Received: 10 October 2021 | Revised: 5 July 2022 | Accepted: 11 July 2022

DOI: 10.1111/1755-0998.13689

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Chromosome-level genome assembly defines female-biased genes associated with sex determination and differentiation in the human blood fluke *Schistosoma japonicum*

Xindong Xu[1] | Yifeng Wang[2] | Changhong Wang[1] | Gangqiang Guo[1] | Xinyu Yu[3] | Yang Dai[3] | Yaobao Liu[3] | Guiying Wei[1] | Xiaohui He[1] | Ge Jin[4] | Ziqiu Zhang[4] | Qingtian Guan[5] | Arnab Pain[5] | Shengyue Wang[6] | Wenbao Zhang[7] | Neil D. Young[8] | Robin B. Gasser[8] | Donald P. McManus[9] | Jun Cao[3,10] | Qi Zhou[2,11,12] | Qingfeng Zhang[1]

[1]Laboratory of Molecular Parasitology, Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Tongji Hospital, and Clinical Center for Brain and Spinal Cord Research School of Medicine, School of Medicine, Tongji University, Shanghai, China

[2]MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China

[3]National Health Commission Key Laboratory of Parasitic Disease Control and Prevention, Jiangsu Provincial Key Laboratory on Parasite and Vector Control Technology, Jiangsu Institute of Parasitic Diseases, Wuxi, China

[4]Novogene Bioinformatics Institute, Beijing, China

[5]Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia

[6]National Research Center for Translational Medicine, State Key Laboratory of Medical Genomics, Ruijin Hospital Affiliated to Shanghai Jiao Tong University (SJTU) School of Medicine, Shanghai, China

[7]State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Medical Research Institute, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China

[8]Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[9]Department of Immunology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[10]Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

[11]Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria

[12]Center for Reproductive Medicine, the Second Affiliated Hospital School of Medicine and Life Sciences Institute, Zhejiang University, Hangzhou, China

Correspondence
Donald P. McManus, Department of Immunology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia.
Email: don.mcmanus@qimrberghofer.edu.au

Jun Cao, National Health Commission Key Laboratory of Parasitic Disease Control and Prevention, Jiangsu Provincial Key Laboratory on Parasite and Vector Control Technology, Jiangsu Institute of Parasitic Diseases, Wuxi 214064, China.
Email: caojuncn@hotmail.com

Qi Zhou, MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang

Abstract

Schistosomiasis is a neglected tropical disease of humans caused by blood flukes of the genus *Schistosoma*, the only dioecious parasitic flatworm. Although aspects of sex determination, differentiation and reproduction have been studied in some *Schistosoma* species, almost nothing is known for *Schistosoma japonicum*, the causative agent of schistosomiasis japonica. This mainly reflects the lack of high-quality genomic and transcriptomic resources for this species. As current genomes for *S. japonicum* are highly fragmented, we assembled and report a chromosome-level reference genome (seven autosomes, the Z-chromosome and partial W-chromosome), achieving a substantially enhanced gene annotation. Utilizing this genome, we discovered that the

Xindong Xu and Yifeng Wang contributed equally to this work.

Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou 310058, China.
Email: zhouqi1982@zju.edu.cn

Qingfeng Zhang, Laboratory of Molecular Parasitology, Key Laboratory of Spine and Spinal Cord Injury Repair and Regeneration of Ministry of Education, Tongji Hospital, and Clinical Center for Brain and Spinal Cord Research School of Medicine, School of Medicine, Tongji University, Shanghai 200092, China.
Email: qfzhang@tongji.edu.cn

Funding information
Australian Research Council; European Research Council Starting Grant, Grant/Award Number: 677696; National Health and Medical Research Council of Australia; National Key R&D Program of China, Grant/Award Number: 2018YFA0507300 and 2020YFC1200105; National Natural Science Foundation of China, Grant/Award Number: 81630063, 81971959, 81971967, 31671319 and 31722050; Natural Science Foundation of Zhejiang Province, Grant/Award Number: LD19C190001

Handling Editor: Isheng J. Tsai

sex chromosomes of *S. japonicum* and its congener *S. mansoni* independently suppressed recombination during evolution, forming five and two evolutionary strata, respectively. By exploring the W-chromosome and sex-specific transcriptomes, we identified 35 W-linked genes and 257 female-preferentially transcribed genes (FTGs) from our chromosomal assembly and uncovered a signature for sex determination and differentiation in *S. japonicum*. These FTGs clustering within autosomes or the Z-chromosome exhibit a highly dynamic transcription profile during the pairing of female and male schistosomula, thereby representing a critical phase for the maturation of the female worms and suggesting distinct layers of regulatory control of gene transcription at this development stage. Collectively, these data provide a valuable resource for further functional genomic characterization of *S. japonicum*, shed light on the evolution of sex chromosomes in this highly virulent human blood fluke, and provide a pathway to identify novel targets for development of intervention tools against schistosomiasis.

KEYWORDS
genome assembly, *Schistosoma japonicum*, schistosomiasis, sex chromosomes, sex differentiation

1 | INTRODUCTION

Schistosomiasis is a disease caused by blood flukes of the genus *Schistosoma* (phylum Platyhelminthes; family Schistosomatidae). Schistosomiasis is a highly significant, neglected tropical disease that remains a serious public health problem mainly in tropical and subtropical areas (World Health Organization, 2010). Currently, >250 million people in 78 countries are affected (McManus et al., 2018). The People's Republic of China has achieved considerable success in preventing human infections by *Schistosoma japonicum* (the Asian schistosome; Cao et al., 2020). In some parts of sub-Saharan Africa, efforts to control the two other most common disease-causing species, *S. mansoni* and *S. haematobium*, through large-scale preventive chemotherapy with praziquantel, have also reduced morbidity and improved human health in underprivileged communities (Webster et al., 2014). However, there is evidence that schistosomiasis is more prevalent than previously recognized (Colley et al., 2017). For instance, schistosomiasis has re-emerged in some endemic areas where control efforts had been successful, or has emerged due to the transmission of zoonotic or hybrid variants of schistosomes in nonendemic areas (Noel et al., 2018; Rudge et al., 2013). Moreover, reduced praziquantel efficacy observed in human populations under high preventive chemotherapy pressure warrants attention in relation to control (Crellen et al., 2016). These issues emphasize the complexities associated with the epidemiology and control of schistosomiasis, and the need to improve understanding of fundamental aspects of the biology (e.g., hybridization) and reproduction of schistosomes, in order to achieve improved prevention, treatment and effective control.

In contrast to most trematodes (flukes), which are hermaphrodites, schistosomes are dioecious with dimorphic adults (Mone & Boissier, 2004). Schistosomes have seven pairs of autosomes and one pair of sex chromosomes (ZZ and ZW); the female worms are heterogametic (ZW) and males are homogametic (ZZ; Basch, 1990; Kunz, 2001; Loker & Brant, 2006; Popiel et al., 1984). Although dimorphic at the adult stage, the sexes of the larval stages (eggs, miracidia, sporocysts, cercariae and schistosomula) are morphologically indistinguishable.

The acquisition of separate sexes in the schistosomes during evolution seems to be concomitant with their invasion of warm-blooded animals (Loker & Brant, 2006). However, the separation and establishment of sex chromosomes appear to be insufficient for complete female development. A unique feature of schistosome biology is that the female worm needs to be permanently coupled with a male for gonad differentiation, sexual maturation and effective reproduction (Beckmann et al., 2010; Kunz, 2001). In the paired adult female, the ovary produces oocytes, and the vitellarium provides the vitelline cells and egg-shell proteins needed for embryogenesis and egg production (Beckmann et al., 2010; Kunz, 2001). The eggs are critical for maintaining the life cycle and, in the human or other mammalian host, cause granulomatous disease (schistosomiasis) usually in enterohepatic or urogenital tissues (Ross et al., 2002).

Due to the medical importance and unique sexual biology of schistosomes, there has been extensive investigation of their

genomes and transcriptomes. Reference genomes have been published for eight distinct species of *Schistosoma*, including the three main human-infecting schistosome species (*S. japonicum*, *S. mansoni* and *S. haematobium*; Berriman et al., 2009; International Helminth Genomes Consortium, 2019; Luo et al., 2019; Oey et al., 2019; *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009; Young et al., 2012). Assemblies with chromosomal or near-chromosomal contiguity have been achieved for the genomes of *S. mansoni* and *S. haematobium* (Protasio et al., 2012; Stroehlein et al., 2019), but not for other species. Recently, the assembly of the genome of *S. japonicum* was improved by combining PacBio long sequencing with Illumina paired-end reads (Luo et al., 2019). This new assembly (V2, BioProject PRJNA520774) reflects a genome size of ~370.5 Mb, with contig and scaffold N50 lengths of 871.9 kb and 1.09 Mb, respectively, representing 142.4- and 6.2-fold improvements over the earlier-released assembly (V1, BioProject PRJEA34885; *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009); however, the V2 genome assembly still lacks considerable chromosome information.

Here, we utilized PacBio, 10× Genomics and Hi-C as well as Illumina sequencing data from a clonal line of *S. japonicum* to generate a chromosome-level genome for female *S. japonicum*. The improved genome allowed us to comprehensively probe the evolutionary strata on the Z-chromosome, identify genes on the W-chromosome and accurately map the genes preferentially transcribed in female worms to autosomes. Our improved assembly and annotation provide a valuable genomic resource for future functional genomic analysis of *S. japonicum* and for studies of the emergence of sex chromosomes in schistosomes.

2 | MATERIALS AND METHODS

2.1 | Procurement of *S. japonicum*

Eggs of *Schistosoma japonicum* were isolated from the livers of infected mice and then hatched in mineral water under light (Xu & Dresden, 1990). To establish clonal *S. japonicum*, individual *Oncomelania hupensis hupensis* snails were exposed to individual miracidia in 24-well cell culture plates (2 ml mineral water per well) overnight at room temperature. After cultivation for 3 months at 25°C on moist straw paper in enamel-coated dishes, each snail was examined individually for cercarial release under a bright white light (Theron et al., 1997). Clonal cercariae were collected from each snail, and the sex of the cloned cercariae was determined by PCR using female-specific primers (Figure S1; Zhao et al., 2010). Female BALB/c mice (8–10 weeks of age) were each infected with both 10 clonal female cercariae and 10 clonal male cercariae. Adult worms of *S. japonicum* were collected by perfusion from the mesenteric veins surrounding the gut at 56 dpi (days post-infection), and eggs from female worms were used to establish inbred lines of *S. japonicum* (Xu & Dresden, 1990). All animal experimentation was approved by the Internal Review Board of Tongji University, School of Medicine, China.

2.2 | Nucleic acid isolation and sequencing

After two rounds of inbreeding, an inbred line—designated SjF4M4—was used for genomic DNA and total RNA extractions. Genomic DNA (10 μg for each sample) was isolated separately from male and female adult worms using a MagAttract HMW DNA Kit (Qiagen), according to the manufacturer's instructions. The DNA was quantified using a Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific) and a NanoDrop One Spectrophotometer (Thermo Fisher Scientific), and then examined for quality by agarose gel electrophoresis and by using a Bioanalyzer 2100 (Agilent Technologies) prior to sequencing using PacBio and 10× Genomics using standardized protocols (Eid et al., 2009; Zheng et al., 2016). Gravid adult females will harbour some eggs of the opposite sex, indicating the genomic DNA of adult female worms contains male egg DNA. However, a small amount of male eggs in female worms can be ignored as female worms are heterogametic (ZW) and males are homogametic (ZZ). Total RNA (5 μg for each stage) was isolated separately from eggs, cercariae, and female and male worms, using the Qiagen RNeasy Midi Kit (Qiagen). Paired-end (PE) RNA-seq libraries (insert sizes: 200 and 500 bp) were constructed using the TruSeq RNA Library Preparation Kit version 2 and sequenced employing the Illumina platform.

