Genomics 113 (2021) 1605–1615

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

High-quality reference genome for *Clonorchis sinensis*

Neil D. Young [a,*], Andreas J. Stroehlein [a], Liina Kinkar [a], Tao Wang [a], Woon-Mok Sohn [b], Bill C. H. Chang [a], Parwinder Kaur [c], David Weisz [d], Olga Dudchenko [d,e], Erez Lieberman Aiden [c,d,e,f], Pasi K. Korhonen [a], Robin B. Gasser [a]

[a] Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia
[b] Department of Parasitology and Institute of Health Sciences, School of Medicine, Gyeongsang National University, Jinju, Republic of Korea
[c] UWA School of Agriculture and Environment, Faculty of Science, University of Western Australia, Perth, Western Australia 6009, Australia
[d] The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[e] Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA
[f] Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech, Pudong 201210, China

ARTICLE INFO

Keywords:
Chinese liver fluke
Clonorchis sinensis
Third-generation sequencing
Advanced bioinformatics
Reference genome

ABSTRACT

The Chinese liver fluke, *Clonorchis sinensis*, causes the disease clonorchiasis, affecting ~35 million people in regions of China, Vietnam, Korea and the Russian Far East. Chronic clonorchiasis causes cholangitis and can induce a malignant cancer, called cholangiocarcinoma, in the biliary system. Control in endemic regions is challenging, and often relies largely on chemotherapy with one anthelmintic, called praziquantel. Routine treatment carries a significant risk of inducing resistance to this anthelmintic in the fluke, such that the discovery of new interventions is considered important. It is hoped that the use of molecular technologies will assist this endeavour by enabling the identification of drug or vaccine targets involved in crucial biological processes and/or pathways in the parasite. Although draft genomes of *C. sinensis* have been published, their assemblies are fragmented. In the present study, we tackle this genome fragmentation issue by utilising, in an integrated way, advanced (second- and third-generation) DNA sequencing and informatic approaches to build a high-quality reference genome for *C. sinensis*, with chromosome-level contiguity and curated gene models. This substantially-enhanced genome provides a resource that could accelerate fundamental and applied molecular investigations of *C. sinensis*, clonorchiasis and/or cholangiocarcinoma, and assist in the discovery of new interventions against what is a highly significant, but neglected disease-complex.

1. Introduction

Parasitic trematodes (flukes), including species of *Schistosoma*, *Opisthorchis* and *Clonorchis*, infect hundreds of millions of people worldwide [1]. In particular, the Chinese liver fluke, *C. sinensis*, causes the disease clonorchiasis, which affects ~35 million people in parts of China, Vietnam, Korea and the Russian Far East [2–5]. People become infected with this fluke by ingesting raw or undercooked fish containing metacercariae (arrested larval stage); juvenile flukes emerge from individual metacercariae in the lumen of the small intestine and then migrate via the common bile duct into the hepatobiliary system or pancreas where they develop to adult worms [4]. Chronic infection and re-infection over years leads to cholangitis and can induce cholangiocarcinoma – a fatal cancer of the biliary system [6,7]. Despite this link between clonorchiasis and cancer, the cultural practice of eating raw fish impedes efforts to control this disease-complex in endemic regions [8]. Current control relies largely on targeted or mass treatment of people for *C. sinensis* infection with one drug – praziquantel [9]. However, the use of this drug alone has not been effective at controlling clonorchiasis, and mass treatment has a significant risk of inducing resistance in the fluke to praziquantel [9], such that the discovery and development of new interventions (drugs or vaccines) remain a priority [10].

Genomic explorations of *C. sinensis* can strongly underpin this focus by identifying essential processes, pathways or molecules in this parasite and/or involved in the disease pathogenesis, which could represent possible intervention targets [11,12]. Draft genomes of *C. sinensis* from Korea (Cs-k2) and China (Cs-c1 and Cs-c2) have allowed the inference of

* Corresponding author.
E-mail address: nyoung@unimelb.edu.au (N.D. Young).

https://doi.org/10.1016/j.ygeno.2021.03.001
Received 10 September 2020; Received in revised form 18 January 2021; Accepted 1 March 2021
Available online 4 March 2021

genes, encoded proteins as well as gene ontology networks and metabolic pathways [13–15], enabling transcriptomic investigations of *C. sinensis* [13–17] and other, comparative studies of flukes [11,18,19]. The sequencing and assembly of new genomic and transcriptomic resources have been accompanied by the development of new informatic tools and workflows to accelerate molecular research of this and related carcinogenic flukes [15,20–22].

Despite these significant advances, a major limitation for profound genomic and associated molecular and biochemical investigations of *C. sinensis* has been the fragmented nature of available draft genomes [13–15] assembled employing data produced using first- and/or second-generation sequencing methods [reviewed by [23]]. Long-read and long-range, so-called third-generation sequencing technologies are now enabling substantially enhanced genome assemblies [24–28], achieving improved scaffold contiguity and gene models, and allowing the discovery of novel, structurally-complex and repetitive, non-coding DNA regions, not adequately represented in previous draft genomes [26,27,29,30]. These advances 'open the door' to tackling the problem of fragmentation in the draft genome of *C. sinensis*. In the present study, we employed a combination of Oxford Nanopore long-read (third-generation) and in situ Hi-C [31] (second-generation) sequencing technologies to produce data that, when combined with short-read data sets, achieved a chromosome-level genomic assembly for *C. sinensis*.

2. Materials and methods

2.1. Parasite material

In a previous study [32], adult worms of *C. sinensis* were obtained from Syrian golden hamsters (*Mesocricetus auratus*), which had been experimentally infected with metacercariae isolated from naturally infected cyprinid fish (*Pseudorasbora parva*) in the Jinju-si, South Gyeongsang Province, South Korea; one of the authors (W.-M. S.) conducted this study [32], which had been approved (permit no. GNU-2009-1001) by the Animal Care and Use Committee of Gyeongsang National University, School of Medicine, Jinju, Korea. The worms collected had been washed extensively in physiological saline and then stored at −80 °C until use; some of them were used previously for transcriptomic and genomic studies of *C. sinensis* [15,17].

2.2. Isolation of genomic DNA, and construction and sequencing of Oxford Nanopore and in situ Hi-C libraries

High quality genomic DNA was isolated from 100 adults of *C. sinensis* using the Circulomics Tissue Kit (Circulomics, Baltimore, MD, USA). The integrity of the DNA was verified using an Agilent 4200 TapeStation system (ThermoFisher) and using Genomic DNA ScreenTape (ThermoFisher). Low molecular weight DNA was removed using a 5 or 20 kb short-read eliminator (SRE) kit (Circulomics, Baltimore, MD, USA), and the remaining, high molecular weight DNA was used to construct the following genomic DNA libraries: Nanopore Rapid Sequencing (SQK-RAD004; Oxford Nanopore Technologies; 5 kb SRE) and Ligation Sequencing (SQK-LSK109; Oxford Nanopore Technologies; 5 and 20 kb SRE), according to the manufacturer's protocols. The SQK-RAD004 (5 kb SRE) and SQK-LSK109 (5 kb SRE) libraries were each sequenced in separate flow cells (v.9.4.1; Oxford Nanopore Technologies). The flow cell used to sequence the SQK-LSK109 (5 kb SRE) library was washed using a Flow Cell Wash Kit (EXP-WSH003; Oxford Nanopore Technologies) and re-used to sequence the SQK-LSK109 (20 kb SRE) library. All libraries were sequenced using a MinION sequencer (Oxford Nanopore Technologies). Following sequencing, bases were 'called' from raw FAST5 reads using the program Guppy v.3.1.5 (Oxford Nanopore Technologies) and stored in the FASTQ format [33].

In situ Hi-C sequencing was performed as described previously [31]. The high molecular weight DNA from 100 *C. sinensis* adults was restriction-digested with equal concentrations of *Cvi*AII/*Mse*I (New England Biolabs); the library was constructed and then sequenced using the NextSeq500 platform (Illumina). The sequence data are available via the DNA Zoo repository (Sequence Read Archive accession number: SRX8933307) in the National Center for Biotechnology Information (NCBI) database.

2.3. Scaffolding of C. sinensis genomic contigs, and removal of contaminants

The new *C. sinensis* genome (designated Cs-k2.v2) was assembled using the following methodology: Sequences from the published Korean genome (Cs-k2.v1; NCBI BioProject PRJNA386618) [15] were combined with the in situ Hi-C data using Juicer v.1.6 [34], 3D-DNA v.180922 [35] and Juicebox Assembly Tools v.1.9.8 (http:github.com/aidenlab/Juicebox/wiki/Juicebox-Assembly-Tools/) to achieve chromosome-length scaffolds. Unanchored sequences were examined and scaffolded using available paired-end and mate-pair libraries (NCBI BioProject PRJNA386618) [15] and Oxford Nanopore long-read libraries (Supplementary Table S1) using platanus_allee v.2.0.2 [36], requiring a minimum of 15 links for scaffold-joining. Error-corrected long reads were then used to close gaps in chromosome-length and unanchored scaffolds using TGS-GapCloser v.1.1.1 (https://github.com/BGI-Qingdao/TGS-GapCloser). Redundancy was removed by mapping long-read data to the genomic scaffolds, and haplotigs were eliminated employing purge_haplotigs v.1.1.0 [37] using depths of 5, 10 and 95 reads (low, medium and high, respectively).

To remove contaminants, genomic scaffolds were fragmented into 1000 bp, non-overlapping regions, and subjected to homology searching and taxonomic classification using centrifuge v.1.04beta [38] utilising an indexed version of the NCBI non-redundant nucleotide (nt) database (https://ccb.jhu.edu/software/centrifuge/; 3 March 2018). Contigs containing non-overlapping regions that aligned over <10% to nucleotide sequences from species within the phylum Platyhelminthes, and > 10% to sequences derived from members of the Chordata (representing the hamster host of *C. sinensis*) or any other taxa were subjected to BLASTn v. 2.5.0 [39] homology searches to sequences in the NCBI nt database (accessed March 2020) [40]. Sequences were removed if the matches to nucleotide sequences in the NCBI nt database did not represent platyhelminths, or if they were contigs of ≤10,000 nucleotides in length and did not contain aligned protein-coding genes from Cs-k2.v1 [15] using gmap v.2019-09-12 [41].

