hough et al. 2014). Over time, these genes are lost, and only increasingly  
372 important (and highly expressed) genes are maintained on the W (Bellott et al. 2014;  
373 Crowson et al. 2017). Finally, the few genes that remain on very ancient strata become stably  
374 maintained over long periods of time (Bachtrog 2013).

375 The presence of old and young strata makes schistosomes a promising model for  
376 studying how such a stepwise loss of recombination can occur. Local loss of recombination

377 between sex chromosomes was originally thought to be driven by inversions, as these prevent  
378 homologous pairing and crossing over during meiosis (Charlesworth et al. 2005). Although it  
379 is supported by the order of XY gene pairs on mammalian sex chromosomes (Lemaitre et al.  
380 2009), and inversions have been detected on young sex chromosomes (Wang et al. 2012;  
381 Roesti et al. 2013), this model has been brought into question by the discovery of unstable  
382 boundaries between the recombining and non-recombining regions of sex chromosomes  
383 (Cotter et al. 2016; Campos et al. 2017; Wright et al. 2017). Instead, changes in epigenetic  
384 state, potentially driven by the accumulation of transposable elements, may first repress  
385 recombination (Lepesant et al. 2012), with inversions accumulating later (Charlesworth 2017;  
386 Furman et al. 2020). A chromosome-level assembly of the *S. japonicum* and *S. mansoni* W  
387 and Z, as well as a thorough investigation of their chromatin state (Picard et al. 2019), will  
388 make it possible to compare ancestral and derived gene order and chromatin landscape for S1  
389 and S2 strata, and potentially provide answers as to how ZW recombination was repressed in  
390 this clade.

### 391 **3.3 Sex determination in African and Asian schistosomes**

392 Our comparative analysis of the content of the W-chromosome of the two main  
393 schistosome lineages yielded very few genes shared between them, one of them an interesting  
394 candidate for sex determination: U2AF2, a conserved house-keeping gene involved in pre-  
395 mRNA splicing from yeast to *Drosophila* (Kanaar et al. 1993; Potashkin et al. 1993).  
396 Although U2AF2 does not play a direct role in sex determination in other clades, homologs  
397 are known to be involved in meiosis-mitosis fate decision in *C. elegans* (*uaf-1*) (Kerins et al.  
398 2010), and in sex-specific splicing in insects (*U2AF-50*) (Verhulst et al. 2010). A W-linked  
399 copy of U2AF2 was previously identified in *S. japonicum* (Liu et al. 2020), and is highly  
400 expressed throughout the female life cycle in the two schistosome species (Fig. 3 and (Liu et  
401 al. 2020)). We hypothesize that the W copy of U2AF2 may have been co-opted for sex  
402 determination, while its homolog on the Z retained the ancestral general pre-mRNA splicing  
403 function. The identification of this candidate is based on the assumption that the sex  
404 determination pathway is conserved in schistosomes. Since the African and Asian groups  
405 share part of the ZW pair, they must have ancestrally had the same sex determining master  
406 switch, but whether this is still true is at this point unclear, as no gene has been functionally  
407 linked to primary sex determination in either species (Wang et al. 2019). Although estimates  
408 of the age of the clade vary widely (Lawton et al. 2011), the fact that new strata have become



409 differentiated in each lineage, and that the median rate of synonymous divergence between  
410 them is substantial (65%, Picard et al 2018) suggests that they have been separated for long  
411 enough that turnover of the sex determination switch could have occurred. On the other hand,  
412 the sexual development of males and females is similar in the two species, and narrowing the  
413 search to the few shared W-linked genes seems like a reasonable first step. Interestingly, the  
414 two other shared candidates, *ANKHDI* and *Uev*, both show some similarity to genes that are  
415 part of - or interact with - the sex determination pathway of nematodes. The *Caenorhabditis*  
416 *elegans* sex determination gene *Fem1* contains Ankyrin repeats, similar to our W-candidate  
417 ANKHD1 (Spence et al. 1990). *Fem1* interacts with a ubiquitin ligase to regulate *Tra*, the  
418 terminal effector of sex determination (Starostina et al. 2007); our candidate *Uev* is an E2  
419 ubiquitin-conjugating enzyme, which is required by ubiquitin ligases to mark target proteins  
420 for degradation. Aside from providing other possible candidates for sex determination, the  
421 fact that the three shared W-candidates show similarity to genes with sex-related functions  
422 supports the idea that genes that are maintained on W-chromosomes for long periods of time  
423 may be likely to perform female functions. Finally, we only focused here on genes with  
424 homology to known protein-coding genes. Non-coding transcripts and microRNAs present on  
425 the W-chromosome could also play a role in sex determination and differentiation (Marco et  
426 al. 2013; Kiuchi et al. 2014; Zhu et al. 2016; Maciel et al. 2020), and need further  
427 investigation.

## 428 **4. Methods**

### 429 **4.1 Data**

430 All analyses were performed on publicly available data. The *S. mansoni* genomic  
431 libraries were downloaded from Bioproject PRJEB2320 (Wellcome Sanger Institute), and the  
432 *S. japonicum* genomic libraries from the three following BioProjects: PRJNA432803 (IST  
433 Austria), PRJNA354903 (Wuhan University), PRJNA650045 (University of Texas at  
434 Arlington). The RNA-seq libraries for *S. mansoni* were downloaded from Bioprojects  
435 PRJNA312093 (Université of Perpignan *Via Domitia*) and PRJEB1237 (Wellcome Sanger  
436 Institute), and the *S. japonicum* RNA-seq libraries from PRJNA343582 (National Institute of  
437 Parasitic Diseases), PRJNA252904 (Chinese Academy of Medical Sciences and Peking  
438 Union Medical College), and PRJNA504625 (Wuhan University). A detailed list of  
439 individual accession numbers, and of the steps for which they were used, is provided as Sup.  
440 Table 1.