2.3 | Genome assembly

The *S. japonicum* genome size was estimated based on *k*-mer analysis ($k = 17$) of sequence data using KMERGENIE (Chikhi & Medvedev, 2014). Murine host DNA contamination was removed before scaffolding using BOWTIE 2 version 2.3.3.1 (Langmead & Salzberg, 2012). All raw reads obtained by PacBio sequencing were aligned to each other using DALIGNER, executed using the FALCON (version 0.7) assembler (Chin et al., 2016). Overlapping reads and raw subreads were processed to produce consensus sequences. After error-correction, overlaps were detected in the pre-assembled, error-corrected reads and used to construct a directed fragment assembly string graph. Contigs were constructed by finding the paths from the string graph. Consensus calling of the preceding assembly was performed using the consensus-calling algorithm Quiver (Chin et al., 2013). Subsequently, the PE, clean Illumina reads were used for further error-correction of the assembly via PILON (version 1.22; Walker et al., 2014). The 10× Genomics–scaffold extension was performed using FRAGSCAFF software (Adey et al., 2014), in which the linked reads produced from the 10× genomic library were aligned to the consensus sequence of the PacBio assembly. To obtain superscaffolds, only consensus sequence with linked-read support was used for assembly. The genomic Illumina data were aligned to the assembled reference genome for *S. japonicum* (V3) using BWA-MEM software (version 0.7.17; Li, 2013). Core Eukaryotic Genes Mapping Approach (CEGMA; Parra et al., 2007) and Benchmarking Universal Single Copy Orthologs (BUSCO version 3; Simao et al., 2015) were used to estimate genome completeness. Murine host genome contamination was analysed using BLOBTOOLS2 (https://blobtoolkit.genomehubs.org/blobtools2/).

2.4 | Scaffolding of the assembly using Hi-C data to reach chromosome contiguity

Paired adult female and male worms of *S. japonicum* ($n = 80$) were collected, washed three times in phosphate-buffered saline (PBS) and fixed in 2% formaldehyde/PBS solution for 30 min at room temperature. Fixed worms were homogenized in liquid nitrogen, resuspended in nuclei isolation buffer and filtered using a 40-nm cell strainer. Parasite membranes were further disrupted by passing the lysate through a 30-gauge needle using a syringe. The extracted nuclei were digested with 100 U *Dpn*II (NEB, R0543) restriction enzyme by incubating overnight at 37°C, and the digested chromatin was blunt-ended with 1 μl of 10 mM dCTP, dTTP, dGTP, 25 μl of 0.4 mM biotin-14-dATP (Invitrogen) and 100 U of Klenow fragment (NEB, M0210) for 3 h at 37°C. Next, the ligation reaction was performed in 10× the volume of ligation production buffer under constant shaking with 50 μl of 10× ligation buffer and 10% Triton X-100, 2.5 μl of 20 mg ml^{-1} bovine serum albumin and 2 μl of T4 DNA ligase (NEB, M0202). The ligation reaction was performed at 16°C overnight. Following ligation, the nuclei were digested with proteinase K at 65°C for 12 h. The DNA was purified and sheared to a size of 200–300 bp using a sonicator (Bioruptor™ UCD-200), and then size-selected using 0.8–1.1× AMpure XP beads (Beckman Coulter, A63881). The 1.1× biotin-tagged Hi-C DNA was subjected to end-repair and then bound to Dynabeads MyOne Streptavidin C1 beads (Invitrogen). The C1 beads with bound DNA were resuspended, and the DNA extended with 3′ A-overhangs and ligated to barcoded NextFlex adapters (Bio Scientific). To prepare sequencing libraries, the purified DNA was end-repaired, extended with 3′ A-overhangs and ligated to barcoded NextFlex adapters (Bio Scientific). Libraries were amplified using KAPA HiFi HotStart ready mix (KAPA Biosystems) employing the following PCR protocol: 98°C for 3 min, followed by 14 cycles of 98°C for 20 s; 65°C for 30 s; 68°C for 30 s, and a final extension of 68°C for 5 min. Amplified libraries were size-selected for 300–600 bp using 2% (w/v) agarose gels and PE sequenced using an Illumina Hiseq X Ten sequencer. The cleaned Hi-C data were aligned to the primary assembly using BWA-MEM 0.7.17 software. Only read-pairs, for which both reads in the pair aligned to contigs, were considered for scaffolding. Sequences of >100 bp were selected by LACHESIS (version 201701; Burton et al., 2013) to scaffold the assembly to the chromosome level.

2.5 | Defining autosomal and sex chromosomes

Illumina reads representing each adult male and adult female *S. japonicum* were mapped separately to the V3 genome assembly of *S. japonicum* using BWA-MEM 0.7.17. The resultant alignments were filtered using SAMTOOLS (Li et al., 2009) to retain only reads with a mapping quality of ≥33. Then, each chromosome and scaffold was divided into 50-kb windows, with 10- or 5-kb windows overlapping by 1 kb to calculate the depth of male and female Illumina sequences separately in each window using SAMTOOLS and BEDTOOLS version 2.27.1 (Quinlan & Hall, 2010). As the insert sizes of the libraries representing female and male *S. japonicum* differed from one another, we selected the longest autosome (chromosome 1) to normalize depth. We estimated depths of 33 (female) and of 44 (male) using the ratio $S_0/2L_0$, where S_0 is the sum of nucleotides that mapped, and L_0 is the length of the chromosome, excluding gaps (Ns). In each window of each chromosome or scaffold, the normalized depth value was calculated using S/dL, where S is the sum of bases that mapped in a window, L is the corresponding window-length after removing the N bases (gaps) and d is the particular absolute sequencing depth observed for a given sex (33 for females; 44 for males). Based on normalized depth, windows with male-to-female depth ratio of 1.5–2 were classified as being Z-linked, whereas windows with male-to-female depth ratios of <0.5 were classified as W-linked. A chromosome/scaffold was assigned as autosomal if the majority of windows (>50%) had a male-to-female depth ratio of ~1.0. A scaffold was assigned as W-chromosomal if the "female" depth, following normalization of the majority of windows, was >0.3, whereas the corresponding normalized "male" depth was less than half of the "female" depth.

2.6 | Chimeric sequence reassembly

As the Z chromosome, as initially defined, contained several W-linked regions, we designated it as a ZW-chromosome. We divided the ZW-chromosome into 5-kb windows to eliminate contamination by W fragments. Windows with a normalized female depth of ~1.0 and a normalized male depth of ~0 were termed W-linked regions, and others were termed Z-linked regions. The same method was employed for scaffolds representing W-linked fragments. Thirteen W-linked regions (Figure S2; W_region_1 to W_region_13) and seven W-linked scaffolds (Figure S2; W_scaf_1 to W_scaf_7) were combined and re-assembled into the W chromosome, and the Z-linked regions were re-assembled into the Z chromosome (Figure S3) using the same approach. The scaffolding used for the reassembly of each sex chromosome was performed using SCAFF10X (https://github.com/wtsi-hpag/Scaff10X), HIC-PRO version 2.10.0 (Servant et al., 2015) and SALSA version 2.0 (Ghurye et al., 2019) using data from the 10× Genomics linked-read library and the Hi-C library. Then, fragments were ordered and oriented in relation to the initial *S. japonicum* V3 ZW-chromosome using RAGOO (Alonge et al., 2019). The final W-linked (ChrW_fragment) sequence assembly of 4.8 Mb does not contain the PAR region.

2.7 | Gene prediction

Genes were predicted by combining de novo, homology-based and transcriptome-based predictions. For the de novo prediction, AUGUSTUS (version 3.2.3; Hoff & Stanke, 2019), GLIMMERHMM (version 3.0.4; Majoros et al., 2004), SNAP (version 2013.11.29; Korf, 2004), GENEID (version 1.2; Alioto et al., 2018) and GENSCAN (version 1.0; Burge & Karlin, 1997) were employed. For the homology-based approach, genes of *S. mansoni* (Buddenborg et al., 2021; Protasio et al., 2012), *S. haematobium* (Stroehlein et al., 2019), *Clonorchis sinensis* (Young

et al., 2021) and *Opisthorchis viverrini* (Young et al., 2014) were used as queries to search against *S. japonicum* V3 using BLAST (Johnson et al., 2008) and GENEWISE (Birney et al., 2004). For the transcriptome-based gene prediction TOPHAT mapper (version 2.0.8), along with CUFFLINKS assembler (version 2.1.1), was employed for mapping (Trapnell et al., 2012). Pseudo-expressed sequence tags were generated from RNA-sequencing data by TRINITY version 2.7.0 (https://github.com/trinityrnaseq/trinityrnaseq/releases) and mapped to the assembled genome. PASA (version 2.3.3; Haas et al., 2003) was employed to predict gene models. The gene prediction results were merged using EVIDENCEMODELER (version 1.1.1) software to generate a consensus gene set. Translated coding sequences were compared with those in known databases such as SWISSPROT (version 20180824), a subset of Uniprot (UniProt Consortium, 2021), the nonredundant NCBI protein database (NR; version 20180716; Sayers et al., 2021), KEGG (version 20160503; https://www.kegg.jp/), INTERPRO (version 5.31-70.0; Blum et al., 2021), GO (Harris et al., 2004) and PFAM (version 31.0; El-Gebali et al., 2019). All protein-coding genes were annotated using information from SWISSPROT, NR, KEGG, INTERPRO, GO and PFAM.

2.8 | Identifying one-to-one orthologues

The *S. mansoni* annotation (schistosoma_mansoni.PRJEA36577.WBPS13.annotations.gff3) and protein (schistosoma_mansoni.PRJEA36577.WBPS13.protein.fa) files were obtained from the WormBase ParaSite database (https://parasite.wormbase.org/index.html, Smansoni_v7, WBPS15, contributed by the Wellcome Sanger Institute; Bolt et al., 2018; Buddenborg et al., 2021). Protein BLAST (BLASTP; Camacho et al., 2009) was performed using predicted proteins encoded by the *S. japonicum* Z chromosome as queries against predicted proteins encoded by the *S. mansoni* ZW-chromosome, with only the longest amino acid isoform sequence used for each *S. japonicum* and *S. mansoni* gene. In the output file, genes on the Z chromosome of *S. japonicum*, whose amino acid sequences aligned <30% or whose alignment identity was <50% were filtered out, retaining only the best matches. The orthologues were assigned to groups based on their genomic locations relating to *S. japonicum*-specific strata (SjapS1, SjapS2, SjapS3, SjapS4), *S. mansoni*-specific strata (SmanS1) and shared strata (SjapS0 and the PAR region).

2.9 | Demarcation of evolutionary strata

We combined the read depth and single nucleotide polymorphism (SNP) density calculated from reads from both sexes of *S. japonicum*, the sequence divergence levels between chromosomes W and Z, and orthology information between *S. japonicum* and *S. mansoni*, to demarcate the evolutionary strata. Based on the depth differences between males and females, we assigned the highly divergent sexually differentiated regions (SDRs) in the genomes of *S. japonicum* V3 and *S. mansoni*, which revealed a 2-fold depth difference between sexes. Based on the one-to-one orthologous relationship between *S. japonicum* and *S. mansoni*, their shared SDR was assigned as the "ancient" stratum (SjapS0), and species-specific SDRs were assigned to two separated strata for *S. japonicum* and SmanS1 for *S. mansoni*. Although the current chrZW assembly of *S. mansoni* may contain W-linked sequences, our coverage analyses confirmed that there were few fragments showing a female-specific read coverage pattern. This suggested that we could still use the chrZW assembly to compare the syntenic relationship between *S. mansoni* and *S. japonicum*. The W chromosome sequences were aligned to the Z chromosome sequence using NUCMER version 3.23 (Kurtz et al., 2004). Sequences whose alignment lengths were <5 kb or alignment identities were <70% were discarded. The aligned W chromosome sequences were oriented along the Z chromosome (colour-coded according to sequence divergence) such that they defined the boundaries or ages between strata; for instance, in our nomenclature, SjapS2 exhibited higher Z/W homology than SjapS1. Illumina reads representing adult female and male *S. japonicum* were mapped separately to the *S. japonicum* V3 genome using BWA-MEM 0.7.17. SNPs were identified from the resultant alignment files using the genomic analysis toolkit (GATK 3.8; McKenna et al., 2010). To estimate Z/W divergence (dS), we extracted female-specific SNPs from genes with SNPGENIE (Nelson et al., 2015). The bilateral regions of *S. japonicum* SjapS2 was assigned as the more "recent" stratum, with the similar levels of male and female read depths, but distinct levels of female SNPs densities relative to male ones. SjapS4 exhibited lower SNPs heterozygosity in females and lower dS value compared with SjapS3, suggesting it is the most "recent" stratum.