2.4. Gene model prediction and functional annotation

Gene models predicted from Cs-k2.v1 and submitted to NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/003/604/175/GCA_003604175.1_ASM360417v1/GCA_003604175.1_ASM360417v1_genomic.gff.gz) [15] were transferred to the new scaffolds of genome Cs-k2.v2 using liftOver [42]. First, a chain file was created using the genomes Cs-k2.v1 and Cs-k2.v2 using the doSameSpeciesLiftOver.pl script. Then,

gene models were transformed from the genome feature format (GFF) to the gene prediction (GP) format, and transferred to the Cs-k2.v2 genome using the liftOver chain file. Transferred gene models were updated using funannotate v.1.7.4 (https://github.com/nextgenusfs/funannotate/) and using available RNA-seq data (NCBI BioProject PRJNA386618) [15] and inferred protein sequence data sets for other flatworm species available via ParaSite v.WBPS14 in WormBase [43]. The evidence modeler (EVM) v.1.1.1 [44] matrix was weighted as follows: Cs-k2.v1 liftOver GFF file: 10, PASA v.2.4.1 [44]: 5, augustus v.3.3.3 [45]: 2, location of aligned proteins of flatworms other than *C. sinensis*: 1 and StringTie v2.1.2 [46]: 1. New gene models were refined using funannotate, and 5′- and 3′-untranslated regions inferred by PASA were included. The final Cs-k2.v2 GFF file was compared with the Cs-k2.v1 liftOver and gmap GFF files, to assign gene models that were retained (beginning with CSKR_1 and retaining the Cs-k2.v1 gene number), that were new, or that resulted from the combination or fragmentation of Cs-k2.v1 gene models (beginning with CSKR_2). The completeness of the gene set was assessed (protein-mode) using the tool BUSCO v 4.0.2 [47]. The annotation of each inferred amino acid sequence was achieved using InterPro v5.35 [48], EggNOG mapper v.5.0 [49] and/or homology to proteins in the databases Swiss-Prot within UniProtKB (accessed March 2020) [50] and/or MEROPS release 12 [51] using DIAMOND (v. 0.9.21) BLASTp (*E*-value $\leq 10^{-5}$) [52]. Evidence of gene transcription was inferred by mapping available RNA-seq data (NCBI BioProject PRJNA386618) [15] to the genome using HISAT2 (v. 2.1.0) [53]. Read counts from each library were inferred using FeatureCounts (v.1.6.4) [54] and normalised to 'counts per million' (CPM) using the edgeR package (v.3.26.8) [55]. Gene models were inferred to have transcriptional support if one or more library had a CPM value of >0.5. Evidence of gene ontology (GO) enrichment in transferred (CSKR_1) or revised/new (CSKR_2) gene models was assessed using topGO (v.2.36.0; bioconductor.org/packages/release/bioc/html/topGO.html).

2.5. Assessing genome completeness and contiguity

The completeness of the Cs-k2.v2 genome was assessed (in genome-mode) using BUSCO v 4.0.2 [47]. Genome-wide synteny between the repeat-masked Cs-k2.v2 and previous, published repeat-masked scaffolds for Cs-k2.v1 and Cs-c2 [13,15] were assessed using CACTUS v.2019.03.01–1 [56] and QUAST v.5.0.2 [57]. For both comparisons, synteny was inferred using the hierarchical alignment (HAL) output file and the halSyteny program in the CACTUS package, employing the following settings: –maxAnchorDistance, 1,000,000 –minBlockSize 10,000. Links between scaffolds were bundled using bundlelinks in circos tools v.0.23 [58] using the following settings: min_bundle_size = $2e^5$, min_bundle_membership = 1 and max_gap = $1e^6$. Scaffolds were reordered and visualised using circos v.0.69-8 [58]. Protein-encoding single copy orthologs (SCOs) of *C. sinensis* (Cs-k2.v2), *Schistosoma mansoni* [59] and *Echinococcus multilocularis* [60] from WormBase Parasite vWBPS14 [43] were inferred using OrthoFinder [61]. The genomic locations of SCOs in Cs-k2.v2, *S. mansoni* or *E. multilocularis* were bundled separately using bundlelinks in circos tools using the following settings: min_bundle_size = $1e^4$, min_bundle_membership = 5 and max_gap = $1e^6$. Scaffolds were reordered and visualised using circos v.0.69-8 [58].

3. Results

3.1. Genome assembly

The long-read and in situ Hi-C sequence data (24.8 Gb, 44-fold coverage; Supplementary Table S1) were combined with available paired-end and mate-pair read data for Cs-k2.v1 [15] to produce a genome assembly of 558 Mb (designated Cs-k2.v2; scaffold N50/N90 = 168.7/30 Mb; scaffold L50/L90 = 2/6; Table 1). This assembly contained 78 contiguous sequences, compared with 2776 and 4348 in previous assemblies of Cs-k2.v1 and Cs-c2, respectively [13,15]. The BUSCO metrics for genome completeness were consistent with Cs-k2.v1 (Table 1). Most aligned, unique in situ Hi-C reads inferred 211,627 'inter-chromosomal' and 1,217,745 'intra-chromosomal' contacts (Fig. 1A; Supplementary Table S2).

Table 1
Features of the *Clonorchis sinensis* reference genome (Cs-k2.v2) compared with those of previous assemblies (C-k2.v1 and Cs-c2) [13,15].

Assembly	*C. sinensis* Korea	*C. sinensis* Korea	*C. sinensis* China
	Cs-k2.v2	C-k2.v1	Cs-c2
Number of scaffolds	78	2776	4348
Total size of scaffolds	558,124,894	562,768,885	547,288,241
Longest scaffold	222,349,635; 10,571	8,861,937; 501	20,508,42; 300
Number of scaffolds >100 K, 1 M; 10 M	22; 7; 7	542; 204; 0	1411; 44; 0
N50 scaffold length; L50 scaffold count	168,711,085; 2	1,628,761; 103	417,486; 408
Scaffold %GC	43.84	42.56	44.03
Scaffold %N	0.52	3.16	0.03
Number of contigs	1604	15,580	6190
Longest contig	3,786,446	793,193	1,089,335
Number of contigs >100 K; 1 M	923; 150	1465; 0	1815; 1
N50 contig length; L50 contig count	844,777; 202	89,578; 1746	233,037; 719
Contig %GC	44.07	43.95	44.05
Genome completeness and accuracy:			
Complete BUSCOs[a]	618 (64.8%)	618 (64.8%)	597 (62.6%)
Complete single-copy BUSCOs	612 (64.2%)	611 (64.1%)	588 (61.6%)
Complete and duplicated BUSCOs	6 (0.6%)	7 (0.7%)	9 (0.9%)
Fragmented BUSCOs	62 (6.5%)	66 (6.9%)	83 (8.7%)
Genome-wide differences[b]		241	3187
Contigs containing differences[b]		147	1328
Total length of contigs containing differences[b]		237,316,409	432,813,610

[a] Number of Benchmarking Universal Single-Copy Orthologs (BUSCOs) identified (genome-mode) and percentage of the 954 genes within the Metazoa_odb10 dataset.
[b] Metrics predicted using the program QUAST [57]. Genome-wide differences based on comparison with the reference genome (Cs-k2.v2).

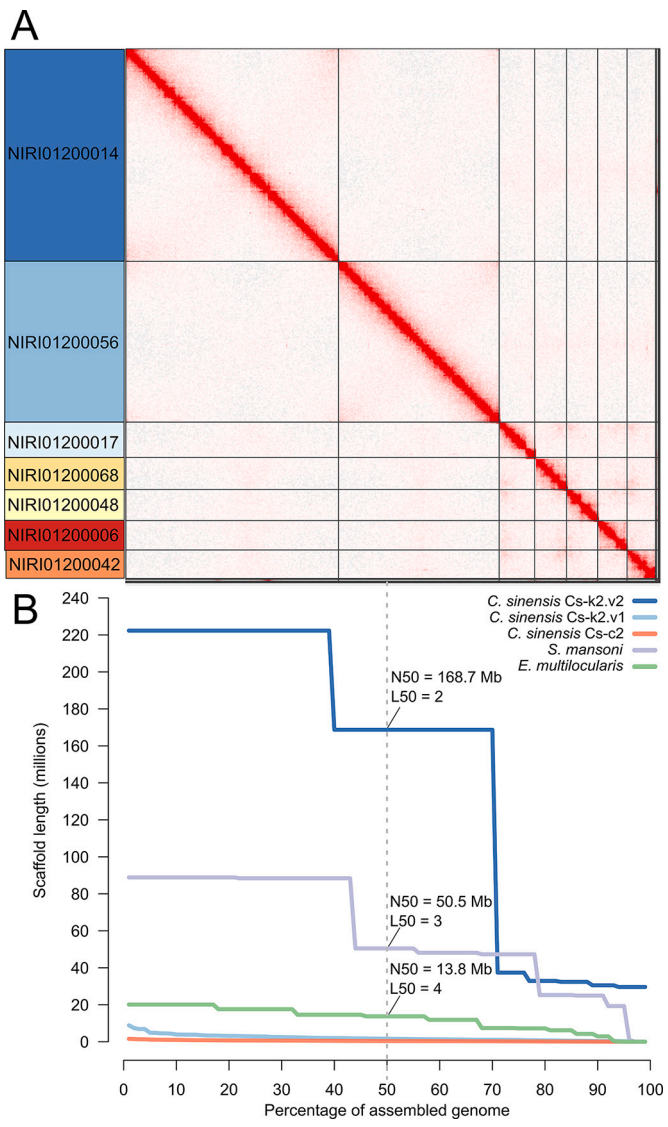

Fig. 1. Completeness of the reference genome (Cs-k2.v2) of *Clonorchis sinensis*. Panel A: Hi-C matrix of the spatial clustering of Hi-C reads to the seven inferred chromosomes (NIRI012-codes; left) in Cs-k2.v2. An interactive Hi-C contact map [94] is available at www.dnazoo.org/assemblies/Clonorchis_sinensis. Panel B: Comparison of the scaffold lengths of the Cs-k2.v2 genome with those of draft genomes of *C. sinensis* from Korea (Cs-k2.v1), *C. sinensis* from Chinese (Cs-c2) and reference genomes available for the blood fluke *Schistosoma mansoni* and for the taeniid tapeworm *Echinococcus multilocularis*. The N50 and L50 values for the reference genomes are indicated.