## 441 4.2 k-mer based assembly of W-linked transcripts

442 **k-mer analysis** For each species, we used one female and one male genomic library,  
443 as well as one (*S. japonicum*) or two (*S. mansoni*) replicates of pooled female RNA-seq  
444 libraries, and several individual male RNA-seq libraries. In the case of *S. mansoni*, the two  
445 female RNA-seq replicates were made by pooling the first replicates (PRJNA312093) of  
446 female cercariae, schistosomula, immature and mature adult libraries together, and the second  
447 replicates in the same way (such that transcripts with stage specific expression are  
448 represented in each of the samples). In the case of *S. japonicum*, a single RNA-seq pool was  
449 made by merging two female schistosomula libraries (PRJNA343582) and one mature female  
450 adult library (PRJNA252904). In both species, the equivalent RNA-seq libraries were  
451 available for males (*S. mansoni*: 2 replicates each of the 3 developmental stages,  
452 PRJNA312093, and of mature adults, PRJEB1237; *S. japonicum*: one replicate each of two  
453 schistosomula stages, PRJNA343582, and mature adults, PRJNA252904), and used  
454 individually.

455 The read libraries for both species were trimmed with the Trimmomatic package  
456 (Bolger et al. 2014). Our k-mer based pipeline utilizes the tools included in the BMAP  
457 package (Bushnell 2014), and was run separately for each species. kmercountexact.sh was  
458 first used to output the unique 31 base pair k-mers in each of the female libraries separately,  
459 yielding one set of unique k-mers per female RNA/DNA library. We then used the same  
460 function to extract the k-mers that are shared between all the female DNA and RNA k-mer  
461 sets, by setting the mincount parameter to the total number of libraries. The resulting set of  
462 shared female k-mers was then filtered by removing any 31-mers found in any of the male  
463 DNA and RNA libraries using bbduk.sh. This set of female specific 31-mers was then used as  
464 input for bbduk.sh to recover female RNA reads with at least 40% female-specific k-mers  
465 (“minkmerfraction” parameter set to 0.4; this threshold is not very stringent in order to allow  
466 for the transcripts to assemble properly, but requires downstream filtering of assembled  
467 transcripts). Those reads were then assembled using SOAPdenovo-trans (Xie et al. 2014), and  
468 fafilter (UCSC source code collection) was used to remove all transcripts shorter than 200 bp  
469 in both species.

470 **Filtering the *S. mansoni* candidates** Bowtie2 (Langmead and Salzberg 2013)  
471 (options --no-unal --no-hd --no-sq) was used to map the female (ERR562990) and male  
472 (ERR562989) genomic reads to the output of the k-mer pipeline (175 transcripts). The

473 number of perfectly matching male and female reads (reads with 0 edit distance “NM:i:0”) to  
474 each transcript were counted. All transcripts that had fewer than 20 perfectly matching female  
475 reads and/or a ratio of (male:(male+female)) perfect matches of more than 0.1 were removed.  
476 In order to have a more comprehensive set of W-transcripts for downstream analyses, the  
477 coding sequences of the 32 annotated W genes were added to our set of candidates. In order  
478 to collapse transcripts that were both in the k-mer and annotated sets, we then used a perl  
479 script (SpliceFinder.pl) to produce clusters of transcripts with >100 bp of alignment and less  
480 than 1% divergence (putative isoforms), and kept only the longest isoform per cluster. The  
481 final set had 97 W-candidates.

482 **Filtering and improvement of assembly of the *S. japonicum* candidates** Bowtie2  
483 (options --no-unal --no-hd --no-sq) was used to map a new set of four male only genomic  
484 libraries (SRR5054524, SRR5054649, SRR5054671, SRR5054674) and three mixed  
485 (males+females) libraries (SRR5054672, SRR5054673, SRR5054701) to the output of the k-  
486 mer pipeline (1041 transcripts). All transcripts with a sum of perfect matches from the mixed  
487 genomic libraries of less than 15 reads and/or a ratio of (male:(male+mixed)) of less than 0.1  
488 were removed. In order to obtain longer sequences for downstream analyses of the resulting  
489 157 *S. japonicum* candidates, we originally mapped them to annotated *S. japonicum* coding  
490 sequences (<https://parasite.wormbase.org/index.html>, [Howe et al. 2016; Howe et al. 2017]):  
491 these appeared to often contain hybrids of W-linked genes and of their Z-homologs (as parts  
492 of it were completely identical to our candidates while other parts were clearly diverged, data  
493 not shown). We therefore mapped our candidates to a long k-mer female transcriptome  
494 assembly (SOAPdenovo-trans assembly of all reads in Bioproject PRJNA343582, with  
495 K=65, Sup. Dataset 15), which should be largely devoid of ZW hybrid assemblies. W-  
496 candidates were mapped to the transcriptome with BLAT, and scaffolds with a minimum  
497 match score of 50 and less than 1% divergence were retrieved. Cap3 (Huang and Madan,  
498 1999) was used to further assemble them. We merged the output of Cap3 with our 157  
499 transcripts and used Splicefinder.pl to keep the longest transcript per gene. We once again  
500 mapped the four male-only samples and the three mixed samples to the resulting set and  
501 followed the same filtering approach described above, yielding a final set of 96 W-specific  
502 candidates.