2.10 | Prediction of repeats and noncoding RNA

The method of repeat sequences annotation was divided into homologous sequence alignment and de novo prediction. Based on the RepBase library (http://www.girinst.org/repbase/), the homologous sequence alignment method used REPEATMASKER and REPEATPROTEINMASK (http://www.repeatmasker.org/) software to identify sequences similar to known repeat sequences. The de novo prediction method first established the de novo repeat sequence libraries by REPEATMODELER (http://www.repeatmasker.org/RepeatModeler.html) software, and then REPEATMASKER was used for prediction again. In addition, we also used the TRF (http://tandem.bu.edu/trf/trf.html) software to look for tandem repeat sequences in the genome. Based on the above method, we provide four repeat sequences annotation files, namely Sjap.known.repeat.gff, Sjap.denovo.repeat.gff, Sjap.protein.repeat.gff and Sjap.tandem.repeat.gff.

Noncoding RNA annotations contained transfer RNA (tRNA), ribosomal RNA (rRNA), micro-RNA (miRNA) and small nuclear (snRNA). According to the structure characteristics of tRNA, TRNASCAN-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) software was used to find tRNA in the genome sequence. As rRNA is highly conserved, rRNA sequences of closely related species can be selected as reference sequences to search for rRNA in the genome by the BARRNAP software (https://github.com/nickp60/barrnap). Using the

Rfam family model, the INFERNAL (http://eddylab.org/infernal/) software predicted the miRNA and snRNA sequence information in the genome.

2.11 | SMRT-based RNA sequencing

Total RNAs were extracted from eggs, cercariae, schistosomula, and adult female and male *S. japonicum* parasites. They were then equally mixed for library preparation. The Iso-Seq library was prepared according to the Isoform Sequencing protocol (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). Sequence data were processed using the SMRTLINK 5.0 software. Circular consensus sequence (CCS) was generated from subread BAM files, with parameters: min_length 200, max_drop_fraction 0.8, no_polish TRUE, min_zscore −9999, min_passes 1, min_predicted_accuracy 0.8, max_length 18,000. CCS.BAM files were output, which were then classified into full-length and nonfull-length reads using pbclassify.py script, ignore polyA false, minSeq Length 200. Nonfull-length and full-length fasta files produced were then fed into the cluster step, which performs isoform-level clustering (ICE), followed by final ARROW polishing, with hq_quiver_min_accuracy 0.99, bin_by_primer false, bin_size_kb 1, qv_trim_5p 100, qv_trim_3p 30. Aligning consensus reads to the reference used GMAP with parameters: --no-chimeras --cross-species --expand-offsets 1 -B 5 -K 50000 -f samse -*n* 1 against the newly assembled genome in this study.

2.12 | Illumina sequencing-based transcriptome analyses

Total RNAs were extracted from eggs, cercariae, schistosomula, and adult female and male *S. japonicum* parasites. Each RNA sample was subjected to Illumina-based RNA-seq. RNA-seq reads from both sexes at eight time points (14–28 dpi) used in dynamic differential transcriptome analysis between female and male parasites were obtained from GenBank (PRJNA343582; Wang et al., 2017). Biological replicates of RNA-seq data for males and females at each stage were combined with 16 merged libraries, and RNAseq reads were then mapped to the *S. japonicum* genome V3 using HISAT2 version 2.0.4 (Kim et al., 2015). Reads were counted using HTSEQ version 0.11.0 (Anders et al., 2015), and transcription values—in reads per kilobase of transcript per million mapped reads (RPKM)—were calculated for each gene for each library (Mortazavi et al., 2008). All of these values are provided in the source data file (Table S19). Data on dosage compensation (DC) analysis and sex-biased genes analysis were filtered using the same threshold (RPKM >1) for each sex and stage of *S. japonicum*. For W chromosome-linked transcription, the filtration threshold was RPKM-mean >1 (sum of RPKMs of all libraries for the same gene was >16). We used EDGER to estimate gene expression changes between sexes or stages.

2.13 | Gametologue analysis

First, we used BLASTP to identify gene gametologues in the W chromosome of *S. japonicum* and to compare them with the predicted proteomes of *S. mansoni* and *Fasciola hepatica*. Homologues between *S. japonicum* and *S. mansoni* were identified that had a sequence alignment score of >50 and aligned over ≥30% of the query gene sequence, whereas homologues between *S. japonicum* and *F. hepatica* were assigned if they had an alignment score of ≥20, and aligned over ≥10% of the query gene sequence. Following filtration, we reconstructed ML (maximum likelihood) phylogenetic trees using MEGA X (Kumar et al., 2018) to establish the relationships among the groups of gametologues, with nodal support values indicated.

2.14 | Gene ontology analysis

Functional annotation of protein sequences was conducted using BLAST2GO software (Conesa et al., 2005) against the entire nonredundant NCBI NR database using default parameters.

2.15 | Inference of gene clusters with preferential transcription in females

We used the program CROC (Pignatelli et al., 2009), selecting a window of 50 genes, three or more sequential genes representing a cluster, and a *p*-value of <.05. A filtration threshold of RPKM of >1 was used for adult female genes.

3 | RESULTS

3.1 | Chromosome-level genomic assembly of *Schistosoma japonicum*

To reduce heterozygosity, an inbred line (SjF4M4) of *S. japonicum* was established (Figures 1a and S1), and the resultant female worms were used for genome sequencing. A genomic survey sequence based on *k*-mer frequency data revealed an estimated genome size for *S. japonicum* of 418.8 Mb, a repetitive sequence content of 38.5% and heterozygosity of 0.9% (Table S1); this level of heterozygosity for the SjF4M4 line is less than that of the published V2 genome (1.1%; Luo et al., 2019).

To construct a chromosome-scale reference genome for *S. japonicum*, we employed a comprehensive de novo assembly strategy (Figure 1b), combining Illumina PE reads (~60× coverage of the genome), PacBio single-molecule long reads (~105× coverage), 10× Genomics (~363× coverage) and Hi-C sequence data (~243× coverage; Table S2). First, we produced gapless contig sequences using PacBio reads from genomic DNA from female worms. Then, we ordered and oriented the contigs into longer scaffolds employing the linkage information obtained from the 10× and Hi-C data sets.

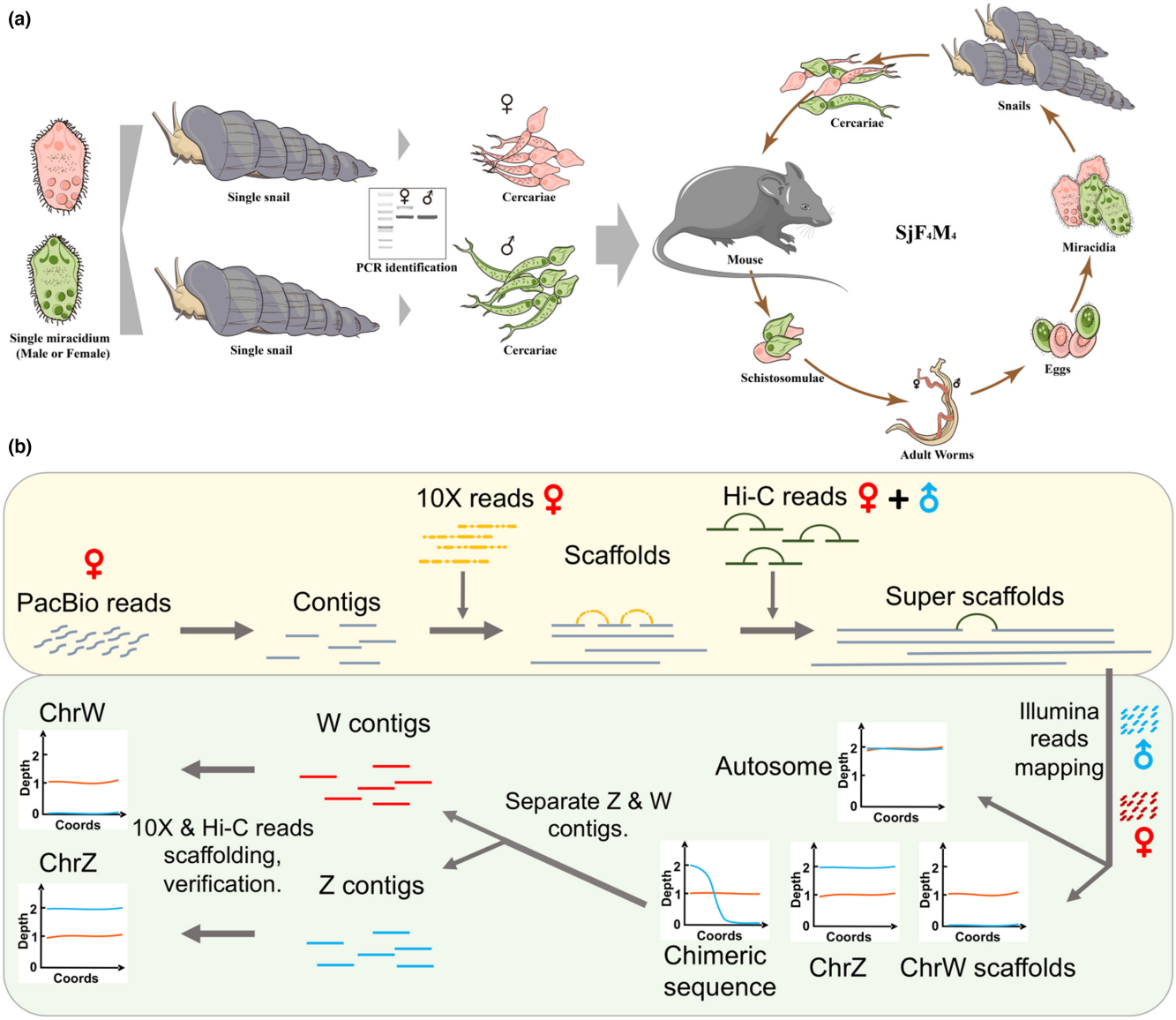

FIGURE 1 Establishment of an inbred line of *Schistosoma japonicum* and the genome assembly pipeline used. (a) Establishment of a clonal line of *S. japonicum*. Individual *Oncomelania hupensis* snails were infected with a single miracidium of *S. japonicum*. Then, the sex of the clonal cercariae shed from each individual snail was determined by PCR (Figure S1). Ten female and 10 male cercariae were used to infect individual mice. Eggs were collected to establish a clonal parasite (SjF4M4) after two rounds of inbreeding; the resultant parasite line was used for DNA and RNA sequencing. (b) A schematic overview of the *S. japonicum* assembly pipeline. Two major aspects are shown: the initial assembly procedure (yellow box) and then sex chromosome validation and reassembly (green box). Female PacBio long reads were used to generate contigs; then, with 10× and Hi-C data linkage information, we generated a chromosome-length genome assembly (yellow box). Using male and female Illumina reads mapping to the assembly, these scaffolds were assigned to autosomal or sexual types based on their normalized depth value. Chimeric sequences were extracted from the Z chromosome, and split into Z and W regions, which were then combined with other Z and W sequences and reassembled into the W chromosome and the new Z chromosome, respectively (green box).