3.2. Inference of seven chromosomes

The final assembly comprised two large (designated NIRI01200014 and NIRI0120005) and five small (NIR01200068, NIR01200048, NIR01200042, NIR01200017 and NIR01200006) chromosome-length scaffolds containing 99.2% of all assembled scaffolds, with abundant, off-diagonal peaks being indicative of telomeres (Fig. 1A). The two large chromosome-length scaffolds contained >50% of the entire 558 Mb genome (L50 = 2), as distinct from chromosomes inferred for the genomes of *S. mansoni* (N50 = 50.5 Mb; L50 = 3) and *E. multilocularis* (N50 = 13.8 Mb; L50 = 4) (Fig. 1B).

Table 2
Features of the *Clonorchis sinensis* reference gene set (Cs-k2.v2) compared with those previously reported (C-k2.v1 and Cs-c2) [13,15].

Features	*C. sinensis* Korea	*C. sinensis* Korea	*C. sinensis* China
	Cs-k2.v2	C-k2.v1	Cs-c2
Number of genes/mRNA	13,489/ 14,408	14,936/ 14,936	13,634/ 13,634
Retained/new gene models	11,256/2233		
Gene length[a]	18,637 ± 21,269	15,143 ± 17,316	17,725 ± 19,238
mRNA length	1748 ± 1700	1437 ± 1464	1592 ± 1569
Coding domain length	1497 ± 1559	1437 ± 1464	1592 ± 1569
Exon length	254 ± 348	233 ± 323	231 ± 297
Intron length	2893 ± 4371	2745 ± 3732	2755 ± 3599
Protein length	498 ± 520	479 ± 488	529 ± 523
Completeness:			
Complete BUSCOs[b]	699 (73.3%)	609 (63.8%)	564 (59.1%)
Complete single-copy BUSCOs	691 (72.4%)	603 (63.2%)	554 (58.1%)
Complete and duplicated BUSCOs	8 (0.8%)	6 (0.6%)	10 (1.1%)
Fragmented BUSCOs	42 (4.4%)	76 (8.0%)	116(12.2%)

[a] Lengths presented as mean ± standard deviation.
[b] Number of Benchmarking Universal Single-Copy Orthologs (BUSCOs) identified (protein-mode), and percentage of the 954 genes within the Metazoa_odb10 dataset.

Table 3
Features of transferred and revised/new gene models in the *Clonorchis sinensis* reference gene set (Cs-k2.v2).

	Transferred gene models (CSKR_1)	Revised/new gene models (CSKR_2)
Number of genes	11,256	2233
mRNA length	1743 ± 1603[a]	1433 ± 1876
Coding domain length	1515 ± 1478	1212 ± 1721
Untranslated regions (UTRs)		
Number of genes with 5′–UTRs	5718 (51%)[b]	1134 (51%);1181 (53%)
5′–UTR length	104 ± 137	128 ± 326
Number of genes with 3′–UTRs	6314 (56%)	1134 (51%);1181 (53%)
3′–UTR length	313 ± 279	295 ± 335
Evidence of transcription[c]	8288 (74%)	1765 (79%)
Homology searches		
InterProScan	7399 (66%)	1078 (48%)
PFAM domains	6259 (56%)	863 (39%)
Gene Ontology results	5512 (49%)	750 (34%)
Biological process	3212 (29%)	481 (22%)
Molecular function	4424 (39%)	599 (27%)
Cellular component	1873 (17%)	319 (14%)

[a] Lengths presented as mean ± standard deviation.
[b] Number of genes and percentage of total gene models.
[c] Genes with transcription support in metacercariae, juvenile and/or adult stages (more than 0.5 counts per million).

Most (11,256 of 13,489; 83.4%) gene models for Cs-k2.v2 were transferred successfully from Cs-k2.v1 and updated (CSKR_1 accession numbers), and 2233 new or revised gene models (CSKR_2 accession numbers) were added (Tables 2 & 3; Supplementary Table S3). The characteristics of the gene models were usually consistent with previous data sets for Cs-k2.v1 and Cs-c2 (Table 2; Fig. 2), and more complete, with fewer fragmented or duplicate BUSCOs being present in the gene set for Cs-k2.v2. The messenger RNA sequences inferred were markedly

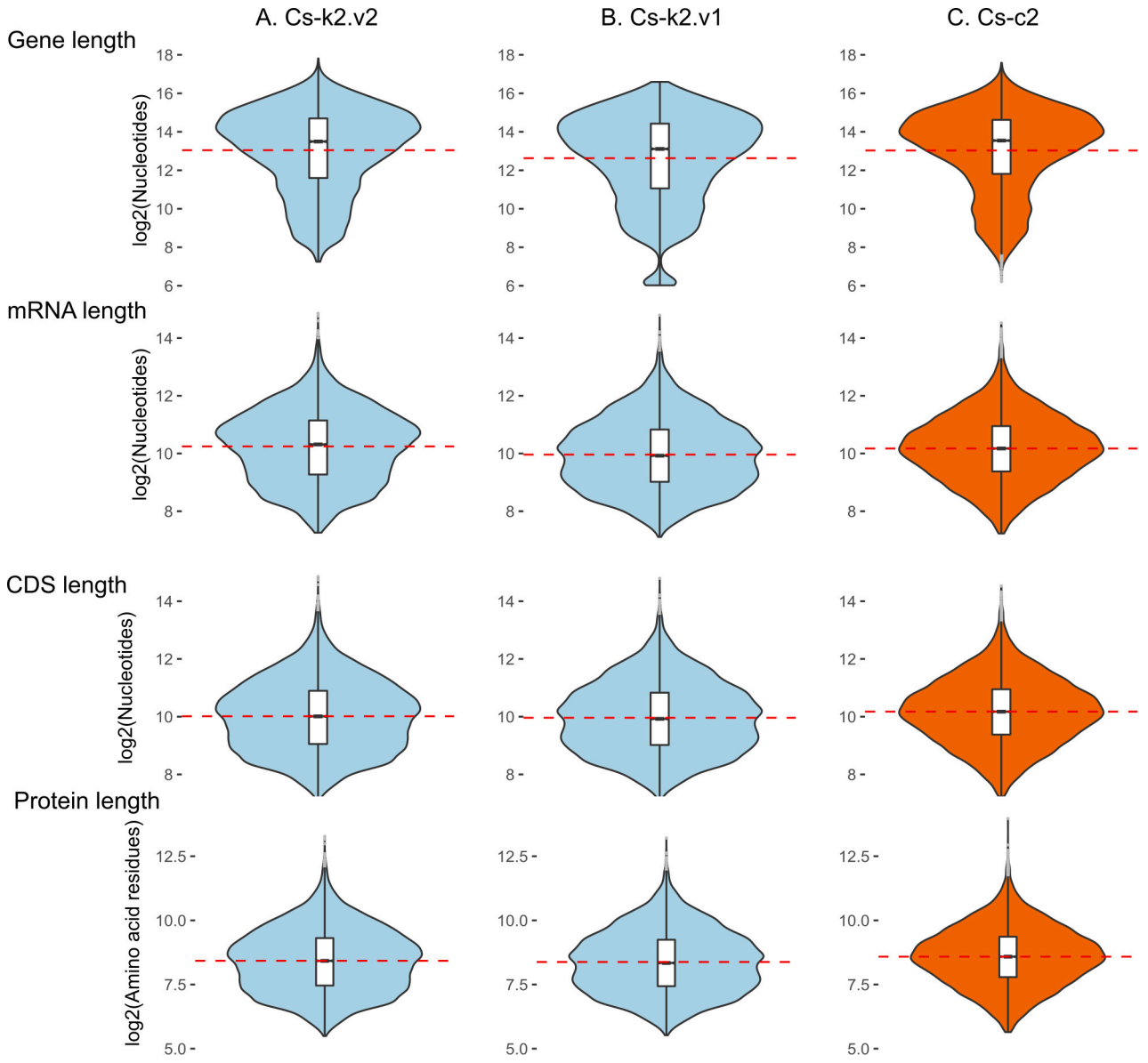

Fig. 2. Characteristics of the gene sets inferred for *Clonorchis sinensis*. Distribution of gene, mRNA, coding domain and protein lengths for gene sets representing *C. sinensis* from Korea (Cs-k2.v2 and Cs-k2.v1) and *C. sinensis* from China (Cs-cn2). Length distributions are displayed using violin and box plots, with red dashed lines indicating average lengths. Coding sequence (CDS); messenger RNA (mRNA). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

longer as ~50% of gene models included untranslated regions (UTRs) (Tables 2 & 3; Fig. 2).