503 To identify the Z-homologs for the *S. japonicum* W-candidates, we assembled a male  
504 transcriptome from male reads (SRR8175618, Sup. Dataset 15). The reads were trimmed

505 with the Trimmomatic package (Bolger et al. 2014) and then assembled using Trinity  
506 (Grabherr et al. 2011) followed by Cap3 (Huang and Madan 1999). We mapped our  
507 candidates to the male assembly using BLAT (with a translated query and database and a  
508 minimum match score of 50), and selected only the transcript with the highest match score  
509 for each W-candidate, which was used as its homolog. The completeness of both the female  
510 and male transcriptomes was assessed using BUSCO (Felipe et al. 2015) in the transcriptome  
511 mode with the metazoa-specific set (metazoa\_odb10) and compared with the BUSCO scores  
512 for the published *S. japonicum* and *S. mansoni* transcriptomes (Sup. Figure 13).

### 513 **4.3 Mapping of W-candidates to *S. mansoni* coding sequences**

514 We mapped the filtered sets of candidates of both species to the published *S. mansoni*  
515 coding sequences (<https://parasite.wormbase.org/index.html>, [Howe et al. 2016; Howe et al.  
516 2017]) using BLAT with a translated query and dataset, and a minimum match score of 100  
517 (Table 1). Only the CDS with the highest matching score was kept for each candidate. In the  
518 case of *S. mansoni*, we performed three different mappings: 1. the filtered original candidates  
519 against the full set of *S. mansoni* CDS; 2. the final combined set against the full *S. mansoni*  
520 CDS; 3. the final combined set against the *S. mansoni* CDS, but with all the different  
521 transcript isoforms of the 32 annotated W genes removed using a perl script, in order to  
522 detect close homologs of W-linked genes in the genome (Table 1).

### 523 **4.4 Definition of shared and lineage-specific strata based on coverage analysis**

524 ***De novo* determination of Z-specific regions in *Schistosoma mansoni*** Z-specific  
525 regions were determined for the latest version of the *S. mansoni* genome based on female and  
526 male genomic coverage, following the bioinformatic pipeline described in Picard et al. 2018.  
527 The reference genome, coding sequences (CDS) and their respective chromosomal locations  
528 (BioProject PRJEA36577, version WBPS14) were obtained from the WormBase Parasite  
529 database (<https://parasite.wormbase.org/index.html>, [Howe et al. 2016; Howe et al. 2017]),  
530 on the 30<sup>th</sup> of October 2020. We estimated the genomic coverage for females (ERR562990)  
531 and males (ERR562989) by mapping the DNA reads to the genome with Bowtie2 (v2.3.4.1)  
532 and by using the uniquely mapped reads as SOAPcoverage input (SOAPcoverage v2.7.7).  
533 The coverage analysis was performed for windows of 10kb. We classified as  
534 pseudoautosomal (PAR) any window on the ZW linkage group that had a F:M coverage ratio  
535 higher than the 2.5 percentile of the autosomes. The remaining windows were considered to

536 be putatively Z-specific. The 1% of Z-specific windows with the highest F:M coverage ratio  
537 were further marked as “ambiguous” and not considered for strata assignments (Sup. Figure  
538 14). The classification of Z-specific and PAR regions, and of the CDS that they contain, is  
539 reported in Sup. Dataset 5.

540 **Inference of the location of *S. japonicum* scaffolds along the *S. mansoni* Z-**  
541 **chromosome** *S. japonicum* reference genome were obtained from the WormBase Parasite  
542 database (<https://parasite.wormbase.org/index.html>, [Howe et al. 2016; Howe et al. 2017]),  
543 on the 30<sup>th</sup> of October 2020 (BioProject PRJEA34885, version WBPS14). We mapped *S.*  
544 *mansoni* CDS to *S. japonicum* scaffolds using BLAT (Kent 2002) with a translated query and  
545 database. The BLAT alignment was then filtered to keep only the mapping hit with the  
546 highest score for each *S. mansoni* CDS. In a second filtering step, when several *S. mansoni*  
547 genes overlapped on the *S. japonicum* genome by more than 20 bp, we kept only the highest  
548 mapping score. The position of each *S. japonicum* scaffold on the *S. mansoni* Z-chromosome  
549 was inferred from the position of the *S. mansoni* genes that mapped to it.

550 **Strata definition** The coverage-based assignment of *S. japonicum* scaffolds to Z-  
551 specific (Z) or pseudoautosomal (PAR) regions was obtained from Picard et al. 2018, and is  
552 reported in Sup. Dataset 5. The *de novo* strata assignment shown in Figure 1 was based on the  
553 comparison of the classification of the genomic location of each *S. mansoni* gene as Z-  
554 specific or PAR, and the classification of the scaffold that their CDS mapped to in *S.*  
555 *japonicum* (see above). Genes were assigned to the stratum S1man if they were located in a  
556 Z-specific window in *S. mansoni* but their CDS mapped to a scaffold classified as PAR in *S.*  
557 *japonicum*, and vice versa for S1jap. Genes that mapped to Z-specific genomic regions in  
558 both species were assigned to the stratum S0. Finally, the analysis of  $F_{ST}$  between males and  
559 females (see below) revealed a new stratum specific to the *S. japonicum* lineage (S2jap).  
560 Genes were assigned to this stratum if they mapped to windows/scaffolds classified as PAR  
561 in both species, but the mean  $F_{ST}$  value of the *S. japonicum* scaffold was above the 90th  
562 percentile of the distribution of the entire genome. Final coordinates of the strata were  
563 defined by the first base of the first *S. mansoni* CDS assigned to a specific stratum and the  
564 last base before the start of the first gene assigned to a different stratum (Sup. Dataset 5, Sup.  
565 Table 7). More than five consecutive assignments to a given region were needed to change  
566 strata. If in those five first CDS one was assigned as “ambiguous”, the original assignment of  
567 the windows was considered.