Hi-C scaffolding resulted in eight chromosomes (Figure S4). After gap-filling and assembly-polishing, we achieved a genome assembly (406.6 Mb) with a scaffold N50 of 49.5 Mb and a total of 107 scaffolds. The eight longest scaffolds, spanning 99.4% of the estimated genome size (Table 1), constitute a chromosome-scale genome assembly for *S. japonicum*, consistent with the reported karyotype of $2n = 16$ (*Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009). The genome size of 406.6 Mb is similar to that estimated by *k*-mer analysis (418.8 Mb), and is markedly larger than the reported V2 genome (370.5 Mb; Luo et al., 2019). The extent of SNP or heterozygosity in the reference genome (0.7%; Table S3) was also similar to that estimated in the *k*-mer analysis (0.9%; Table S1). In addition, we performed genome contamination analysis using BLOBTOOLS2, which used BLAST and DIAMOND for alignment with the NCBI nt and Uniprot database at the nucleic acid and protein level, and further verified by read realignment to the assembled genome. Based on sequence similarity to the host and read coverage of the genome sequence, the results indicated that this schistosoma

TABLE 1 Features of the chromosome-level assembled genome of *Schistosoma japonicum* (version V3) compared with the previously published genome (V2, BioProject PRJNA520774) for the same species, *S. mansoni* (V9) and *S. haematobium* (V2, BioProject PRJNA78265)

	S. japonicum V3	*S. japonicum* V2	*S. mansoni* V9[a]	*S. haematobium* V2
Genome size (Mb)	406.6	369.9	391.4	371.4
Number of scaffolds	107	1789	9	666
Number of contigs	628	2108	368	12,937
Longest scaffold (Mb)	90.6	6.3	89.1	14.3
Average scaffold length (bp)	3,799,739	210,145	43,488,369	557,649
Number of scaffolds: >10 kb	98	1052	9	–
Scaffold N50 (bp)	49,539,924	1,093,989	45,716,228	4,779,868
Contig N50 (bp)	2,378,822	871,911	4,473,015	21,744
GC content (%)	34.0	33.8	35.5	34.5
Repeat content (%)	46.3	46.9	54.5	47.2
CEGMA assembly ($n = 248$)	86.3	–	–	–
Putative coding genes (*N*)	10,158[b]	10,089	9794	9314
BUSCO annotation ($n = 1658$)	68.2	–	53.3	63.9
Chromosomes	7+Z+W	NA	7+Z+W	7+Z/W

[a]*S. mansoni* V9 genome data were download from NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/237/925/GCA_000237925.5_SM_V9/).

[b]9760 full-length CDS.

genome assembly was not contaminated with the murine host genome (Table S4).

3.2 | Identification of sex chromosome sequences in *S. japonicum*

To identify autosomal and sex-linked sequences in *S. japonicum*, we also produced Illumina reads (with 80× coverage) using DNA samples from male worms of the SjF4M4 line (Figure 1b). We assumed that equal or similar numbers of female and male reads would map to the autosomes and the pseudo-autosomal regions (PARs) of the sex chromosomes—where homologous recombination occurs (Otto et al., 2011). Furthermore, we anticipated that (i) lower numbers of female than male reads would map to SDRs of the Z chromosome; (ii) a lower number of female reads would map to the SDR of the Z chromosome than to individual autosomes; and (iii) a much higher number of female reads would map to the W chromosome.

With these considerations in mind, we identified seven autosomal sequences, and one sex chromosome sequence, although many regions showed a chimeric assembly of Z and W chromosomes, that is a female-specific read coverage pattern within the adjacent regions showing a similar coverage level between sexes (Figure S3). Chimeric assemblies of Z- and W-linked sequences were also reported for the previously published genomes of *S. japonicum* (Picard et al., 2018), suggesting that some sex-linked regions are not clearly demarcated from each other, thereby confounding the genome assembly. To circumvent this issue, we extracted the W-linked sequence fragments from the genome assembly, according to their female-specific read mapping patterns, and performed separate scaffold assemblies for the Z- or W-linked sequences, again using the 10× and Hi-C linkage data. The final assembly contains one Z-chromosome of 87 Mb, and a W-linked (ChrW) scaffold of 4.8 Mb (Figure 2). Overall, the autosomes, the Z chromosome and the W chromosome constitute 76.0%, 21.4% and 1.2% of the *S. japonicum* genome assembly (version 3, V3), respectively; only 1.4% of the assembled sequences (~5.9 Mb in length) could not be assigned to chromosomes due to their repetitive nature.

3.3 | Genome assembly quality and gene annotation

The present *S. japonicum* genome (V3) is significantly more intact than both previously published versions (V1 and V2; Luo et al., 2019; *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009), as it shows a higher percentage (97.5% for V3 vs. 90.3% for V2) of concordantly mapped Illumina reads (Table S5); 98.6% of V3 had a sequence depth of >10 (Table S5), indicating high accuracy at the nucleotide level. There was a high degree of synteny between versions V2 and V3 (Figure 3a), and between V3 and the *S. mansoni* genome V7 (Figure 3b). The W chromosome sequence matched the combined ChrZ/W sequence for *S. mansoni*. As *S. mansoni* chromosomes were not used to order or orient the scaffolds for

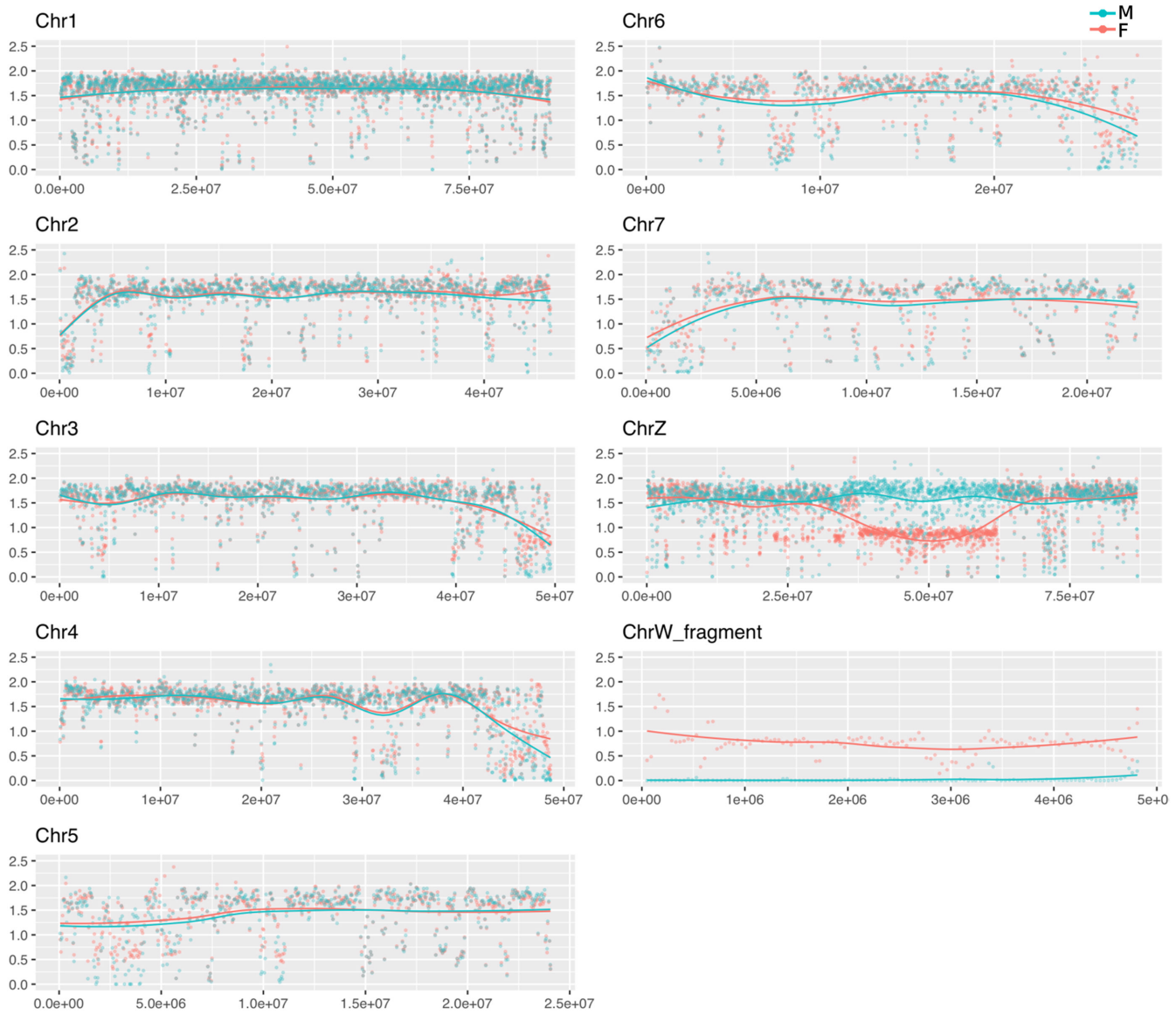

FIGURE 2 Coverage plots of *Schistosoma japonicum* V3 autosomal chromosomes and sex chromosomes Z and W. Sequence read-coverage plots of *S. japonicum* chromosomes. Male (blue dot) and female (red dot) normalized depth values (*y*-axis) along *S. japonicum* chromosomes (*x*-axis) are shown across 50-kb windows.

the *S. japonicum* genome, these findings provide independent evidence that the contig assemblies and scaffolding are accurate.

Approximately 46.0% of the V3 genome of *S. japonicum* represent repetitive sequences, comparable to 46.6% reported previously for the V2 assembly (Table S6). Due to the suppression of recombination (Charlesworth & Charlesworth, 2000), the W chromosome of *S. japonicum* has accumulated more simple repeats and LTR retrotransposons than the Z and autosomal chromosomes (Figure S5). Using de novo-, homology- and transcriptome- (SMRT- or Illumina-based RNA-sequencing) based approaches, we predicted 10,158 protein-coding genes in the *S. japonicum* V3 assembly, which is similar to the number (10,089) estimated for V2 (Tables 1 and S7). CEGMA analysis identified 214 of the 248 core eukaryotic genes (86.3%; Tables 1 and S8). An analysis by BUSCO showed that 62.2% of complete genes (i.e., 58.4% single-copy genes and 3.8% duplicated genes) of *S. japonicum* are represented in its transcriptome, compared with 59.8% for *S. mansoni* and 55.1% for *S. haematobium* (Tables 1 and S9). The average gene and coding DNA sequence (CDS) lengths were 16,941 and 1573 bp, respectively (Table S10). Numerous noncoding RNAs were identified, including 3892 miRNAs, 176 tRNAs, 29 snRNAs and 40 rRNAs (Table S11).