Average mRNA (CSKR_1: 1743 ± 1603 bp; CSKR_2: 1433 ± 1876 bp), coding sequence (CSKR_1: 1515 ± 1478 bp; CSKR_2: 1212 ± 1721 bp) and 5′- (CSKR_1: 104 ± 137 bp; CSKR_2: 128 ± 326 bp) and 3′- UTR lengths (CSKR_1: 313 ± 279 bp; CSKR_2: 295 ± 335 bp) of 11,256 transferred (CSKR_1) and 2233 revised/new gene models (CSKR_2) were similar, with considerable length variation being a feature common to both gene sets (Table 3). There was support for gene transcription in metacercarial, juvenile and/or adult *C. sinensis* stages for most transferred (CSKR_1: 74%) and revised/new gene models (CSKR_2: 79%), although a higher proportion of transferred gene models encoded proteins with homology to conserved InterProScan (66%) and PFAM (56%) domains than the revised/new gene models (InterProScan: 48%; PFAM: 39%). Similarly, more transferred gene models (CSKR_1: 49%) could be annotated with Gene Ontology terms than in the revised/new gene set (CSKR_2: 39%). Specifically, 29%, 39% and 17% of transferred gene models and 22%, 27% and 14% of revised/new gene models were associated with 'biological process', 'molecular function' and 'cellular component' terms, respectively. There was no significant enrichment in ontology terms for either gene set.

3.3. Synteny

Upon pairwise comparison, there was marked synteny between the Cs-k2.v2 assembly and previous draft genome assemblies for *C. sinensis* (Cs-k2.v1 and Cs-c2) as well as the genomes of *S. mansoni* and *E. multilocularis* (Fig. 3). Synteny was greatest between Cs-k2.v2 and Cs-k2.v1 (Fig. 3A; Table 4), with 96.1% ($n = 459$) of scaffolds in the latter aligning to 99.3% ($n = 9$) of the former in 527 circos-bundles containing 541 syntenic blocks (62.0% of all syntenic blocks inferred). Synteny between Cs-k2.v2 and Cs-c2 was somewhat limited (Fig. 3B; Table 4), with 64.8% of the Cs-c2 scaffolds ($n = 686$) aligning to 99.2% of the Cs-k2.v2 genome ($n = 7$) in 838 circos-bundles containing 900 syntenic

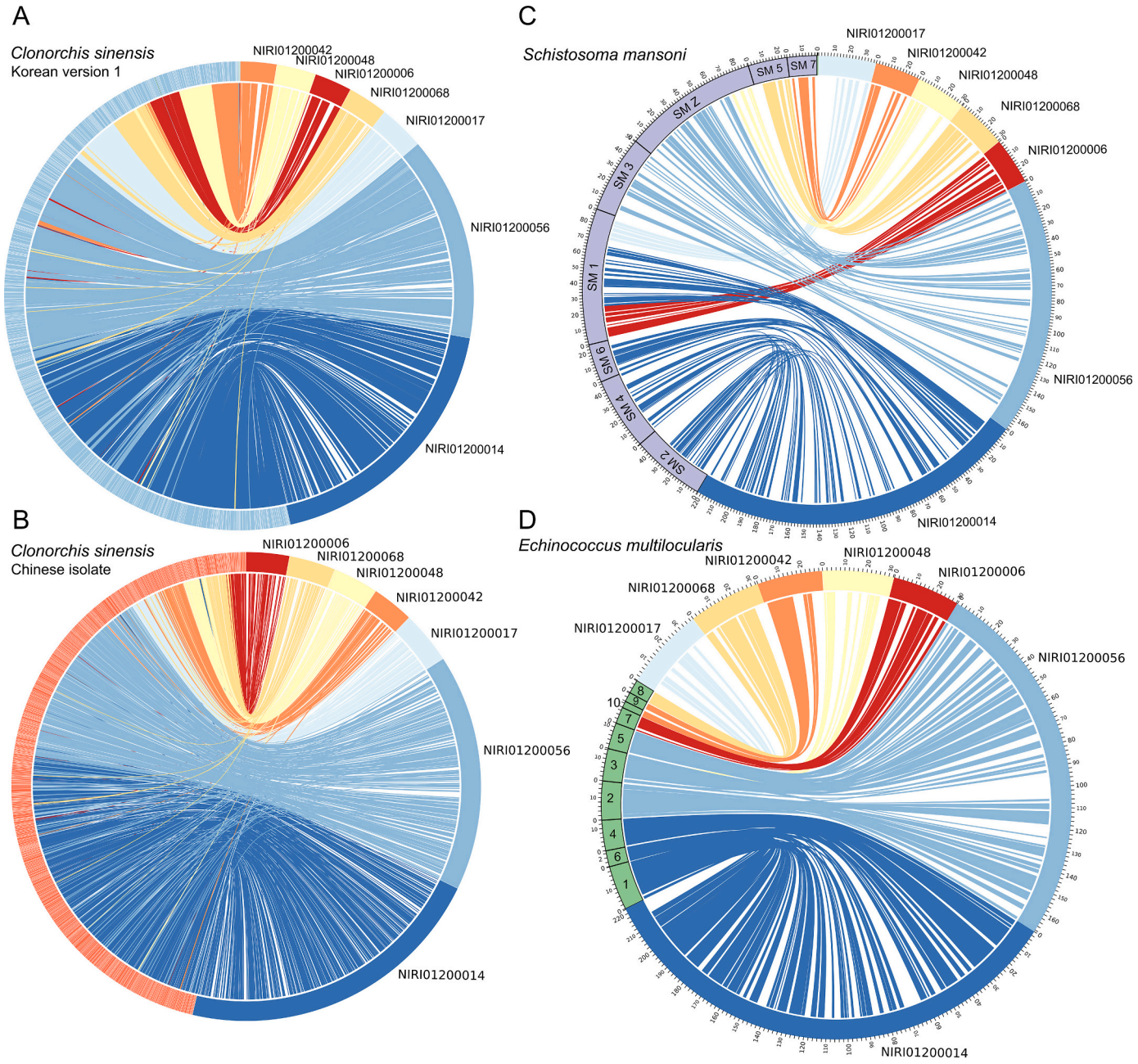

Fig. 3. Synteny and contiguity of the *Clonorchis sinensis* reference genome (Cs-k2.v2) with respect to the draft genomes of *C. sinensis* from Korea (Cs-k2.v1) or China (Cs-c2), and the chromosomal-level reference genomes of the blood fluke *Schistosoma mansoni* and the taeniid tapeworm *Echinococcus multilocularis*. Scaffolds are arranged in circular (circos) plots with seven reference scaffolds (=inferred chromosomes with NIRI012-codes; right) for *C. sinensis* and linked syntenic blocks in distinctly-coloured bars. Panel A: Alignment of the masked draft genome Cs-k2.v1 (light blue; version 1) with reference genome Cs-k2.v2; scaffolds are aligned in 527 bundles containing 541 syntenic blocks (62.0% of the alignment). Panel B: Alignment of the masked draft genome Cs-c2 (light red) with reference genome Cs-k2.v2; scaffolds are aligned in 838 bundles containing 900 blocks (18.7% of the alignment). Panel C: Chromosomal synteny between the reference genomes of *S. mansoni* and *C. sinensis* (Cs-k2.v2) established based on the positions of 163 syntenic blocks each containing five or more single-copy orthologs (SCOs) (i.e. 1108 of a total of 4917). Panel D: Chromosomal synteny between the reference genomes of *E. multilocularis* and *C. sinensis* (Cs-k2.v2) established based on the position of 177 syntenic blocks containing five or more SCOs (i.e. 1613 of a total of 4136). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

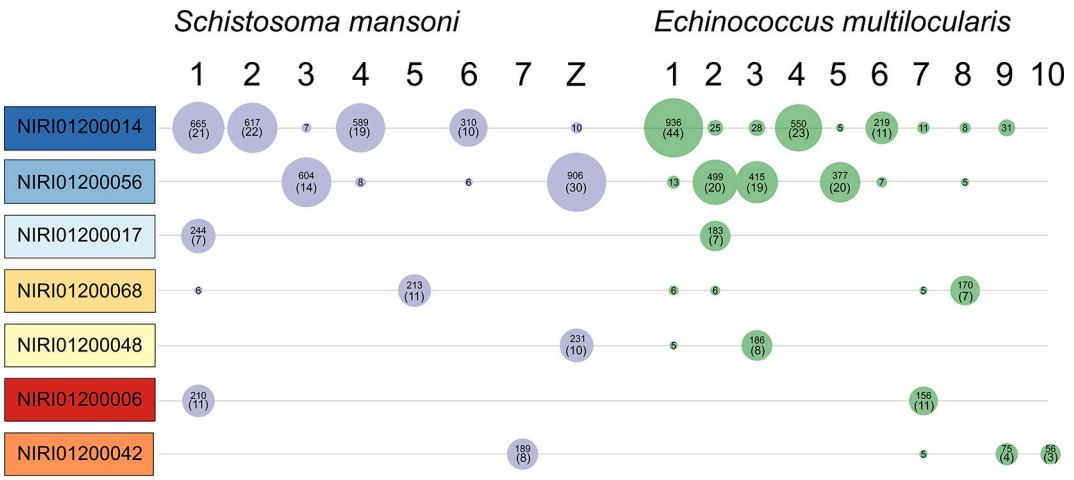

Fig. 4. Bubble plot of single copy orthologs (SCOs) and syntenic bundles, each with five or more SCOs (in parentheses), between the seven individual scaffolds (=inferred chromosomes with NIRI012-codes; left) of the *Clonorchis sinensis* genome (Cs-k2.v.2) and individual chromosomes of *Schistosoma mansoni* [59] or *Echinococcus multilocularis* [60].

Table 4
Genome-wide synteny comparisons of the genomes of *Clonorchis sinensis* (Cs-k2.v2), *Schistosoma mansoni* (Sman) and *Echinococcus multilocularis* (Emul) in a pairwise manner.