## 568 4.5 $F_{ST}$ analysis

569 We downloaded whole genome sequencing data of 22 *S. japonicum* individual  
570 Miracidia samples (PRJNA650045). The sex of the samples was not identified, so we used  
571 Bowtie2 (v2.3.4.1, --end-to-end --sensitive mode) to map the reads to the *S. japonicum*  
572 genome. The resulting SAM files were filtered to keep only uniquely mapped reads, and the  
573 genomic coverage for each scaffold was estimated from the filtered SAM files with  
574 SOAPcoverage (v2.7.7, <http://soap.genomics.org.cn>). We used the output to calculate the  
575 median Z:Autosomal genomic coverage for each of the samples based on the scaffold  
576 assignments from (Picard et al. 2018). Furthermore, we used bowtie2 (--no-unal --no-hd --no-  
577 sq) to map the forward reads of all the samples to our W-candidates and three randomly  
578 chosen Autosomal controls; we filtered for 0 edit distance “NM:i:0” to find the number of  
579 perfectly-matched reads. We removed three of the samples due to low coverage, and the  
580 remaining 19 were identified as 11 males ( $\text{Log}_2(\text{Z:A}) > 0.9$ ) and 8 females ( $\text{Log}_2(\text{Z:A}) < 0.75$   
581 and ratio of reads mapping to W versus autosomal controls  $> 100$ , Sup. Figure 4).

582 We used Bowtie2 to index the *S. japonicum* genome and map the genomic reads from  
583 the 19 libraries to it (--end-to-end --sensitive mode). The SAM files were converted to sorted  
584 BAM files using Samtools. We detected the genetic variants (SNPs) in the samples using the  
585 BCFtools mpileup function, filtered the output for minimum and maximum read coverage  
586 and mapping quality using VCFtools, and removed multi-allelic sites using BCFtools. The  
587 calculation of the per-site  $F_{ST}$  between males and females was performed using VCFtools, and  
588 the output was used to calculate the mean  $F_{ST}$  values between the male and female samples  
589 for each scaffold.

## 590 4.6 CNV analysis in *S. japonicum* female.

591 After read mapping with Bowtie2 against *S. japonicum* genome, Copy Number  
592 Variation (CNV) analysis was performed using the control-FREEC prediction tool (Boeva et  
593 al. 2012), for genomic windows of 1kb, 5kb and 10kb, testing the female genomic library  
594 (SRR6841388) against the male genomic library as reference (SRR6841389). Only deletions  
595 with statistical support (wilcoxon P-value  $< 0.05$ ) were further analysed. *S. japonicum* genes  
596 that overlapped with a deleted window were classified as putatively lost (Sup. Datasets 7 to  
597 9).

## 598 4.7 Estimating the rates of evolution of ZW homologs

599 The coding sequences of W-candidates of both species (and of their closest homologs  
600 in *S. japonicum*) were obtained with the command-line version of Genewise (Wise2 package  
601 v2.4.1), using the protein sequence of the closest *S. mansoni* homologs (section 4.2) as input.  
602 The coding sequences of W-candidates and their closest homologs (within the same species)  
603 were aligned with the TranslatorX package (Abascal et al. 2010) with the “gblocks” option to  
604 filter out unreliable sections of the alignment. The dN and dS values were obtained with  
605 KaKs\_calculator2.0 (Wang et al. 2010) using the Yang Nielsen algorithm (YN). For the  
606 dN/dS and dS between ZW pairs of *S. mansoni* and *S. japonicum* (Figure 2A), we only  
607 considered the KaKs\_calculator2.0’s estimates for pairs with 300 sites or more, and dS above  
608 0 (Sup. Dataset 4).

609 To estimate dN and dS between W-genes and their Z-homologs in the outgroup  
610 species (W vs Z comparisons), as well as between the corresponding Z-genes and their Z-  
611 homologs in the outgroup species (Z vs Z), we considered only the instances in which the  
612 ZW pair and at least one sequence from the outgroup species clustered in the same  
613 Orthogroup (section 4.7). When multiple orthologs were present in the outgroup species, the  
614 sequence with the highest BLAT match score to the W-candidate was used. Sequence  
615 alignments (of the three sequences) and pairwise dN and dS estimation were then performed  
616 as before (Sup. Dataset 10).

617 To obtain branch-specific estimates of dN/dS for W-candidates shared between the  
618 two schistosome lineages (Fig. 4), we aligned W- and Z-homologs from both species, as well  
619 the closest *Clonorchis sinensis* sequence (an outgroup of schistosomes) placed in the  
620 respective Orthogroup (section 4.7). Trees were constructed with the maximum likelihood  
621 method (Felsenstein, 1973) implemented in the PhyML package with default parameters  
622 (Guindon and Gascuel, 2003) and 100 bootstraps, and visualized on the IToL web server  
623 (Letunic and Bork, 2016). Branch-specific divergence rates were estimated with PAML 4.9  
624 (Yang, 2007), and compared to a null model presuming the same dN/dS on all branches  
625 (likelihood-ratio test, Jeffares et al., 2015).

#### 626 **4.8 Transcriptome curation and gene expression analysis**

627 **Curation** To avoid multi-mapping reads resulting from having multiple transcript  
628 isoforms of the same gene in the expression analysis, we used the in-house script  
629 (Splicefinder.pl) to remove all the isoforms from the published transcriptomes of both

630 species. In addition to that, we mapped our set of W-candidates and their homologs to their  
631 respective transcriptomes and removed any transcripts with matches > 100 bp and less than  
632 5% divergence. This is important to remove all the existing isoforms of our transcripts from  
633 the published transcriptomes, some of which are possibly hybrid assemblies between the Z  
634 and the W. Following that, we added our candidate set to the transcriptomes in order to  
635 perform gene expression analysis to untangle the differences in expression patterns between  
636 males and females.