3.4 | "Evolutionary strata" on the Z chromosome

A previous study inferred the presence of PARs and SDRs in *S. japonicum* using the Z chromosome of *S. mansoni* (Figure S6) as a reference (Picard et al., 2018). Due to the Z-linked rearrangements between these two species, and a lack of W chromosome sequences, our understanding of the evolutionary history of schistosome sex

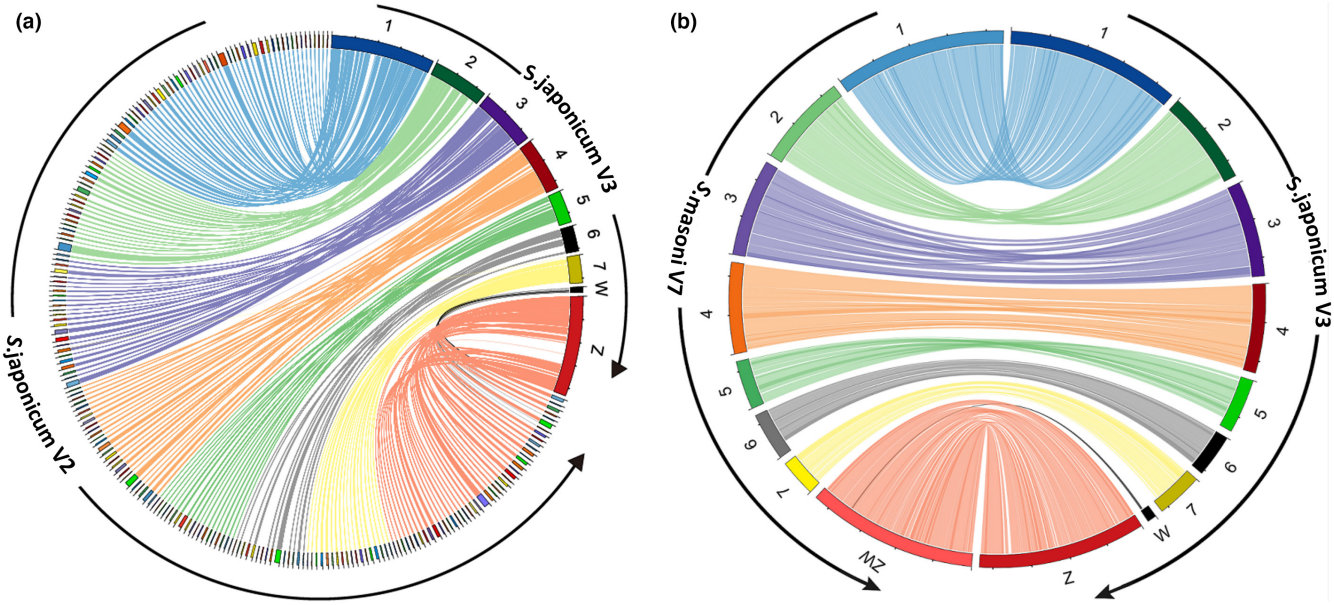

FIGURE 3 Circos plot of synteny between *Schistosoma japonicum* genome versions (V3 and V2) and between *S. japonicum* genome V3 and *S. mansoni* genome version V7. (a) A total of 98.1% of contigs longer than 1 Mb in *S. japonicum* genome V2 (BioProject PRJNA520774) were assembled in *S. japonicum* genome V3. (b) All *S. japonicum* V3 chromosomes matched the corresponding *S. mansoni* chromosomes V7 (BioProject PRJEA36577).

chromosomes remains incomplete. By combining the information on comparative alignment of chrZ between *S. mansoni* and *S. japonicum* (Figure 4a), assembled W-linked sequences and Z/W sequence similarities (Figure 4b), read-coverage and SNP density for each sex of *S. japonicum* (Figure 4c,d), we reconstructed the successive steps of recombination suppression (evolutionary strata; Lahn & Page, 1999; Li et al., 2021; Vicoso et al., 2013; Zhou et al., 2014) for the sex chromosomes of both *S. japonicum* and *S. mansoni*. Evolutionary strata are a sequence pattern between sex chromosomes that reflects stepwise suppression of recombination caused by presumably successive chromosomal inversions. In vertebrates (birds and mammals) this process is recognized as resulting from a stepwise linkage of sexually antagonistic genes to sex-determining genes (Handley et al., 2004; Zhou et al., 2014, 2021). Along the Z-linked SDR that has completely suppressed recombination, we expected a two-fold higher read coverage in males than in females if the W-linked sequences have become fully degenerated, or a female-biased SNP pattern in the Z-linked SDR if the homologous W-linked region has only lost recombination very recently and just started to diverge in sequence. Older evolutionary strata, if any, were expected to contain fewer assembled W-linked sequences, and lower Z/W pairwise sequence similarity due to sequence deletion or more serious accumulation of repeat sequences that hamper the genome assembly, than younger strata.

For the Z-linked SDRs of *S. japonicum*, to which two-fold more male than female reads mapped (Figure 4a), 20.5% were shared with the partial SDR of *S. mansoni* (Figure S6), whereas the remainder were homologous to the PARs of *S. mansoni* (Figure 4a,c). We defined this shared SDR between two schistosome species as the S0 region (designated "evolutionary stratum 0" = SjapS0). As expected, we were able to assemble very few W-linked fragments at S0 compared (Table S12) to other evolutionary strata (Figure 4b),except for some that were possibly unannotated repeat sequences and showed high levels of Z/W sequence identity. This result indicated that only the SjapS0 stratum suppressed recombination on the proto-sex chromosomes in the ancestor of *S. japonicum* and *S. mansoni*, followed by the independent formation of the evolutionary stratum after these species diverged from one another (Picard et al., 2018). One would expect the master female-determining gene(s), if present, to be located on W-linked sequences, or that a dose-dependent Z-linked gene occurs in this SjapS0 region (Table S13). However, we were unable to identify any genes on the very few assembled W-linked sequences at S0. A recent study used de novo assembled transcriptome sequences and identified a spicing factor *U2AF2* at W-linked S0 as a candidate female-determining gene (Elkrewi et al., 2021).

We inferred two younger evolutionary strata, SjapS1 and SjapS2, of *S. japonicum* that were mapped to the PAR of *S. mansoni*, and also exhibited a 2-fold male-biased read coverage. SjapS1 had no homologous W-linked fragments assembled and there were 945 kb of W-linked fragments assembled for SjapS2. A Z-linked inversion probably reversed the order of SjapS2 and SjapS3, the latter of which showed a similar level of read coverage between sexes, but a higher SNP density in females (i.e., the Z/W sequence divergence level, Figure 4d) than in males. The youngest stratum SjapS4 also had a similar level of read coverage between sexes, but a lower degree of female-biased SNP pattern compared to SjapS3. We hypothesized that SjapS3 and SjapS4 represented evolutionary strata that have recently lost homologous recombination. Therefore, the W chromosome regions are expected to show very

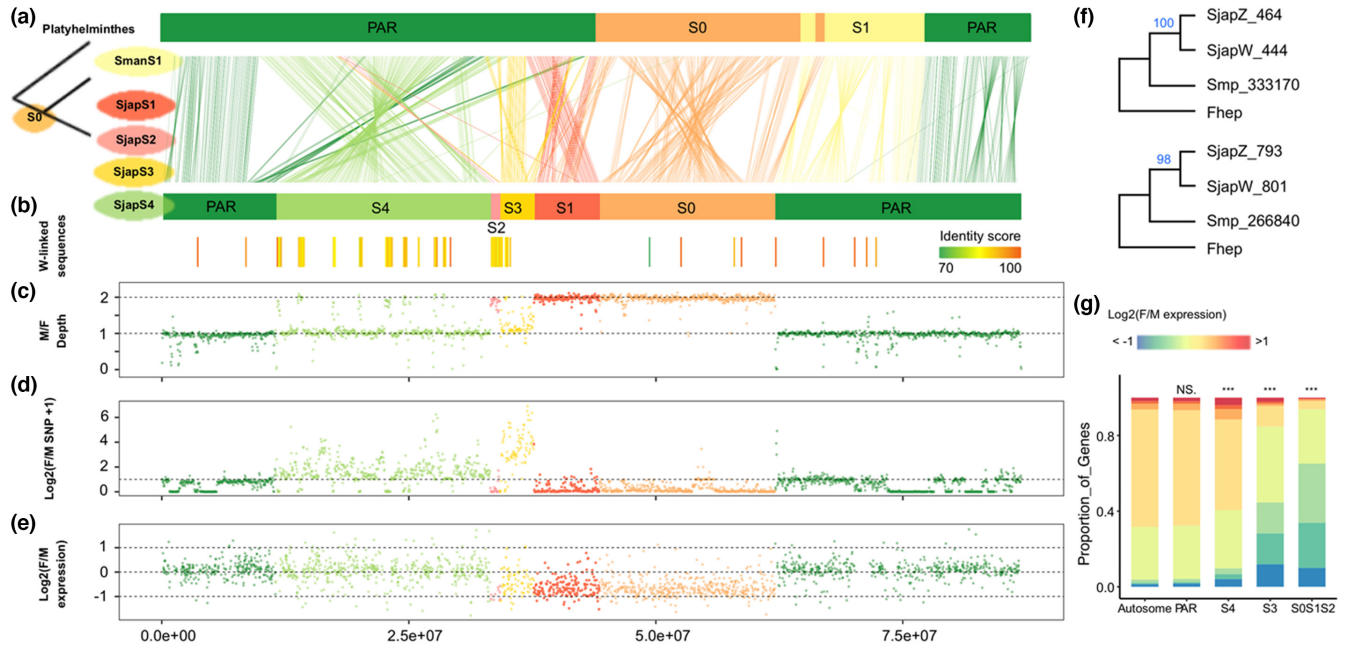

FIGURE 4 Evolutionary strata on the Z chromosome. (a) Genes orthologous between the *Schistosoma japonicum* (Sjap) Z chromosome and the *S. mansoni* (Sman; fused) ZW chromosome were plotted. Each line represents one pair of orthologues, and different line colours represent orthologues with varied genomic locations. The inferred region of each evolutionary stratum of *S. japonicum* and *S. mansoni* (Figure S6), and also PAR of the Z chromosome are shown. Since the W chromosomes are expected to share the PAR with the Z chromosome, the PAR of the W are not shown. (b) *S. japonicum* (Sjap) sequence similarity between the W and Z chromosomes were plotted; colour from green to red represents low (70%) to high (100%) similarity. (c–e) *S. japonicum* read-coverage on the Z chromosome, SNPs and gene transcription plots. Each dot represents a 50-kb window. Coordinates on the Z chromosome of *S. japonicum* are used in these figures. (f) The ML (maximum-likelihood) phylogenetic relationship of *S. japonicum* gametologues; *Fasciola hepatica* (Fhep) was used as an outgroup. (g) Sex-biased transcription of genes shown as proportions. Asterisks illustrate significant numbers of differences between the sex-biased genes (F/M > 2 or M/F > 2) on the sex and autosomal chromosomes: ***$p < .001$, estimated using the chi-squared test.

limited sequence divergence, which would explain the similar coverage level between sexes, and chimeric assembly earlier observed in the *S. japonicum* V3 chimeric assembly (Figure S3). We found very few Z- and W-linked gametologue pairs for the older strata SjapS0-2; however, we did confirm that the divergence level at the silent sites was indeed higher in SjapS3 than in SjapS4, calculated by female-specific SNPs within the Z-linked genes of these two regions (Figure S7).

To search for further evidence of an independent evolution of strata between *S. japonicum* and *S. mansoni*, we reconstructed phylogenetic trees (ML) for all seven genes with Z-linked orthologous gene sequences in *S. mansoni* and both the Z/W gametologues available for *S. japonicum*. If the strata formed after the divergence of the two species, one would expect the gametologues to group according to species; otherwise, the Z- or W-linked gametologues of different species would be expected to group together. Five of seven genes from SjapS4 and the two other genes from SjapS2 supported the independent evolution of strata (Figures 4f and S8 and Table S13).