	Syntenic scaffolds			Scaffolds with bundled links			Length of scaffolds with bundled links (percentage of genome assembly)			Number of links	Number of passed bundles	Number of links in passed bundles
	Cs-c2	Cs-k2.v1	Cs-k2.v2	Cs-c2	Cs-k2.v1	Cs-k2.v2	Cs-c2	Cs-k2.v1	Cs-k2.v2			
Cs-k2.v2 vs. Cs-k2.v1		668	69		459	9		540,610,759 (96.1%)	554,462,252 (99.3%)	872	527	541 (62.04%)
Cs-k2.v2 vs. Cs-c2	2371		61	686		7	354,388,787 (64.8%)		553,904,272 (99.2%)	4699	838	900 (19.15%)
	Sman	Emul	Cs-k2.v2	Sman	Emul	Cs-k2.v2	Sman	Emul	Cs-k2.v2			
Cs-k2.v2 vs. Sman	12		37	8		9	392,668,716 (95.9%)		554,462,252 (99.3%)	4917	163	1108 (22.5%)
Cs-k2.v2 vs. Emul		16	33		10	7		106,004,081 (92.2%)	553,904,272 (99.2%)	4136	177	1613 (39.0%)

blocks – equating to 19.2% of all syntenic blocks inferred. QUAST metrics suggested more genomic rearrangements or mis-assemblies in Cs-c2 than in Cs-k2.v1 (cf. Table 1).

The seven chromosomes inferred for Cs-k2.v2 were syntenic to regions within the *S. mansoni* and *E. multilocularis* genomes containing 392,668,716 (95.9%; $n = 8$ chromosomes) and 106,004,081 (92.2%; $n = 10$ chromosomes) nucleotides, respectively (Table 4). There were 163 syntenic blocks (with 1108 of all 4917 SCOs) between Cs-k2.v2 and *S. mansoni* (Fig. 3C; Table 4), and 177 blocks (with 1613 of 4136 SCOs) between Cs-k2.v2 and *E. multilocularis* (Fig. 3D; Table 4).

In syntenic regions, chromosomal arrangements were then compared (Figs. 3 & 4). Between Cs.k2.v2 and *S. mansoni*, a one–to–one relationship was seen for the five small chromosomes (i.e. NIR01200068–SM5, NIR01200048–Z, NIR01200042–SM7, NIR01200017–SM1 and NIR01200006–SM1), whereas the two large chromosomes inferred for Cs-k2.v2 related to more than one *S. mansoni* chromosome (i.e. NIRI01200056–SM3/Z and NIRI01200014–SM1/SM2/SM4/SM6). A similar pattern was seen between Cs-k2.v2 and *E. multilocularis*, with one–to–one relationships for four small chromosomes (NIR01200068–EM8, NIR01200048–EM3, NIR01200017–EM2 and NIR01200006–EM71), and regions in three inferred Cs-k2.v2 chromosomes, each linking to more than one chromosome (NIRI01200056–EM2/EM3/EM5, NIR012 00042–EM9/EM10 and NIRI01200014–EM1/EM4/EM6) (Fig. 4).

4. Discussion

Despite recent advances in parasite genomics [11,13,15,18,19], many published draft genomes of helminths are fragmented, including those of socioeconomically important liver flukes (see WormBase ParaSite; https://parasite.wormbase.org/index.html). To tackle such fragmentation in the genomic assembly of the key carcinogenic fluke, *C. sinensis* [15], we employed in situ Hi-C and long-read data sets, together with available short-read data, to assemble the first complete reference genome (558 Mb) for this opisthorchid trematode. Presently, this is the largest helminth genome assembled to chromosome-contiguity, now providing an excellent resource for comparative genomic and molecular parasitological research.

Here, we showed that the contiguity of this reference genome led to major improvements in the inferred gene set for *C. sinensis*, including 2233 new or revised gene models. The revised/new gene models resulted from the improved contiguity of genome scaffolds with no evidence of an enrichment of one or more gene ontology terms when compared with the Cs-k2.v2 gene set. The RNA-Seq libraries representing metacercarial, immature and adult *C. sinensis* stages, respectively, provided transcriptional support for most (~75%) genes and 5′– and/or 3′–UTRs for more than ~50% of gene models. Herein, on average, for *C. sinensis*, 5′–UTRs (~104 bp) and 3′–UTRs (~313 bp) were shorter than the curated UTRs in *S. mansoni* (5′ UTR: 348 ± 534 bp; 3′ UTR: 650 ± 819 bp) and other invertebrates (5′–UTR: 222 bp; 3′–UTR: 445 bp) [62]. Shorter UTRs could reflect compact genes in *C. sinensis* or a limitation in accurately predicting full-length UTRs using in silico methods [63]. The latter limitation appears to be reflected in suboptimal annotation of genes of flatworms and many other invertebrates [43,64]. In future, mRNA and/or 5′–end RNA sequencing [65] of multiple developmental stages (including miracidium, sporocyst, redia and cercaria) as well as long-read RNA-Seq data, should assist in curating gene models. Moreover, continued efforts are required to annotate presently unknown genes using orthology-based assignment [66], paving the way for subsequent studies of gene expression and regulation in this worm. For example, elucidating the relationships/interactions of new full-length mRNAs with non-coding RNAs available for *C. sinensis* [67] would be informative. Logically extending previous work [e.g., [18,19,68,69–71]], we expect that this new genomic resource for *C. sinensis* will significantly facilitate studies of the evolution of protein families, population genetics and comparative genomics of flatworms.

Published works [72–76] indicate that *C. sinensis* and related species exhibit significant genetic complexity and variability. Importantly, there is substantial variation in the karyotype of *C. sinensis* between studies. For example, Zadesenets et al. [77] reported seven chromosome pairs ($2n = 14$), whereas Park [78] recorded 28 ($2n = 56$), and other, allozyme electrophoretic investigations [72] provide evidence of cryptic species. The seven largest scaffolds assembled here (i.e. inferred chromosomes) correspond to a karyotype with five pairs of small chromosomes and two pairs of large chromosomes ($2n = 14$) (Figs. 1 and 3), in accordance with the findings of Zadesenets et al. [77].

The synteny of large regions in the two large chromosomes to six chromosomes of *S. mansoni* or *E. multilocularis* suggests an interesting evolution of chromosomes in flatworms. Although presently unknown, we hypothesise that the arrangement of the two large chromosomes (>200 Mb) in *C. sinensis* is consistent with an ancestral flatworm karyotype. There seems to be a trend for more and smaller chromosomes in 'decendant' trematode species, based on current karyotype information [79–81]. For example, karyotypes can vary from as few as six pairs in *Opisthorchis viverrini* [79] to 12 pairs in *Fascioloides magna* [81]. Beyond liver flukes, the karyotype of the (basal) monogenean, *Eudiplozoon nipponicum*, has five pairs, which is the smallest number reported to date for a flatworm [82]. This evidence stimulates the hypothesis that basal trematodes have fewer chromosomes. We believe that comparative cytogenetic and genomic investigations of flatworm species are warranted to study the evolution of chromosomal arrangements, to test this hypothesis and to attempt to establish the evolutionary processes or mechanisms that have led to chromosomal fusion and/or fragmentation events in the Trematoda.

The sex chromosome (Z) of dioecious trematodes – schistosomes – has been the subject of extensive study [59,80,81,83–85]. The synteny observed here between the "core ancestral" Z region (first 10 to 30 Mb) of the chromosome of *S. mansoni* [83] and the beginning of 'chromosome' NIRI01200056 (first 5 to 80 Mb) of *C. sinensis* as well as chromosome 5 of the hermaphroditic tapeworm *E. multilocularis* [60] suggest that genes in this region could be involved in sexual differentiation, development and reproduction. Although almost nothing is known about the sex chromosome system and the evolution of gene dosage-compensation in hermaphroditic trematodes [86], work on the *Z*-chromosome gene content of African (*S. mansoni* and *S. haematobium*) and Asian (*S. japonicum*) schistosomes and lineage-specific evolutionary strata, using genomic, transcriptomic and proteomic approaches, has indicated that schistosomes have a reduced expression of Z-linked genes in females and an upregulation in both females and males [83], in accordance with the first step of Ohno's classic model of the evolution of dosage-compensation [87]. This information appears to provide a basis to start comparing and contrasting the evolution of gene-dosage of the chromosome responsible for sex determination in dioecious and hermaphroditic trematodes as well as the processes of sexual differentiation within hermaphrodites, and of reproductive biology at the genetic and molecular levels. The *C. sinensis* reference genome now provides a sound platform to explore this area.

The completion of a reference genome for *C. sinensis* provides improved insights into the genomic landscape of a cancer-causing parasite, responsible for a significant neglected tropical disease in Asia [3]. The reference genome and revised gene models can now underpin post-genomic explorations of the parasite's molecular biology, disease pathogenesis and the link between *C. sinensis* and cholangiocarcinoma. For instance, the genome will enable detailed analyses of genetic variation within *C. sinensis*, the discovery/characterisation of cryptic species [72] and definition of genetic markers to identify sibling species, if they exist. Importantly, the enhanced genomic resource will underpin functional genomic work on selected stages of *C. sinensis*. Given that gene-specific knockdown by double-stranded RNA interference (RNAi) [88–90] and CRiSPR/Cas systems [22] are known to work in liver flukes, we believe that genome-assisted drug target discovery could provide a complementary approach to the screening of drugs for new anthelmintics [8,81,91]. A future aim will be to identify genes or expressed

proteins whose inactivation by one or more drugs will selectively kill *C. sinensis* and not harm the mammalian host, interfere with reproductive processes and/or disrupt the parasite's life cycle.

Recent advances in the prediction of gene essentiality from functional genomic information (e.g., lethality) for other metazoan organisms employing machine learning-coupled approaches (e.g., [92,93]) can now be combined with RNAi- or CRISPR/Cas9-based screening of *C. sinensis* to construct a functional genomics-informatics workflow to validate predicted and prioritised targets. Focussing on select panels of molecules, such as peptidases, GPCRs, kinases and ion channels, and understanding their involvement in the host-parasite interplay and/or disease, should assist in the design of new drugs or a vaccine against *C. sinensis*.