637 **Expression analysis** For *S. japonicum*, the analysis of expression was carried out  
638 using 48 *S. japonicum* RNA-seq read libraries corresponding to three replicates of RNA-seq  
639 read data collected at 8 time points of male and female *S. japonicum* development in the  
640 definitive host (Wang et al. 2017, bioproject PRJNA343582). For *S. mansoni*, the expression  
641 analysis was carried out using 28 *S. mansoni* RNA-seq libraries corresponding to two  
642 replicates of RNA-seq read data from two different studies (20 libraries from PRJNA312093  
643 (Picard et al. 2016) and 8 libraries from PRJEB1237) for different developmental stages:  
644 cercariae, three schistosomula stages, immature (unpaired) and mature (paired) adults. The  
645 transcriptomes curated specifically for the expression analysis are provided as a Sup. Dataset  
646 16. The raw reads were mapped onto the transcriptome using Kallisto (v0.44.0, Bray et al.  
647 2016), and TPM values were obtained from the Kallisto output and quantile-normalized with  
648 NormalyzerDE (Willforss et al. 2018).

#### 649 **4.9 Identifying the shared candidates using OrthoFinder**

650 The identification of the orthologous genes between the two species was performed  
651 using OrthoFinder (Emms et al. 2019). We followed the same steps detailed in section 4.6 to  
652 curate the transcriptomes of the two species; however, we only included coding sequences of  
653 W-candidates and their Z-homologs in *S. japonicum* obtained in section 4.5 (the two curation  
654 pipelines are outlined in Sup. Figure 15). We downloaded the coding sequences of  
655 *Clonorchis sinensis* from the WormBase Parasite database and used it as an outgroup. As  
656 OrthoFinder takes only protein sequences as input, we used an in-house perl script  
657 "GetLongestAA\_v1\_July2020.pl" to perform 6-frame translation of all the transcripts and  
658 retain only the longest isoforms. Orthofinder was then run using the three sets of protein  
659 sequences to assign proteins to clusters of homologs ("orthogroups"). The output files  
660 "orthogroups.tsv" and "Orthogroups\_UnassignedGenes.tsv" are provided in Sup. Dataset 17,  
661 and the transcriptomes curated specifically for OrthoFinder can be found in Sup. Dataset 18.



## 662 4.10 Functional annotation of W-candidates

663 Protein sequences were extracted from the longest open reading frame of each W-  
664 candidate with “GetLongestAA\_v1\_July2020.pl”, and their functional annotation was  
665 performed using the web-based version of PANNZER2 (Protein ANnotation with Z-scoRE)  
666 (Törönen et al. 2018). The annotations and gene ontology predictions are provided in Sup.  
667 Datasets 13 and 14.

668 **Code availability:** All supplementary codes are available in supplementary materials (Sup.  
669 Codes 1 to 14), as well as on: [https://github.com/Melkrewi/Schisto\\_project](https://github.com/Melkrewi/Schisto_project). Supplementary  
670 datasets are available at: <https://seafiler.ist.ac.at/d/2d9ce33a4e0c45aeadd1/>.

671 **Data availability:** All analyses were performed on previously published and publicly  
672 available data: accession numbers are provided as Sup. Table 1.

673 **Acknowledgements:** The authors thank IT support at IST Austria for providing an optimal  
674 environment for bioinformatic analyses.

675 **Funding:** This work was supported by an Austrian Science Foundation FWF grant (project  
676 P28842) to Beatriz Vicoso.

677 **Competing interests:** The authors declare no competing interests.

## 678 References

679 Bachtrog D. 2013. Y-chromosome evolution: Emerging insights into processes of Y-chromosome  
680 degeneration. *Nat Rev Genet* 14(2), 113-124.

681 Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y  
682 chromosome in *Drosophila miranda*. *Genome Biol* 9(2), 1-10.

683 Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, Koutseva N, Zaghul S,  
684 Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive  
685 regulators. *Nature* 508(7497), 494-499.

686 Bellott DW, Page DC. 2020. Dosage-sensitive functions in embryonic development drove the survival  
687 of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Research* 31(2), 198-  
688 210.

689 Bellott DW, Skaletsky H, Cho TJ, Brown L, Locke D, Chen N, Galkina S, Pyntikova T, Koutseva N,

- 690 Graves T, et al. 2017. Avian W and mammalian Y chromosomes convergently retained dosage-  
691 sensitive regulators. *Nat Genet* 49(3), 387.
- 692 Bergero R, Forrest A, Kamau E, Charlesworth D. 2007. Evolutionary strata on the X chromosomes of  
693 the dioecious plant *Silene latifolia*: Evidence from new sex-linked genes. *Genetics* 175(4), 1945-1954.
- 694 Bernardo Carvalho A, Clark AG. 2013. Efficient identification of Y chromosome sequences in the  
695 human and drosophila genomes. *Genome Res* 23(11), 1894-1907.
- 696 Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome Res* 14(5), 988-995.
- 697 Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre  
698 O, Barillot E. 2012. Control-FREEC: A tool for assessing copy number and allelic content using next-  
699 generation sequencing data. *Bioinformatics* 28(3), pp.423-425.
- 700 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data.  
701 *Bioinformatics* 30(15), 2114-2120.
- 702 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq  
703 quantification. *Nat Biotechnol* 34(5), 525-527.
- 704 Bull JJ. 1983. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing  
705 Company Menlo Park CA.
- 706 Bushnell B. 2014. BBLMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.  
707 (LBNL), Berkeley, CA (United States).
- 708 Campos JL, Qiu S, Guirao-Rico S, Bergero R, Charlesworth D. 2017. Recombination changes at the  
709 boundaries of fully and partially sex-linked regions between closely related *Silene* species pairs.  
710 *Heredity* 118(4), 395-403.
- 711 Charlesworth D. 2017. Evolution of recombination rates between sex chromosomes. *Philos Trans R  
712 Soc B Biol Sci* 372(1736), 20160456.
- 713 Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex  
714 chromosomes. *Heredity* 95(2), 118-128.
- 715 Le Clec'h W, Chevalier FD, McDew-White M, Allan F, Webster BL, Gouvras AN, Kinunghi S,  
716 Tchuem Tchuenté LA, Garba A, Mohammed KA, et al. 2018. Whole genome amplification and  
717 exome sequencing of archived schistosome miracidia. *Parasitology* 145(13), 1739-1747.
- 718 Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H.  
719 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508(7497), 488-  
720 493.
- 721 Cotter DJ, Brotman SM, Wilson Sayres MA. 2016. Genetic diversity on the human X chromosome  
722 does not support a strict pseudoautosomal boundary. *Genetics* 203(1), 485-492.
- 723 Criscione CD, Valentim CLL, Hirai H, LoVerde PT, Anderson TJC. 2009. Genomic linkage map of  
724 the human blood fluke *Schistosoma mansoni*. *Genome Biol* 10(6), 1-13.
- 725 Crowson D, Barrett SCH, Wright SI. 2017. Purifying and positive selection influence patterns of gene  
726 loss and gene expression in the evolution of a plant sex chromosome system. *Mol Biol Evol* 34(5),  
727 1140-1154.
- 728 Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics.