For Z-linked genes of strata SjapS0/1/2/3, we revealed overall male-biased transcription profiles (~2-fold difference between sexes; Figure 4e,g), which may be explained by a lack of global dosage compensation (Vicoso & Bachtrog, 2011) and/or an almost complete loss of W-linked genes (Charlesworth & Charlesworth, 2000). For stratum SjapS4, we inferred a largely unbiased transcription profile (with slightly more male-biased genes compared with PARs and autosomes; Figures S9, S10c and S11). Based on these findings, we concluded that the sex chromosomes of *S. japonicum* shared one recombination suppression event (SjapS0) with those of *S. mansoni*, followed by three other recombination suppression events occurring specifically in *S. japonicum*.

3.5 | W-linked transcription profiles

We annotated 35 W-linked genes on the 4.8 Mb of W-linked sequences (Figure 5a; Table S14), 20 of which were confirmed by qPCR using (separately) female and male genomic DNA and cDNA (Figure S12 and Table S15). The dynamic expression pattern of these genes was analysed separately using the transcriptome data for each male and female worm of *S. japonicum* at eight time points (14–28 dpi) throughout the sexual developmental process (Wang et al., 2017). As expected, these W-linked genes were specifically or highly transcribed in female worms. The transcription patterns of W-linked genes in both the egg and cercarial stages were similar to those in the schistosomula and adult stages. Although we were unable to differentiate the transcription of these genes in female

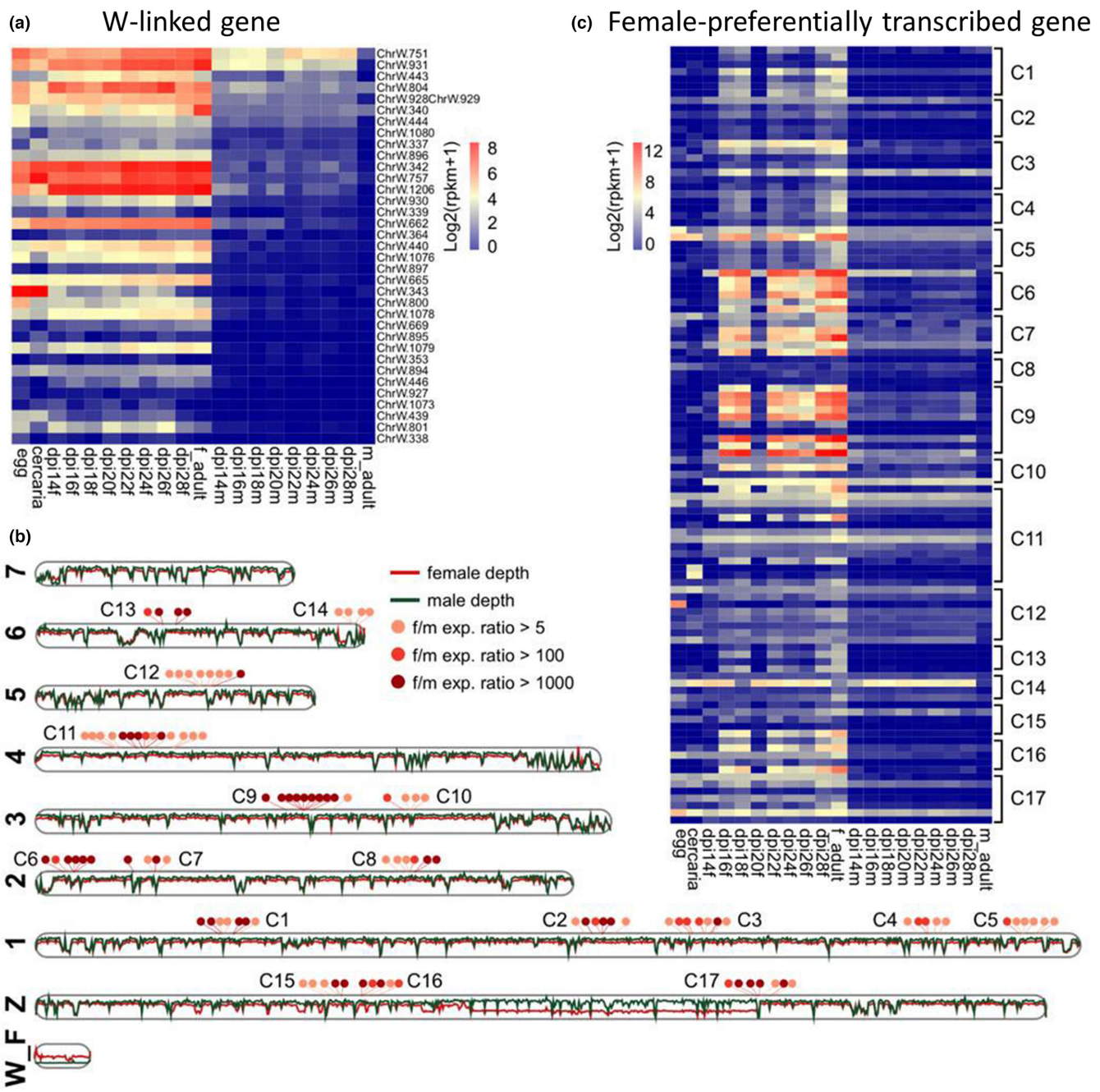

FIGURE 5 Sex-linked gene expression and female-preferentially transcribed gene clusters. (a) Transcription of W-chromosomal genes in different developmental stages. (b) Female-preferentially transcribed gene (FTG) cluster distribution (C1–C17); each solid circle represents one gene, and F/M expression differences are shown by colour variation. (c) Female-preferentially transcribed gene (FTG) cluster expression across various developmental stages. Dpi, days post-infection.

eggs and cercariae, our results indicated that they might be stably transcribed in females of all stages. Gene ontology analysis assigned the W-linked genes to metabolic and cellular processes, binding and cellular anatomy (Figure S13). Of the 35 W-linked genes, 11 were linked to metabolic enzymes (such as glycogen phosphorylase, beta-1, 3-galactosyltransferase and Ras-related protein Rab-14), three transcription factors or co-activators, two receptors (a rhodopsin orphan GPCR and a growth hormone secretagogue receptor) and one chromatin regulatory protein (Sir2).

3.6 | Female-biased gene clusters on autosomes and the Z chromosome

We did not identify any genes associated with egg-production encoded in the W-chromosome, except for a gene belonging to the vitellogenin-N superfamily. We found 257 genes on seven autosomes or the Z chromosome which were transcribed at a higher level (≥5-fold) in adult females than in adult males (Table S16); notably, 42.0% (108 genes) of them formed 17 gene clusters (Figure 5b; Table S16);

four genes on chromosome 2 (cluster 6) and six on chromosome 3 (cluster 9) were located in arrays (Figure S14). These genes are short and contain only one or two exons. All four adjacent genes of cluster 6 encode eggshell proteins (ESPs); the six adjacent genes in cluster 9 encode a chorion protein, a trematode eggshell synthesis protein, two DNA damage-responsive proteins, an asparagine-rich antigen and a gene without annotation.

Unlike W-linked genes that were almost constitutively transcribed in females in all developmental stages studied, the 257 non-W-chromosomal genes with female-biased transcription were preferentially and highly transcribed in schistosomula (from 16 dpi) and in adult worms. It is noteworthy that the transcription of 156 (60.7%) of these 257 genes was markedly down-regulated at 20 dpi (Figure 5c), when ~50% of worms were paired. This transcriptional down-regulation was unique, and was not observed for genes on chromosomes 1, or the Z and W chromosomes (Figures 5a and S15).

We used EDGER to estimate the significant gene expression changes. For 108 female-biased gene clusters on autosomes and the Z chromosome, 104 of them showed a significantly higher expression in adult females compared with adult males ($p < .05$); the other four genes were filtered out caused by low reads count (Table S17, sheet 1). For the 156 female down-regulated genes at 20 dpi, 146 showed a significantly lower expression at 20 dpi compared with 18 and 22 dpi ($p < .05$); the other 10 genes were filtered out as a result of low read counts (Table S17, sheet 2).

4 | DISCUSSION

Here, by combining PacBio, 10× Genomics, Hi-C and Illumina sequence data from a clonal line of *Schistosoma japonicum*, we produced an improved high-quality chromosome-level genome for this species. We created a new pipeline and separately assembled the Z- and W-chromosomes. Then, we reconstructed the evolutionary history of the sex chromosomes in *Schistosoma* spp., revealing a partially overlapping pattern of evolutionary strata and found considerably more regions with suppressed recombination between sex chromosomes than inferred previously (Figure S16; Picard et al., 2018). Published genomes for *S. mansoni* and *S. japonicum* are chimeric, with some mixed sequences representing both the Z and W chromosomes. The order of sequences within the two chromosomes reported here for both species is significantly improved (Picard et al., 2018), allowing us to reveal unique observations regarding the evolutionary strata of schistosome sex chromosomes.

We inferred that, besides one evolutionary stratum, or recombination suppression event, shared by *S. mansoni* and *S. japonicum*, the latter species has formed four independent evolutionary strata (Figure 4). Two of four strata (SjapS3 and SjapS4) probably formed recently, given the similarity in read coverage between the sexes, but significantly more heterozygosity in females than in males. However, as the W chromosome sequences are not completely assembled, the mechanism(s) of formation of the evolutionary strata cannot be established. A specific reduction in read coverage in females, but not in males, at the boundary between SjapS3 and SjapS2 and between SjapS4 and SjapS2, suggests that, similar to the sex chromosomes of humans and birds (Lahn & Page, 1999; Zhou et al., 2014), a chromosomal inversion on the W chromosome might account for the formation of SjapS2 (Figure S16a).

The W-chromosomal genes of *S. japonicum* were observed to be transcribed in several, and perhaps all, female developmental stages. Although our study did not include an analysis of the developmental stages in the snail intermediate host, we observed a stable transcription of these genes in eggs, cercariae, female schistosomula and female adults, providing some evidence for sex determination by the W chromosome. The continuous transcription/expression of these W chromosome-specific genes could determine the sex of the parasite after the egg is fertilized.

Recently, Elkrewi et al. (2021) applied a *k*-mer-based pipeline to assemble female-specific transcripts in *S. mansoni* and *S. japonicum*. We searched the reported W-linked transcripts of the two Schistosome species annotated by Elkrewi et al. against the *S. japonicum* genome V3 by BLAST (Table S18, sheet 1). In total, 81 of the 94 or 86.1% of the W-linked *S. japonicum* transcripts found by Elkrewi et al. can be found in our *S. japonicum* genome V3 (Table S18, sheet 1). Besides the 35 W-linked genes we annotated and analysed on the chromosomal assembly, another 80 W-linked genes can be found on the contig sequences that cannot be incorporated into the chromosomal sequences, due to the lack of 10× or Hi-C linkage evidence. For the 90 *S. mansoni* W-linked transcripts identified by Elkrewi et al., we found 12 of them have homologous sequences in *S. japonicum* genome V3 located on chrZ or autosomes (Table S18, sheet 2). For *U2AF2*, we only found its gametologue on *S. japonicum* genome V3 chrZ.

We showed that some genes encoding transcription factors, RNA-binding proteins, GPCRs and chromatin regulatory proteins are located on the W chromosome, features which merit further study to determine their precise functions in the regulation of gene transcription during female reproductive development. Of these, ChrW.662 encodes a chromatin regulatory protein, Sir2. Sir2 is a member of the class of histone deacetylases, which coordinate the DNA transcription by deacetylation of histones (Strahl & Allis, 2000). Sir2s are widely present in all phyla, and they play different roles depending on the organism (Gomes et al., 2021; Religa & Waters, 2012). In *S. mansoni*, *Sm*Sir2 was first reported by Lancelot et al. (2013). *Sm*Sir2 was considered as a novel drug target. Inhibitors of *Sm*Sir2 showed potency against both larval schistosomes (viability) and adult worms (pairing, egg-laying) in culture (Monaldi et al., 2019). A role for Sir2 in schistosome sex determination and differentiation has not been reported. However, *Drosophila* Sir2 was found to be an essential gene whose loss of function resulted in both segmentation defects and skewed sex ratios, associated with reduced activities of the Hairy and Deadpan bHLH repressors (Rosenberg & Parkhurst, 2002).