Although this study focussed on the genome of an isolate of *C. sinensis* from Korea, the exquisite effectiveness of our approach of using combined sequence data sets and advanced informatics to finish a large genome (>500 Mb) should be readily applicable to other flatworm parasites of major animal and human health importance globally. We believe that the present, substantially-improved *C. sinensis* genome will accelerate both fundamental and applied investigations of clonorchiasis and cholangiocarcinoma, and enable the development of new interventions for this important neglected tropical disease-complex.

Acknowledgements

This project was supported through grants LP180101334 (N.D.Y. and P.K.K.) and LP180101085 (R.B.G. and BCCH) from the Australian Research Council (ARC). P.K.K. held an Early Career Research Fellowship from the National Health and Medical Research Council (NHMRC) of Australia. The in situ Hi-C sequencing and analysis was supported by DNA Zoo (Baylor College of Medicine) and DNA Zoo – Australia (The University of Western Australia), with additional computational resources and support from the Pawsey Supercomputing Centre, and funding from the Australian Government and the Government of Western Australia. E.L.A. was supported by an NSF Physics Frontiers Center Award (PHY1427654), the Welch Foundation (Q-1866), a USDA Agriculture and Food Research Initiative Grant (2017-05741), and an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.03.001.

References

[1] P.J. Hotez, M.E. Bottazzi, U. Strych, L.Y. Chang, Y.A. Lim, M.M. Goodenow, et al., Neglected tropical diseases among the Association of Southeast Asian Nations (ASEAN): overview and update, PLoS Negl. Trop. Dis. 9 (4) (2015), e0003575, https://doi.org/10.1371/journal.pntd.0003575 (Epub 2015/04/17. PubMed PMID: 25880767; PubMed Central PMCID: PMCPMC4400050).

[2] D. Wang, N.D. Young, P.K. Korhonen, R.B. Gasser, *Clonorchis sinensis* and Clonorchiasis: the relevance of exploring genetic variation, Adv. Parasitol. 100 (2018) 155–208, https://doi.org/10.1016/bs.apar.2018.03.006 (Epub 2018/05/14; PubMed PMID: 29753338).

[3] Z.R. Lun, R.B. Gasser, D.H. Lai, A.X. Li, X.Q. Zhu, X.B. Yu, et al., Clonorchiasis: a key foodborne zoonosis in China, Lancet Infecious Disease 5 (1) (2005) 31–41, https://doi.org/10.1016/S1473-3099(04)01252-6 (Epub 2004/12/29. PubMed PMID: 15620559).

[4] M.B. Qian, J. Utzinger, J. Keiser, X.N. Zhou, Clonorchiasis, Lancet 387 (10020) (2016) 800–810, https://doi.org/10.1016/S0140-6736(15)60313-0 (PubMed PMID: 26299184).

[5] T.S. Kim, J.H. Pak, J.B. Kim, Y.Y. Bahk, *Clonorchis sinensis*, an oriental liver fluke, as a human biological agent of cholangiocarcinoma: a brief review, BMB Rep. 49 (11) (2016) 590–597 (PubMed PMID: 27418285; PubMed Central PMCID: PMCPMC5346318).

[6] V. Bouvard, R. Baan, K. Straif, Y. Grosse, B. Secretan, F. El Ghissassi, et al., A review of human carcinogens–Part B: biological agents, Lancet Oncol. 10 (4) (2009) 321–322, https://doi.org/10.1016/s1470-2045(09)70096-8 (Epub 2009/04/08. PubMed PMID: 19350698).

[7] B.I. Choi, J.K. Han, S.T. Hong, K.H. Lee, Clonorchiasis and cholangiocarcinoma: etiologic relationship and imaging diagnosis, Clin. Microbiol. Rev. 17 (3) (2004) 540–552, https://doi.org/10.1128/CMR.17.3.540-552.200417/3/540 (Epub 2004/07/20. PubMed PMID: 15258092).

[8] M.B. Qian, C.H. Zhou, H.H. Zhu, T.J. Zhu, J.L. Huang, Y.D. Chen, et al., From awareness to action: NIPD's engagement in the control of food-borne clonorchiasis, Adv. Parasitol. 110 (2020) 245–267, https://doi.org/10.1016/bs.apar.2020.04.004 (Epub 2020/06/22. PubMed PMID: 32563327).

[9] M.J. Doenhoff, P. Hagan, D. Cioli, V. Southgate, L. Pica-Mattoccia, S. Botros, et al., Praziquantel: its use in control of schistosomiasis in sub-Saharan Africa and current research needs, Parasitology. 136 (13) (2009) 1825–1835, https://doi.org/10.1017/S0031182009000493 (Epub 2009/03/14. PubMed PMID: 19281637).

[10] Z.L. Tang, Y. Huang, X.B. Yu, Current status and perspectives of *Clonorchis sinensis* and clonorchiasis: epidemiology, pathogenesis, omics, prevention and control, Infect. Dis. Poverty. 5 (1) (2016) 71, https://doi.org/10.1186/s40249-016-0166-1 (PubMed PMID: 27384714; PubMed Central PMCID: PMCPMC4933995).

[11] International Helminth Genomes C. Comparative genomics of the major parasitic worms, Nat. Genet. 51 (1) (2019) 163–174, https://doi.org/10.1038/s41588-018-0262-1 (Epub 2018/11/07. PubMed PMID: 30397333; PubMed Central PMCID: PMCPMC6349046).

[12] G. Rinaldi, N.D. Young, J.D. Honeycutt, P.J. Brindley, R.B. Gasser, M.H. Hsieh, New research tools for urogenital schistosomiasis, J. Infect. Dis. 211 (6) (2015) 861–869, https://doi.org/10.1093/infdis/jiu527 (PubMed PMID: 25240172; PubMed Central PMCID: PMC4416124).

[13] Y. Huang, W. Chen, X. Wang, H. Liu, Y. Chen, L. Guo, et al., The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes, PLoS ONE 8 (1) (2013), e54732, https://doi.org/10.1371/journal.pone.0054732 (Epub 2013/02/06. PubMed PMID: 23382950; PubMed Central PMCID: PMCPMC3559784).

[14] X. Wang, W. Chen, Y. Huang, J. Sun, J. Men, H. Liu, et al., The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*, Genome Biol. 12 (10) (2011) R107, https://doi.org/10.1186/gb-2011-12-10-r107 (Epub 2011/10/26. PubMed PMID: 22023798).

[15] D. Wang, P.K. Korhonen, R.B. Gasser, N.D. Young, Improved genomic resources and new bioinformatic workflow for the carcinogenic parasite *Clonorchis sinensis*: biotechnological implications, Biotechnol. Adv. 36 (4) (2018) 894–904, https://doi.org/10.1016/j.biotechadv.2018.02.008 (Epub 2018/02/20. PubMed PMID: 29454982).

[16] W.G. Yoo, D.-W. Kim, J.-W. Ju, P.Y. Cho, T.I. Kim, S.-H. Cho, et al., Developmental transcriptomic features of the carcinogenic liver fluke, *Clonorchis sinensis*, PLoS Negl. Trop. Dis. 5 (6) (2011), e1208.

[17] N.D. Young, B.E. Campbell, R.S. Hall, A.R. Jex, C. Cantacessi, T. Laha, et al., Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*, PLoS Negl. Trop. Dis. 4 (6) (2010), e719, https://doi.org/10.1371/journal.pntd.0000719 (Epub 2010/06/29. PubMed PMID: 20582164; PubMed Central PMCID: PMC2889816).

[18] N.D. Young, N. Nagarajan, S.J. Lin, P.K. Korhonen, A.R. Jex, R.S. Hall, et al., The *Opisthorchis viverrini* genome provides insights into life in the bile duct, Nat. Commun. 5 (2014) 4378, https://doi.org/10.1038/ncomms5378 (Epub 2014/07/10. PubMed PMID: 25007141; PubMed Central PMCID: PMCPMC4104445).

[19] N.I. Ershov, V.A. Mordvinov, E.B. Prokhortchouk, M.Y. Pakharukova, K.V. Gunbin, K. Ustyantsev, et al., New insights from *Opisthorchis felineus* genome: update on genomics of the epidemiologically important liver flukes, BMC Genomics 20 (1) (2019) 399, https://doi.org/10.1186/s12864-019-5752-8 (Epub 2019/05/24. PubMed PMID: 31117933; PubMed Central PMCID: PMCPMC6530080).

[20] P.K. Korhonen, R.S. Hall, N.D. Young, R.B. Gasser, Common workflow language (CWL)-based software pipeline for *de novo* genome assembly from long- and short-read data, Gigascience 8 (4) (2019), https://doi.org/10.1093/gigascience/giz014 (Epub 2019/03/02. PubMed PMID: 30821816).

[21] P.K. Korhonen, N.D. Young, R.B. Gasser, Making sense of genomes of parasitic worms: tackling bioinformatic challenges, Biotechnol. Adv. 34 (5) (2016) 663–686, https://doi.org/10.1016/j.biotechadv.2016.03.001 (PubMed PMID: 26956711).

[22] P. Arunsan, W. Ittiprasert, M.J. Smout, C.J. Cochran, V.H. Mann, S. Chaiyadet, et al., Programmed knockout mutation of liver fluke granulin attenuates virulence of infection-induced hepatobiliary morbidity, Elife. 8 (2019), https://doi.org/10.7554/eLife.41463 (Epub 2019/01/16. PubMed PMID: 30644359).

[23] E.R. Mardis, DNA sequencing technologies: 2006–2016, Nat. Protoc. 12 (2) (2017) 213–218, https://doi.org/10.1038/nprot.2016.182 (Epub 2017/01/06. PubMed PMID: 28055035).