- 729 Genome Biol 20(1), 1-14.
- 730 Felsenstein J. 1973. Maximum likelihood estimation of evolutionary trees from continuous characters.  
731 Am J Hum Genet 25(5), 471.
- 732 Fraïsse C, Puixeu Sala G, Vicoso B. 2019. Pleiotropy modulates the efficacy of selection in  
733 *Drosophila melanogaster*. Mol Biol Evol 36(3), 500-515.
- 734 Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE, Fraser  
735 B. 2020. Sex chromosome evolution: So many exceptions to the rules. Genome Biol Evol 12(6), 750-  
736 763.
- 737 Gammerding WJ, Conte MA, Acquah EA, Roberts RB, Kocher TD. 2014. Structure and decay of a  
738 proto-Y region in Tilapia, *Oreochromis niloticus*. BMC Genomics 15(1), 1-9.
- 739 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,  
740 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data  
741 without a reference genome. Nat Biotechnol 29(7), 644.
- 742 Grossman AI, Short RB, Cain GD. 1981. Karyotype evolution and sex chromosome differentiation in  
743 schistosomes (Trematoda, Schistosomatidae). Chromosoma 84(3), 413-430.
- 744 Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by  
745 maximum likelihood. Syst Biol 52(5), 696-704.
- 746 Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. 2014. Genetic degeneration of old and  
747 young Y chromosomes in the flowering plant *Rumex hastatulus*. Proc Natl Acad Sci USA 111(21),  
748 7713-7718.
- 749 Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, Done J, Down T, Gao S, Grove C, et al. 2016.  
750 WormBase 2016: Expanding to enable helminth genomic research. Nucleic Acids Res 44(D1), D774-  
751 D780.
- 752 Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. WormBase ParaSite – a comprehensive  
753 resource for helminth genomics. Mol Biochem Parasitol 215, 2-10.
- 754 Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res 9(9), 868-877.
- 755 Jeffares DC, Tomiczek B, Sojo V, dos Reis M. 2015. A beginners guide to estimating the non-  
756 synonymous to synonymous rate ratio of all protein-coding genes in a genome. In: Parasite Genomics  
757 Protocols. Humana Press, New York (NY). p. 65-90.
- 758 Kabir M, Barradas A, Tzotzos GT, Hentges KE, Doig AJ. 2017. Properties of genes essential for  
759 mouse development. PLoS One 12(5), e0178273.
- 760 Kaiser VB, Zhou Q, Bachtrog D. 2011. Nonrandom gene loss from the *Drosophila miranda* neo-Y  
761 chromosome. Genome Biol Evol 3, 1329-1337.
- 762 Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. 1993. The conserved pre-mRNA splicing factor  
763 U2AF from *Drosophila*: Requirement for viability. Science 262(5133), 569-573.
- 764 Katsuma S, Kiuchi T, Kawamoto M, Fujimoto T, Sahara K. 2018. Unique sex determination system  
765 in the silkworm, *Bombyx mori*: Current status and beyond. Proc Japan Acad Ser B Phys Biol Sci  
766 94(5), 205-216.
- 767 Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. Genome Res 12(4), 656-664.