Some autosomal genes that are highly transcribed in females are clustered and stage-specific. We identified four adjacent single-exon genes encoding ESPs on autosomal chromosome 2, that is Chr2.57 (encoding ESP2a), Chr2.58 (encoding ESP1a), Chr2.59 (encoding

ESP1b) and Chr2.60 (encoding ESP2b). Henkle et al. (1990) reported that a female-specific component in nuclear extracts of *S. japonicum* bound to conserved DNA sequences (i.e., TCACGT, TCAGCT and GTAGAAT) upstream of the transcription start sites of the ESP genes, indicating that clustering of genes preferentially transcribed in females, including ESPs, might facilitate transcriptional regulation.

Direct male–female interaction in schistosomes usually occurs after 14 dpi (Wang et al., 2017), such that the high level of transcription in female worms at 16 and 18 dpi (and low transcription in eggs and cercariae) indicates a dependence on worm pairing. The marked down-regulation of most of these genes at 20 dpi we observed was unexpected. We hypothesize that mRNAs transiently accumulate in young females in large quantities and are required for egg production and/or for the eggs themselves. Notably, from 22 dpi onward, as egg production stabilizes, these genes are upregulated and regain high levels of transcription.

5 | CONCLUSION

The enhanced genome assembly (V3) for *Schistosoma japonicum* and the genetic analyses of the Z and W chromosomes presented here provide a solid platform for future functional investigations of *S. japonicum* to unravel the processes of sex determination, differentiation and pairing, which could enable the discovery and development of new interventions for the control of schistosomiasis.

AUTHOR CONTRIBUTIONS

X.X. cloned parasites and prepared genomic DNAs or RNAs for sequencing. Y.W., Q.G., G.G., G.J. and Z.Z. performed the high-throughput sequencing, genome assembly, gene annotation and other bioinformatic analysis. Y.D., Y.L., G.W., C.W. and X.H. constructed Hi-C-seq and RNA-seq libraries. X.X., Y.W., D.P.M., J.C., Q.Z. and Q.Z. wrote the manuscript with crucial contributions from R.B.G., N.D.Y., W.Z., S.W., A.P. and other co-authors. Q.Z., Q.Z., J.C. and D.P.M. managed the project. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Prof. Zhao-Rong Lun for reading and commenting on a draft of the manuscript. This work was supported by the National Key R&D Program of China 2018YFA0507300 and 2020YFC1200105; the National Natural Science Foundation of China (NSFC) 81630063, 81971959, 81971967, 31671319, 31722050, 32061130208 and 81830066; the Natural Science Foundation of Zhejiang Province LD19C190001; the European Research Council Starting Grant 677696; and the Shanghai Blue Cross Brain Hospital Co., Ltd and Shanghai Tongji University Education Development Foundation. Research at the University of Melbourne was supported by the Australian Research Council (ARC; R.B.G. and N.D.Y.). Schistosomiasis research at QIMR Berghofer (D.P.M.) is supported by the National Health and Medical Research Council of Australia.

CONFLICT OF INTEREST

No competing interests.

DATA AVAILABILITY STATEMENT

All the data have been deposited in the Genome Warehouse in the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under BioProject ID PRJCA010213 that is publicly accessible at https://ngdc.cncb.ac.cn/gwh.

ORCID

Neil D. Young https://orcid.org/0000-0001-8756-229X
Qi Zhou https://orcid.org/0000-0002-7419-2047
Qingfeng Zhang https://orcid.org/0000-0002-8759-9102

REFERENCES

Adey, A., Kitzman, J. O., Burton, J. N., Daza, R., Kumar, A., Christiansen, L., Ronaghi, M., Amini, S., Gunderson, K. L., Steemers, F. J., & Shendure, J. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Research*, *24*, 2041–2049. https://doi.org/10.1101/gr.178319.114

Alioto, T., Blanco, E., Parra, G., & Guigó, R. (2018). Using geneid to identify genes. *Current Protocols in Bioinformatics*, *64*, e56. https://doi.org/10.1002/cpbi.56

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, *20*, 224. https://doi.org/10.1186/s13059-019-1829-6

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Basch, P. F. (1990). Why do schistosomes have separate sexes. *Parasitology Today*, *6*, 160–163. https://doi.org/10.1016/0169-4758(90)90339-6

Beckmann, S., Quack, T., Burmeister, C., Buro, C., Long, T., Dissous, C., & Grevelding, C. G. (2010). *Schistosoma mansoni*: Signal transduction processes during the development of the reproductive organs. *Parasitology*, *137*, 497–520. https://doi.org/10.1017/S0031182010000053

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A., Dillon, G. P., Cerqueira, G. C., Mashiyama, S. T., Al-Lazikani, B., Andrade, L. F., Ashton, P. D., Aslett, M. A., Bartholomeu, D. C., Blandin, G., Caffrey, C. R., Coghlan, A., Coulson, R., Day, T. A., Delcher, A., DeMarco, R., … El-Sayed, N. M. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature*, *460*, U352–U365. https://doi.org/10.1038/nature08160

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and genomewise. *Genome Research*, *14*, 988–995. https://doi.org/10.1101/gr.1865504

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., … Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*, D344–D354. https://doi.org/10.1093/nar/gkaa977

Bolt, B. J., Rodgers, F. H., Shafie, M., Kersey, P. J., Berriman, M., & Howe, K. L. (2018). Using WormBase ParaSite: An integrated platform for exploring helminth genomic data. *Methods in Molecular Biology*, *1757*, 471–491. https://doi.org/10.1007/978-1-4939-7737-6_15

Buddenborg, S. K., Tracey, A., Berger, D. J., Lu, Z., Doyle, S. R., Fu, B., Yang, F., Reid, A. J., Rodgers, F. H., Rinaldi, G., Sankaranarayanan, G., Böhme, U., Holroyd, N., & Berriman, M. (2021). Assembled chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the evolution of its ZW sex-determination system. *bioRxiv*. https://doi.org/10.1101/2021.08.13.456314

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, *268*, 78–94. https://doi.org/10.1006/jmbi.1997.0951

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, *31*, 1119–1125. https://doi.org/10.1038/nbt.2727

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Cao, C. L., Zhang, L. J., Deng, W. P., Li, Y. L., Lv, C., Dai, S. M., Feng, T., Qin, Z. Q., Duan, L. P., Zhang, H. B., Hu, W., Feng, Z., Xu, J., Lv, S., Guo, J. G., Li, S. Z., Cao, J. P., & Zhou, X. N. (2020). Contributions and achievements on schistosomiasis control and elimination in China by NIPD-CTDR. *Advances in Parasitology*, *110*, 1–62. https://doi.org/10.1016/bs.apar.2020.04.002

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *355*, 1563–1572. https://doi.org/10.1098/rstb.2000.0717

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *30*, 31–37. https://doi.org/10.1093/bioinformatics/btt310

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, *10*, 563–569. https://doi.org/10.1038/nmeth.2474

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, *13*, 1050–1054. https://doi.org/10.1038/nmeth.4035

Colley, D. G., Andros, T. S., & Campbell, C. H. (2017). Schistosomiasis is more prevalent than previously thought: What does it mean for public health goals, policies, strategies, guidelines and intervention programs? *Infectious Diseases of Poverty*, *6*, e63. https://doi.org/10.1186/s40249-017-0275-5

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A niversal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Crellen, T., Walker, M., Lamberton, P. H., Kabatereine, N. B., Tukahebwa, E. M., Cotton, J. A., & Webster, J. P. (2016). Reduced efficacy of praziquantel against *Schistosoma mansoni* is associated with multiple rounds of mass drug administration. *Clinical Infectious Diseases*, *63*, 1151–1159. https://doi.org/10.1093/cid/ciw506

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*, 133–138. https://doi.org/10.1126/science.1162986

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The PFAM protein families database in 2019. *Nucleic Acids Research*, *47*, D427–D432. https://doi.org/10.1093/nar/gky995

Elkrewi, M., Moldovan, M. A., Picard, M. A. L., & Vicoso, B. (2021). Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. *Molecular Biology and Evolution*, *38*, 5345–5358. https://doi.org/10.1093/molbev/msab178

Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A. M., & Koren, S. (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computational Biology*, *15*, e1007273. https://doi.org/10.1371/journal.pcbi.1007273

Gomes, R. A., Fornari, E., Silva Rocha, A. C., Tripodi, G. L., Silva Emery, F. D., & Goulart Trossini, G. H. (2021). Parasitic sirtuin 2 as an opportunity in drug discovery. *Future Medicinal Chemistry*, *13*, 1397–1409. https://doi.org/10.4155/fmc-2021-0091

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Jr., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*, 5654–5666. https://doi.org/10.1093/nar/gkg770

Handley, L.-J. L., Ceplitis, H., & Ellegren, H. (2004). Evolutionary strata on the chicken Z chromosome: Implications for sex chromosome evolution. *Genetics*, *167*, 367–376. https://doi.org/10.1534/genetics.167.1.367

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., … White, R. (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*, D258–D261. https://doi.org/10.1093/nar/gkh036

Henkle, K. J., Cook, G. A., Foster, L. A., Engman, D. M., Bobek, L. A., Cain, G. D., & Donelson, J. E. (1990). The gene family encoding eggshell proteins of *Schistosoma japonicum*. *Molecular and Biochemical Parasitology*, *42*, 69–82. https://doi.org/10.1016/0166-6851(90)90114-2

Hoff, K. J., & Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*, e57. https://doi.org/10.1002/cpbi.57

International Helminth Genomes Consortium. (2019). Comparative genomics of the major parasitic worms. *Nature Genetics*, *51*, 163–174. https://doi.org/10.1038/s41588-018-0262-1

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*, W5–W9. https://doi.org/10.1093/nar/gkn201

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*, 357–360. https://doi.org/10.1038/nmeth.3317

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096

Kunz, W. (2001). Schistosome male-female interaction: Induction of germ-cell differentiation. *Trends in Parasitology*, *17*, 227–231. https://doi.org/10.1016/s1471-4922(01)01893-1

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*, 1–9. https://doi.org/10.1186/gb-2004-5-2-r12

Lahn, B. T., & Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science*, *286*, 964–967. https://doi.org/10.1126/science.286.5441.964

Lancelot, J., Caby, S., Dubois-Abdesselem, F., Vanderstraete, M., Trolet, J., Oliveira, G., Bracher, F., Jung, M., & Pierce, R. J. (2013). *Schistosoma mansoni* Sirtuins characterization and potential as chemotherapeutic targets. *PLoS Neglected Tropical Diseases*, *7*, e2428. https://doi.org/10.1371/journal.pntd.0002428

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with BOWTIE 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Zhang, J., Liu, J., Zhou, Y., Cai, C., Xu, L., Dai, X., Feng, S., Guo, C., & Rao, J. (2021). A new duck genome reveals conserved and convergently evolved chromosome architectures of birds and mammals. *Gigascience*, *10*, giaa142. https://doi.org/10.1093/gigascience/giaa142

Loker, E. S., & Brant, S. V. (2006). Diversification, dioecy and dimorphism in schistosomes. *Trends in Parasitology*, *22*, 521–528. https://doi.org/10.1016/j.pt.2006.09.001

Luo, F., Yin, M., Mo, X., Sun, C., Wu, Q., Zhu, B., Xiang, M., Wang, J., Wang, Y., Li, J., Zhang, T., Xu, B., Zheng, H., Feng, Z., & Hu, W. (2019). An improved genome assembly of the fluke *Schistosoma japonicum*. *PLoS Neglected Tropical Diseases*, *13*, e0007612. https://doi.org/10.1371/journal.pntd.0007612