[24] F. Luo, M. Yin, X. Mo, C. Sun, Q. Wu, B. Zhu, et al., An improved genome assembly of the fluke *Schistosoma japonicum*, PLoS Negl. Trop. Dis. 13 (8) (2019), e0007612, https://doi.org/10.1371/journal.pntd.0007612 (Epub 2019/08/08. PubMed PMID: 31390359; PubMed Central PMCID: PMCPMC6685614).

[25] A.J. Stroehlein, P.K. Korhonen, T.M. Chong, Y.L. Lim, K.G. Chan, B. Webster, et al., High-quality *Schistosoma haematobium* genome achieved by single-molecule and long-range sequencing, Gigascience 8 (9) (2019), https://doi.org/10.1093/gigascience/giz108 (Epub 2019/09/09. PubMed PMID: 31494670; PubMed Central PMCID: PMCPMC6736295).

[26] H. Oey, M. Zakrzewski, K. Gravermann, N.D. Young, P.K. Korhonen, G.N. Gobert, et al., Whole-genome sequence of the bovine blood fluke *Schistosoma bovis* supports interspecific hybridization with *S. haematobium*, PLoS Pathog. 15 (1) (2019), e1007513, https://doi.org/10.1371/journal.ppat.1007513 (Epub 2019/01/24. PubMed PMID: 30673782).

[27] H. Oey, M. Zakrzewski, K. Narain, K.R. Devi, T. Agatsuma, S. Nawaratna, et al., Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*, Gigascience 8 (1) (2019), https://doi.org/10.1093/gigascience/giy146 (Epub

2018/12/07. PubMed PMID: 30520948; PubMed Central PMCID: PMCPMC6329441).
[28] P.K. Korhonen, G. Ma, T. Wang, A.J. Stroehlein, N.D. Young, C.S. Ang, et al., High quality nuclear genome for *Sarcoptes scabiei* – a critical resource for a neglected parasite, PLoS Negl. Trop. Dis. 14 (2020), e0008720 (accepted).
[29] L. Kinkar, N.D. Young, W.-M. Sohn, A.J. Stroehlein, P.K. Korhonen, R.B. Gasser, First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach, PLoS Negl. Trop. Dis. 14 (8) (2020), e0008552, https://doi.org/10.1371/journal.pntd.0008552.
[30] L. Kinkar, P.K. Korhonen, H. Cai, C.G. Gauci, M.W. Lightowlers, U. Saarma, et al., Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1, Parasit. Vectors 12 (1) (2019) 238, https://doi.org/10.1186/s13071-019-3492-x (Epub 2019/05/18. PubMed PMID: 31097022; PubMed Central PMCID: PMCPMC6521400).
[31] S.S. Rao, M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, Cell 159 (7) (2014) 1665–1680, https://doi.org/10.1016/j.cell.2014.11.021 (Epub 2014/12/17. PubMed PMID: 25497547; PubMed Central PMCID: PMCPMC5635824).
[32] W.M. Sohn, H. Zhang, M.H. Choi, S.T. Hong, Susceptibility of experimental animals to reinfection with *Clonorchis sinensis*, Korean J. Parasitol. 44 (2) (2006) 163–166 (Epub 2006/07/01. doi: 200606163 [pii]. PubMed PMID: 16809966).
[33] P.J. Cock, C.J. Fields, N. Goto, M.L. Heuer, P.M. Rice, The sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res. 38 (6) (2010) 1767–1771, https://doi.org/10.1093/nar/gkp1137 (PubMed PMID: 20015970; PubMed Central PMCID: PMC2847217).
[34] N.C. Durand, M.S. Shamim, I. Machol, S.S. Rao, M.H. Huntley, E.S. Lander, et al., Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments, Cell Syst. 3 (1) (2016) 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (Epub 2016/07/29. PubMed PMID: 27467249; PubMed Central PMCID: PMCPMC5846465).
[35] O. Dudchenko, S.S. Batra, A.D. Omer, S.K. Nyquist, M. Hoeger, N.C. Durand, et al., De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds, Science 356 (6333) (2017) 92–95, https://doi.org/10.1126/science.aal3327 (Epub 2017/03/25. PubMed PMID: 28336562; PubMed Central PMCID: PMCPMC5635820).
[36] R. Kajitani, D. Yoshimura, M. Okuno, Y. Minakuchi, H. Kagoshima, A. Fujiyama, et al., Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions, Nat. Commun. 10 (1) (2019) 1702, https://doi.org/10.1038/s41467-019-09575-2 (Epub 2019/04/14. PubMed PMID: 30979905; PubMed Central PMCID: PMCPMC6461651).
[37] M.J. Roach, S.A. Schmidt, A.R. Borneman, Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies, BMC Bioinformatics. 19 (1) (2018) 460, https://doi.org/10.1186/s12859-018-2485-7 (Epub 2018/12/01. PubMed PMID: 30497373; PubMed Central PMCID: PMCPMC6267036).
[38] D. Kim, L. Song, F.P. Breitwieser, S.L. Salzberg, Centrifuge: rapid and sensitive classification of metagenomic sequences, Genome Res. 26 (12) (2016) 1721–1729, https://doi.org/10.1101/gr.210641.116 (Epub 2016/11/18. PubMed PMID: 27852649; PubMed Central PMCID: PMCPMC5131823).
[39] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2 (Epub 1990/10/05. PubMed PMID: 2231712).
[40] E.W. Sayers, R. Agarwala, E.E. Bolton, J.R. Brister, K. Canese, K. Clark, et al., Database resources of the National Center for Biotechnology Information, Nucleic Acids Res. 47 (D1) (2019), D23-D8, https://doi.org/10.1093/nar/gky1069 (Epub 2018/11/06. PubMed PMID: 30395293; PubMed Central PMCID: PMCPMC6323993).
[41] T.D. Wu, J. Reeder, M. Lawrence, G. Becker, M.J. Brauer, GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality, Methods Mol. Biol. 1418 (2016) 283–334, https://doi.org/10.1007/978-1-4939-3578-9_15 (Epub 2016/03/24. PubMed PMID: 27008021).
[42] M. Haeussler, A.S. Zweig, C. Tyner, M.L. Speir, K.R. Rosenbloom, B.J. Raney, et al., The UCSC Genome Browser database: 2019 update, Nucleic Acids Res. 47 (D1) (2019), D853-D8, https://doi.org/10.1093/nar/gky1095 (Epub 2018/11/09. PubMed PMID: 30407534; PubMed Central PMCID: PMCPMC6323953).
[43] K.L. Howe, B.J. Bolt, M. Shafie, P. Kersey, M. Berriman, WormBase ParaSite - a comprehensive resource for helminth genomics, Mol. Biochem. Parasitol. 215 (2017) 2–10, https://doi.org/10.1016/j.molbiopara.2016.11.005 (Epub 2016/12/03. PubMed PMID: 27899279; PubMed Central PMCID: PMCPMC5486357).
[44] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, et al., Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments, Genome Biol. 9 (1) (2008) R7, https://doi.org/10.1186/gb-2008-9-1-r7 (PubMed PMID: 18190707; PubMed Central PMCID: PMCPMC2395244).
[45] K.J. Hoff, M. Stanke, Predicting genes in single genomes with AUGUSTUS, Curr. Protoc. Bioinformatics 65 (1) (2019), e57, https://doi.org/10.1002/cpbi.57 (Epub 2018/11/23. PubMed PMID: 30466165).
[46] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (3) (2015) 290–295, https://doi.org/10.1038/nbt.3122 (Epub 2015/02/19. PubMed PMID: 25690850; PubMed Central PMCID: PMCPMC4643835).
[47] F.A. Simao, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics. 31 (19) (2015) 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (Epub 2015/06/11. PubMed PMID: 26059717).
[48] E.M. Zdobnov, R. Apweiler, InterProScan–an integration platform for the signature-recognition methods in InterPro, Bioinformatics. 17 (9) (2001) 847–848 (Epub 2001/10/09. PubMed PMID: 11590104).
[49] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernandez-Plaza, S.K. Forslund, H. Cook, et al., eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (D1) (2019), D309-D14, https://doi.org/10.1093/nar/gky1085 (Epub 2018/11/13. PubMed PMID: 30418610; PubMed Central PMCID: PMCPMC6324079).
[50] M. Magrane, U. Consortium, UniProt Knowledgebase: a hub of integrated protein data, Database (2011), 2011:bar009, https://doi.org/10.1093/database/bar009 (Epub 2011/03/31. PubMed PMID: 21447597; PubMed Central PMCID: PMC3070428).
[51] N.D. Rawlings, D.P. Tolle, A.J. Barrett, MEROPS: the peptidase database, Nucleic Acids Res. 32 (Database issue) (2004), D160-4, https://doi.org/10.1093/nar/gkh07132/suppl_1/D160 (Epub 2003/12/19. PubMed PMID: 14681384; PubMed Central PMCID: PMC308805).
[52] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (1) (2015) 59–60, https://doi.org/10.1038/nmeth.3176 (Epub 2014/11/18. PubMed PMID: 25402007).
[53] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (8) (2019) 907–915, https://doi.org/10.1038/s41587-019-0201-4 (Epub 2019/08/04. PubMed PMID: 31375807).
[54] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics. 30 (7) (2014) 923–930, https://doi.org/10.1093/bioinformatics/btt656 (Epub 2013/11/15. PubMed PMID: 24227677).
[55] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics. 26 (1) (2010) 139–140, https://doi.org/10.1093/bioinformatics/btp616 (Epub 2009/11/17. PubMed PMID: 19910308; PubMed Central PMCID: PMC2796818).
[56] B. Paten, D. Earl, N. Nguyen, M. Diekhans, D. Zerbino, D. Haussler, Cactus: algorithms for genome multiple sequence alignment, Genome Res. 21 (9) (2011) 1512–1528, https://doi.org/10.1101/gr.123356.111 (Epub 2011/06/15. PubMed PMID: 21665927; PubMed Central PMCID: PMCPMC3166836).
[57] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (8) (2013) 1072–1075, https://doi.org/10.1093/bioinformatics/btt086 (Epub 2013/02/21. PubMed PMID: 23422339; PubMed Central PMCID: PMCPMC3624806).
[58] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, et al., Circos: an information aesthetic for comparative genomics, Genome Res. 19 (9) (2009) 1639–1645, https://doi.org/10.1101/gr.092759.109 (Epub 2009/06/23. PubMed PMID: 19541911; PubMed Central PMCID: PMCPMC2752132).
[59] A.V. Protasio, I.J. Tsai, A. Babbage, S. Nichol, M. Hunt, M.A. Aslett, et al., A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*, PLoS Negl. Trop. Dis. 6 (1) (2012), e1455, https://doi.org/10.1371/journal.pntd.0001455 (Epub 2012/01/19. PubMed PMID: 22253936; PubMed Central PMCID: PMC3254664).
[60] I.J. Tsai, M. Zarowiecki, N. Holroyd, A. Garciarrubio, A. Sanchez-Flores, K. L. Brooks, et al., The genomes of four tapeworm species reveal adaptations to parasitism, Nature. 496 (7443) (2013) 57–63, https://doi.org/10.1038/nature12031 (Epub 2013/03/15. PubMed PMID: 23485966).
[61] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (1) (2019) 238, https://doi.org/10.1186/s13059-019-1832-y (Epub 2019/11/16. PubMed PMID: 31727128; PubMed Central PMCID: PMCPMC6857279).
[62] F. Mignone, C. Gissi, S. Liuni, G. Pesole, Untranslated regions of mRNAs, Genome Biol. 3 (3) (2002), REVIEWS0004, https://doi.org/10.1186/gb-2002-3-3-reviews0004 (Epub 2002/03/19. PubMed PMID: 11897027; PubMed Central PMCID: PMCPMC139023).
[63] S.L. Salzberg, Next-generation genome annotation: we still struggle to get it right, Genome Biol. 20 (1) (2019) 92, https://doi.org/10.1186/s13059-019-1715-2 (Epub 2019/05/18. PubMed PMID: 31097009; PubMed Central PMCID: PMCPMC6521345).
[64] G. Grillo, A. Turi, F. Licciulli, F. Mignone, S. Liuni, S. Banfi, et al., UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs, Nucleic Acids Res. 38 (Database issue) (2010), D75-80, https://doi.org/10.1093/nar/gkp902 (Epub 2009/11/03. PubMed PMID: 19880380; PubMed Central PMCID: PMCPMC2808995).
[65] X. Adiconis, A.L. Haber, S.K. Simmons, A. Levy Moonshine, Z. Ji, M.A. Busby, et al., Comprehensive comparative analysis of 5′-end RNA-sequencing methods, Nat. Methods 15 (7) (2018) 505–511, https://doi.org/10.1038/s41592-018-0014-2 (Epub 2018/06/06. PubMed PMID: 29867192; PubMed Central PMCID: PMCPMC6075671).
[66] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. von Mering, et al., Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper, Mol. Biol. Evol. 34 (8) (2017) 2115–2122, https://doi.org/10.1093/molbev/msx148 (Epub 2017/05/02. PubMed PMID: 28460117; PubMed Central PMCID: PMCPMC5850834).
[67] M.J. Xu, Q. Liu, A.J. Nisbet, X.Q. Cai, C. Yan, R.Q. Lin, et al., Identification and characterization of microRNAs in *Clonorchis sinensis* of human health significance, BMC Genomics 11 (2010) 521, https://doi.org/10.1186/1471-2164-11-521 (Epub 2010/10/06. PubMed PMID: 20920166).
[68] C. Wang, H. Lei, Y. Tian, M. Shang, Y. Wu, Y. Li, et al., *Clonorchis sinensis* granulin: identification, immunolocalization, and function in promoting the metastasis of