- 768 Kerins JA, Hanazawa M, Dorsett M, Schedl T. 2010. PRP-17 and the pre-mRNA splicing pathway  
769 are preferentially required for the proliferation versus meiotic development decision and germline sex  
770 determination in *Caenorhabditis elegans*. *Dev Dyn* 239(5), 1555-1572.
- 771 Kimura M. 1987. Molecular evolutionary clock and the neutral theory. *J Mol Evol* 26(1), 24-33.
- 772 Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, Ishihara G, Kawaoka S, Sugano S,  
773 Shimada T, et al. 2014. A single female-specific piRNA is the primary determiner of sex in the  
774 silkworm. *Nature* 509(7502), 633-636.
- 775 Kunz W. 2001. Schistosome male-female interaction: induction of germ-cell differentiation. *Trends*  
776 *Parasitol* 17(5), 227-231.
- 777 Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Science* 286(5441),  
778 964-967.
- 779 Lahn BT, Pearson NM, Jegalian K. 2001. The human Y chromosome, in the light of evolution. *Nat*  
780 *Rev Genet* 2(3), 207-216.
- 781 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4), 357.
- 782 Lawton SP, Hirai H, Ironside JE, Johnston DA, Rollinson D. 2011. Genomes and geography:  
783 Genomic insights into the evolution and phylogeography of the genus *Schistosoma*. *Parasites and*  
784 *Vectors* 4(1), 1-11.
- 785 Lemaitre C, Braga MD V., Gautier C, Sagot M-F, Tannier E, Marais GAB. 2009. Footprints of  
786 inversions at present and past pseudoautosomal boundaries in human sex chromosomes. *Genome Biol*  
787 *Evol* 1, 56-66.
- 788 Lepesant JMJ, Cosseau C, Boissier J, Freitag M, Portela J, Climent D, Perrin C, Zerlotini A, Grunau  
789 C. 2012. Chromatin structural changes around satellite repeats on the female sex chromosome in  
790 *Schistosoma mansoni* and their possible role in sex chromosome emergence. *Genome Biol* 13(2), 1-  
791 15.
- 792 Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and  
793 annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1), W242-W245.
- 794 Li S, Ajimura M, Chen Z, Liu J, Chen E, Guo H, Tadapatri V, Reddy CG, Zhang J, Kishino H, et al.  
795 2018. A new approach for comprehensively describing heterogametic sex chromosomes. *DNA Res*  
796 25(4), 375-382.
- 797 Liao BY, Scott NM, Zhang J. 2006. Impacts of gene essentiality, expression pattern, and gene  
798 compactness on the evolutionary rate of mammalian proteins. *Mol Biol Evol* 23(11), 2072-2080.
- 799 Liu S, Piao X, Hou N, Cai P, Ma Y, Chen Q. 2020. Duplex real-time pcr for sexing *Schistosoma*  
800 *japonicum* cercariae based on w chromosome-specific genes and its applications. *PLoS Negl Trop Dis*  
801 14(8), e0008609.
- 802 LoVerde PT, Niles EG, Osman A, Wu W. 2004. *Schistosoma mansoni* male-female interactions. *Can*  
803 *J Zool* 82(2), 357-374.
- 804 Maciel LF, Morales-Vicente DA, Verjovski-Almeida S. 2020. Dynamic expression of long non-  
805 coding RNAs throughout parasite sexual and neural maturation in *Schistosoma japonicum*. *Non-*  
806 *coding RNA* 6(2), 15.

- 807 Mahajan S, Bachtrog D. 2017. Convergent evolution of Y chromosome gene content in flies. *Nat*  
808 *Commun* 8(1), 1-13.
- 809 Marco A, Kozomara A, Hui JHL, Emery AM, Rollinson D, Griffiths-Jones S, Ronshaugen M. 2013.  
810 Sex-biased expression of microRNAs in *Schistosoma mansoni*. *PLoS Negl Trop Dis* 7(9), e2402.
- 811 Moghadam HK, Pointer MA, Wright AE, Berlin S, Mank JE. 2012. W chromosome expression  
812 responds to female-specific selection. *Proc Natl Acad Sci USA* 109(21), 8207-8211.
- 813 Morris J, Darolti I, Bloch NI, Wright AE, Mank JE. 2018. Shared and species-specific patterns of  
814 nascent Y chromosome evolution in two guppy species. *Genes* 9(5), 238.
- 815 Nam K, Ellegren H. 2008. The chicken (*Gallus gallus*) Z chromosome contains at least three  
816 nonlinear evolutionary strata. *Genetics* 180(2), 1131-1136.
- 817 Nei M. 1969. Linkage modifications and sex difference in recombination. *Genetics* 63(3), 681.
- 818 Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu I,  
819 Charlesworth D, Monéger F. 2005. A gradual process of recombination restriction in the evolutionary  
820 history of the sex chromosomes in dioecious plants. *PLoS Biol* 3(1), e4.
- 821 Ohno S. 1967. Sex chromosomes and sex-linked genes. Springer Science & Business Media.
- 822 Palmer DH, Rogers TF, Dean R, Wright AE. 2019. How to identify sex chromosomes and their  
823 turnover. *Mol Ecol* 28(21), 4709-4724.
- 824 Picard MAL, Cosseau C, Ferré S, Quack T, Grevelding CG, Couté Y, Vicoso B. 2018. Evolution of  
825 gene dosage on the Z-chromosome of schistosome parasites. *Elife* 7, e35684.
- 826 Picard MAL, Vicoso B, Roquis D, Bulla I, Augusto RC, Arancibia N, Grunau C, Boissier J, Cosseau  
827 C, Mank J. 2019. Dosage compensation throughout the *Schistosoma mansoni* lifecycle: specific  
828 chromatin landscape of the Z chromosome. *Genome Biol Evol* 11(7), 1909-1922.
- 829 Potashkin J, Naik K, Wentz-Hunter K. 1993. U2AF homolog required for splicing in vivo. *Science*  
830 262(5133), 573-575.
- 831 Protasio A V., Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, de Silva N, Velarde GS, Anderson  
832 TJC, Clark RC, et al. 2012. A systematically improved high quality genome and transcriptome of the  
833 human blood fluke *Schistosoma mansoni*. *PLoS Negl Trop Dis* 6(1), e1455.
- 834 Pucholt P, Wright AE, Conze LL, Mank JE, Berlin S. 2017. Recent sex chromosome divergence  
835 despite ancient dioecy in the willow *Salix viminalis*. *Mol Biol Evol* 34(8), 1991-2001.
- 836 Rangavittal S, Harris RS, Cechova M, Tomaszewicz M, Chikhi R, Makova KD, Medvedev P. 2018.  
837 RecoverY: K-mer-based read classification for Y-chromosome-specific sequencing and assembly.  
838 *Bioinformatics* 34(7), 1125-1131.
- 839 Rangavittal S, Stopa N, Tomaszewicz M, Sahlin K, Makova KD, Medvedev P. 2019. DiscoverY: a  
840 classifier for identifying Y chromosome sequences in male assemblies. *BMC Genomics* 20(1), 1-11.
- 841 Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome - Patterns  
842 and consequences. *Mol Ecol* 22(11), 3014-3027.
- 843 Schultheiß R, Viitaniemi HM, Leder EH. 2015. Spatial dynamics of evolving dosage compensation in  
844 a young sex chromosome system. *Genome Biol Evol* 7(2), 581-590.