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics*, *20*, 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

McManus, D. P., Dunne, D. W., Sacko, M., Utzinger, J., Vennervald, B. J., & Zhou, X. N. (2018). Schistosomiasis. *Nature Reviews Disease Primers*, *4*, 13. https://doi.org/10.1038/s41572-018-0013-8

Monaldi, D., Rotili, D., Lancelot, J., Marek, M., Wossner, N., Lucidi, A., Tomaselli, D., Ramos-Morales, E., Romier, C., Pierce, R. J., Mai, A., & Jung, M. (2019). Structure-reactivity relationships on substrates and inhibitors of the lysine deacylase sirtuin 2 from *Schistosoma mansoni* (SmSirt2). *Journal of Medicinal Chemistry*, *62*, 8733–8759. https://doi.org/10.1021/acs.jmedchem.9b00638

Mone, H., & Boissier, J. (2004). Sexual biology of schistosomes. *Advances in Parasitology*, *57*, 89–189. https://doi.org/10.1016/S0065-308X(04)57002-1

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*, 621–628. https://doi.org/10.1038/nmeth.1226

Nelson, C. W., Moncla, L. H., & Hughes, A. L. (2015). SNPGenie: Estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics*, *31*, 3709–3711. https://doi.org/10.1093/bioinformatics/btv449

Noel, H., Ruello, M., Maccary, A., Pelat, C., Sommen, C., Boissier, J., Barre-Cardi, H., Fillaux, J., Termignon, J. L., Debruyne, M., Chyderiotis, G., Durand, J., Ramalli, L., Chiappini, J. D., Malfait, P., & Berry, A. (2018). Large outbreak of urogenital schistosomiasis acquired in Southern Corsica, France: Monitoring early signs of endemicization? *Clinical Microbiology and Infection*, *24*, 295–300. https://doi.org/10.1016/j.cmi.2017.06.026

Oey, H., Zakrzewski, M., Gravermann, K., Young, N. D., Korhonen, P. K., Gobert, G. N., Nawaratna, S., Hasan, S., Martínez, D. M., You, H., Lavin, M., Jones, M. K., Ragan, M. A., Stoye, J., Oleaga, A., Emery, A. M., Webster, B. L., Rollinson, D., Gasser, R. B., … Krause, L. (2019). Whole-genome sequence of the bovine blood fluke Schistosoma bovis supports interspecific hybridization with *S. haematobium*. *PLoS Pathogens*, *15*, e1007513. https://doi.org/10.1371/journal.ppat.1007513

Otto, S. P., Pannell, J. R., Peichel, C. L., Ashman, T.-L., Charlesworth, D., Chippindale, A. K., Delph, L. F., Guerrero, R. F., Scarpino, S. V., & McAllister, B. F. (2011). About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics*, *27*, 358–367.

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genornes. *Bioinformatics*, *23*, 1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Picard, M. A. L., Cosseau, C., Ferre, S., Quack, T., Grevelding, C. G., Coute, Y., & Vicoso, B. (2018). Evolution of gene dosage on the Z-chromosome of schistosome parasites. *eLife*, *7*, e35684. https://doi.org/10.7554/eLife.35684

Pignatelli, M., Serras, F., Moya, A., Guigó, R., & Corominas, M. (2009). CROC: Finding chromosomal clusters in eukaryotic genomes. *Bioinformatics*, *25*, 1552–1553. https://doi.org/10.1093/bioinformatics/btp248

Popiel, I., Cioli, D., & Erasmus, D. A. (1984). The morphology and reproductive status of female *Schistosoma mansoni* following separation from male worms. *International Journal for Parasitology*, *14*, 183–190. https://doi.org/10.1016/0020-7519(84)90047-x

Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., De Silva, N., Velarde, G. S., Anderson, T. J., Clark, R. C., Davidson, C., Dillon, G. P., Holroyd, N. E., LoVerde, P. T., Lloyd, C., McQuillan, J., Oliveira, G., Otto, T. D., Parker-Manuel, S. J., … Berriman, M. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases*, *6*, e1455. https://doi.org/10.1371/journal.pntd.0001455

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Religa, A. A., & Waters, A. P. (2012). Sirtuins of parasitic protozoa: In search of function(s). *Molecular and Biochemical Parasitology*, *185*, 71–88. https://doi.org/10.1016/j.molbiopara.2012.08.003

Rosenberg, M. I., & Parkhurst, S. M. (2002). Drosophila Sir2 is required for heterochromatic silencing and by euchromatic Hairy/E(Spl) bHLH repressors in segmentation and sex determination. *Cell*, *109*, 447–458. https://doi.org/10.1016/S0092-8674(02)00732-8

Ross, A. G. P., Bartley, P. B., Sleigh, A. C., Olds, G. R., Li, Y. S., Williams, G. M., & McManus, D. P. (2002). Current concepts—Schistosomiasis. *New England Journal of Medicine*, *346*, 1212–1220. https://doi.org/10.1056/Nejmra012396

Rudge, J. W., Webster, J. P., Lu, D. B., Wang, T. P., Fang, G. R., & Basanez, M. G. (2013). Identifying host species driving transmission of schistosomiasis japonica, a multihost parasite system, in China. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 11457–11462. https://doi.org/10.1073/pnas.1221509110

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Comeau, D. C., Funk, K., Kim, S., Klimke, W., Marchler-Bauer, A., Landrum, M., Lathrop, S., Lu, Z., Madden, T. L., O'Leary, N., Phan, L., Rangwala, S. H., Schneider, V. A., … Sherry, S. T. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *49*, D10–D17. https://doi.org/10.1093/nar/gkaa892

Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium. (2009). The *Schistosoma japonicum* genome reveals

features of host-parasite interplay. *Nature*, *460*, 345–351. https://doi.org/10.1038/nature08140

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*, 259. https://doi.org/10.1186/s13059-015-0831-x

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*, 41–45. https://doi.org/10.1038/47412

Stroehlein, A. J., Korhonen, P. K., Chong, T. M., Lim, Y. L., Chan, K. G., Webster, B., Rollinson, D., Brindley, P. J., Gasser, R. B., & Young, N. D. (2019). High-quality *Schistosoma haematobium* genome achieved by single-molecule and long-range sequencing. *Gigascience*, *8*, giz108. https://doi.org/10.1093/gigascience/giz108

Theron, A., Mouahid, G., & Mone, H. (1997). *Schistosoma mansoni*: Cercarial shedding patterns from a mixed infection of *Biomphalaria glabrata* with two (early and late) chronobiological variants. *Parasitology Research*, *83*, 356–358. https://doi.org/10.1007/s004360050261

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*, 562–578. https://doi.org/10.1038/nprot.2012.016

UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*, D480–D489. https://doi.org/10.1093/nar/gkaa1100

Vicoso, B., & Bachtrog, D. (2011). Lack of global dosage compensation in *Schistosoma mansoni*, a female-heterogametic parasite. *Genome Biology and Evolution*, *3*, 230–235. https://doi.org/10.1093/gbe/evr010

Vicoso, B., Kaiser, V. B., & Bachtrog, D. (2013). Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6453–6458. https://doi.org/10.1073/pnas.1217027110

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*, e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, J. P., Yu, Y., Shen, H. M., Qing, T., Zheng, Y. T., Li, Q., Mo, X. J., Wang, S. Q., Li, N. N., Chai, R. Y., Xu, B., Liu, M., Brindley, P. J., McManus, D. P., Feng, Z., Shi, L. M., & Hu, W. (2017). Dynamic transcriptomes identify biogenic amines and insect-like hormonal regulation for mediating reproduction in *Schistosoma japonicum*. *Nature Communications*, *8*, 14693. https://doi.org/10.1038/Ncomms14693

Webster, J. P., Molyneux, D. H., Hotez, P. J., & Fenwick, A. (2014). The contribution of mass drug administration to global health: Past, present and future. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *369*, 20130434. https://doi.org/10.1098/rstb.2013.0434

World Health Organization. (2010). *Working to overcome the global impact of neglected tropical diseases: First WHO report on neglected tropical diseases*. Department of Reproductive Health and Research, World Health Organization.

Xu, Y. Z., & Dresden, M. H. (1990). The hatching of schistosome eggs. *Experimental Parasitology*, *70*, 236–240. https://doi.org/10.1016/0014-4894(90)90104-k

Young, N. D., Jex, A. R., Li, B., Liu, S., Yang, L., Xiong, Z., Li, Y., Cantacessi, C., Hall, R. S., Xu, X., Chen, F., Wu, X., Zerlotini, A., Oliveira, G., Hofmann, A., Zhang, G., Fang, X., Kang, Y., Campbell, B. E., … Gasser, R. B. (2012). Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics*, *44*, 221–225. https://doi.org/10.1038/ng.1065

Young, N. D., Nagarajan, N., Lin, S. L. J., Korhonen, P. K., Jex, A. R., Hall, R. S., Safavi-Hemami, H., Kaewkong, W., Bertrand, D., Gao, S., Seet, Q. H., Wongkham, S., Teh, B. T., Wongkham, C., Intapan, P. M., Maleewong, W. C., Yang, X. H., Hu, M., Wang, Z., … Gasser, R. B. (2014). The Opisthorchis viverrini genome provides insights into life in the bile duct. *Nature Communications*, *5*, 4378. https://doi.org/10.1038/ncomms5378

Young, N. D., Stroehlein, A. J., Kinkar, L., Wang, T., Sohn, W. M., Chang, B. C. H., Kaur, P., Weisz, D., Dudchenko, O., Aiden, E. L., Korhonen, P. K., & Gasser, R. B. (2021). High-quality reference genome for *Clonorchis sinensis*. *Genomics*, *113*, 1605–1615. https://doi.org/10.1016/j.ygeno.2021.03.001

Zhao, G. H., Li, J., Lin, R. Q., Zou, F. C., Liu, W., Yuan, Z. G., Mo, X. H., Song, H. Q., Weng, Y. B., & Zhu, X. Q. (2010). An effective sequence characterized amplified region-PCR method derived from restriction site-amplified polymorphism for the identification of female *Schistosoma japonicum* of zoonotic significance. *Electrophoresis*, *31*, 641–647. https://doi.org/10.1002/elps.200900615

Zheng, G. X., Lau, B. T., Schnall-Levin, M., Jarosz, M., Bell, J. M., Hindson, C. M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D. A., Merrill, L., Terry, J. M., Mudivarti, P. A., Wyatt, P. W., Bharadwaj, R., Makarewicz, A. J., Li, Y., Belgrader, P., Price, A. D., Lowe, A. J., Marks, P., … Ji, H. P. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, *34*, 303–311. https://doi.org/10.1038/nbt.3432

Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E. D., Gilbert, M. T., & Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, *346*, 1246338. https://doi.org/10.1126/science.1246338

Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J. C., Peel, E., Cheng, Y., Pajpach, F., Bradley, N., Suzuki, H., Nikaido, M., Damas, J., Daish, T., Perry, T., Zhu, Z., Geng, Y., Rhie, A., … Zhang, G. (2021). Platypus and echidna genomes reveal mammalian biology and evolution. *Nature*, *592*, 756–762. https://doi.org/10.1038/s41586-020-03039-0

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Xu, X., Wang, Y., Wang, C., Guo, G., Yu, X., Dai, Y., Liu, Y., Wei, G., He, X., Jin, G., Zhang, Z., Guan, Q., Pain, A., Wang, S., Zhang, W., Young, N. D., Gasser, R. B., McManus, D. P., Cao, J. … Zhang, Q. (2022). Chromosome-level genome assembly defines female-biased genes associated with sex determination and differentiation in the human blood fluke *Schistosoma japonicum*. *Molecular Ecology Resources*, *00*, 1–17. https://doi.org/10.1111/1755-0998.13689