- 845 Short RB, Grossman AI. 1981. Conventional Giemsa and C-banded karyotypes of *Schistosoma*  
846 *mansoni* and *S. rodhaini*. J Parasitol 67(5), 661-671.
- 847 Sigeman H, Ponnikas S, Chauhan P, Dierickx E, De Brooke M, Hansson B. 2019. Repeated sex  
848 chromosome evolution in vertebrates supported by expanded avian sex chromosomes. Proc R Soc B  
849 Biol Sci 286(1916), 20192051.
- 850 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing  
851 genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19),  
852 3210-3212.
- 853 Smeds L, Warmuth V, Bolivar P, Uebbing S, Burri R, Suh A, Nater A, Bureš S, Garamszegi LZ,  
854 Hogner S, et al. 2015. Evolutionary analysis of the female-specific avian W chromosome. Nat  
855 Commun 6(1), 1-10.
- 856 Spence AM, Coulson A, Hodgkin J. 1990. The product of fem-1, a nematode sex-determining gene,  
857 contains a motif found in cell cycle control proteins and receptors for cell-cell interactions. Cell 60(6),  
858 981-990.
- 859 Starostina NG, Lim J min, Schvarzstein M, Wells L, Spence AM, Kipreos ETT. 2007. A CUL-2  
860 ubiquitin ligase containing three FEM proteins degrades TRA-1 to regulate *C. elegans* sex  
861 determination. Dev Cell 13(1), 127-139.
- 862 Tomaszewicz M, Rangavittal S, Cechova M, Sanchez RC, Fescemyer HW, Harris R, Ye D, O'Brien  
863 PCM, Chikhi R, Ryder OA, et al. 2016. A time- and cost-effective strategy to sequence mammalian Y  
864 chromosomes: An application to the de novo assembly of gorilla Y. Genome Res 26(4), 530-540.
- 865 Törönen P, Medlar A, Holm L. 2018. PANNZER2: A rapid functional annotation web server. Nucleic  
866 Acids Res 46(W1), W84-W88.
- 867 Verhulst EC, van de Zande L, Beukeboom LW. 2010. Insect sex determination: It all evolves around  
868 transformer. Curr Opin Genet Dev 20(4), 376-383.
- 869 Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs\_Calculator 2.0: A toolkit incorporating  
870 gamma-series methods and sliding window strategies. Genomics, Proteomics Bioinformatics 8(1), 77-  
871 80.
- 872 Wang J, Chen R, Collins JJ. 2019. Systematically improved in vitro culture conditions reveal new  
873 insights into the reproductive biology of the human parasite *Schistosoma mansoni*. PLoS Biol 17(5),  
874 e3000254.
- 875 Wang J, Na JK, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, et  
876 al. 2012. Sequencing papaya X and Y<sup>h</sup> chromosomes reveals molecular basis of incipient sex  
877 chromosome evolution. Proc Natl Acad Sci USA 109(34), 13710-13715.
- 878 Wang J, Yu Y, Shen H, Qing T, Zheng Y, Li Q, Mo X, Wang S, Li N, Chai R, et al. 2017. Dynamic  
879 transcriptomes identify biogenic amines and insect-like hormonal regulation for mediating  
880 reproduction in *Schistosoma japonicum*. Nat Commun 8(1), 1-13.
- 881 Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015.  
882 Identification and characterization of essential genes in the human genome. Science 350(6264), 1096-  
883 1101.
- 884 Waterhouse RM, Seppely M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V.,  
885 Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and

- 886 phylogenomics. *Mol Biol Evol* 35(3), 543-548.
- 887 White MA, Kitano J, Peichel CL. 2015. Purifying selection maintains dosage-sensitive genes during  
888 degeneration of the threespine stickleback Y chromosome. *Mol Biol Evol* 32(8), 1981-1995.
- 889 Willforss, J., Chawade, A., & Levander, F. 2018. NormalyzerDE: online tool for improved  
890 normalization of omics expression data and high-sensitivity differential expression analysis. *Journal*  
891 *of proteome research*, 18(2), 732-740.
- 892 Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, Kolm N, Breden F, Vicoso B,  
893 Mank JE. 2017. Convergent recombination suppression suggests role of sexual selection in guppy sex  
894 chromosome formation. *Nat Commun* 8(1), 1-10.
- 895 Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al. 2014.  
896 SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics*  
897 30(12), 1660-1666.
- 898 Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, Zhang G, Blom MPK, Christidis L, Prost S, et al.  
899 2019. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds.  
900 *Nat Ecol Evol* 3(5), 834-844.
- 901 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8), 1586-  
902 1591.
- 903 Zhou Q, Bachtrog D. 2012. Chromosome-wide gene silencing initiates Y degeneration in *Drosophila*.  
904 *Curr Biol* 22(6), 522-525.
- 905 Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MTP, Zhang G. 2014. Complex  
906 evolutionary trajectories of sex chromosomes across bird taxa. *Science* 346(6215), 1246338.
- 907 Zhu L, Zhao J, Wang J, Hu C, Peng J, Luo R, Zhou C, Liu J, Lin J, Jin Y, et al. 2016. MicroRNAs are  
908 involved in the regulation of ovary development in the pathogenic blood fluke *Schistosoma*  
909 *japonicum*. *PLoS Pathog* 12(2), e1005423.