cholangiocarcinoma and hepatocellular carcinoma, Parasit. Vectors 10 (1) (2017) 262, https://doi.org/10.1186/s13071-017-2179-4 (PubMed PMID: 28545547; PubMed Central PMCID: PMCPMC5445496).
[69] Y.A. Bae, J.G. Kim, Y. Kong, Phylogenetic characterization of *Clonorchis sinensis* proteins homologous to the sigma-class glutathione transferase and their differential expression profiles, Mol. Biochem. Parasitol. 206 (1–2) (2016) 46–55, https://doi.org/10.1016/j.molbiopara.2016.01.002 (PubMed PMID: 26792248).
[70] J.M. Kang, J. Lee, H.L. Ju, J.W. Ju, J.H. Kim, J.H. Pak, et al., Characterization of a gut-associated asparaginyl endopeptidase of *Clonorchis sinensis*, Exp. Parasitol. 153 (2015) 81–90, https://doi.org/10.1016/j.exppara.2015.03.015 (PubMed PMID: 25819296).
[71] M. Anari, A.J. Stroehlein, R.S. Hall, B.C.H. Chang, R.B. Gasser, N.D. Young, Expanded complement of Niemann-Pick type C2-like protein genes in *Clonorchis sinensis* suggests functions beyond sterol binding and transport, Parasit. Vectors 13 (1) (2020) 38, https://doi.org/10.1186/s13071-020-3910-0 (Epub 2020/01/25. PubMed PMID: 31973758; PubMed Central PMCID: PMCPMC6979364).
[72] T.N. Petney, R.H. Andrews, W. Saijuntha, A. Wenz-Mucke, P. Sithithaworn, The zoonotic, fish-borne liver flukes *Clonorchis sinensis, Opisthorchis felineus* and *Opisthorchis viverrini*, Int. J. Parasitol. 43 (12−13) (2013) 1031–1046, https://doi.org/10.1016/j.ijpara.2013.07.007 (Epub 2013/08/28. PubMed PMID: 23978669).
[73] L. Kinkar, P.K. Korhonen, D. Wang, X.Q. Zhu, G.N. Chelomina, T. Wang, et al., Marked mitochondrial genetic variation in individuals and populations of the carcinogenic liver fluke *Clonorchis sinensis*, PLoS Negl. Trop. Dis. 14 (8) (2020) e0008480, https://doi.org/10.1371/journal.pntd.0008480 (Epub 2020/08/20. PubMed PMID: 32813714; PubMed Central PMCID: PMCPMC7437864).
[74] N. Laoprom, P. Sithithaworn, R.H. Andrews, K. Ando, T. Laha, S. Klinbunga, et al., Population genetic structuring in *Opisthorchis viverrini* over various spatial scales in Thailand and Lao PDR, PLoS Negl. Trop. Dis. 6 (11) (2012), e1906, https://doi.org/10.1371/journal.pntd.0001906 (PubMed PMID: 23166853; PubMed Central PMCID: PMCPMC3499411).
[75] N. Kiatsopit, P. Sithithaworn, W. Saijuntha, O. Pitaksakulrat, T.N. Petney, J. P. Webster, et al., Analysis of the population genetics of *Opisthorchis viverrini sensu lato* in the Nam Ngum River wetland, Lao PDR, by multilocus enzyme electrophoresis, Parasitol. Res. 113 (8) (2014) 2973–2981, https://doi.org/10.1007/s00436-014-3959-9 (PubMed PMID: 25033811).
[76] G.N. Chelomina, Y.V. Tatonova, N.M. Hung, H.D. Ngo, Genetic diversity of the Chinese liver fluke *Clonorchis sinensis* from Russia and Vietnam, Int. J. Parasitol. 44 (11) (2014) 795–810, https://doi.org/10.1016/j.ijpara.2014.06.009 (PubMed PMID: 25123068).
[77] K.S. Zadesenets, A.V. Katokhin, V.A. Mordvinov, N.B. Rubtsov, Comparative cytogenetics of opisthorchiid species (Trematoda, Opisthorchiidae), Parasitol. Int. 61 (1) (2012) 87–89, https://doi.org/10.1016/j.parint.2011.07.006 (Epub 2011/07/30. PubMed PMID: 21798365).
[78] G.M. Park, K. Im, S. Huh, T.S. Yong, Chromosomes of the liver fluke, *Clonorchis sinensis*, Korean J. Parasitol. 38 (3) (2000) 201–206 (Epub 2000/09/26. PubMed PMID: 11002660; PubMed Central PMCID: PMC2728209).
[79] W. Kaewkong, W. Choochote, P. Kanla, W. Maleewong, P.M. Intapan, S. Wongkham, et al., Chromosomes and karyotype analysis of a liver fluke, *Opisthorchis viverrini,* by scanning electron microscopy, Parasitol. Int. 61 (3) (2012) 504–507, https://doi.org/10.1016/j.parint.2012.03.008 (Epub 2012/04/17. PubMed PMID: 22504416).
[80] R.B. Short, M.Y. Menzel, Chromosomes of nine species of schistosomes, J. Parasitol. 46 (1960) 273–287 (Epub 1960/06/01. PubMed PMID: 14446186).
[81] M. Reblanova, M. Spakulova, M. Orosova, I. Kralova-Hromadova, E. Bazsalovicsova, D. Rajsky, A comparative study of karyotypes and chromosomal location of rDNA genes in important liver flukes *Fasciola hepatica* and *Fascioloides magna* (Trematoda: Fasciolidae), Parasitol. Res. 109 (4) (2011) 1021–1028, https://doi.org/10.1007/s00436-011-2339-y (Epub 2011/04/22. PubMed PMID: 21509448).
[82] E. Koskova, M. Spakulova, B. Koubkova, M. Reblanova, M. Orosova, Comparative karyological analysis of four diplozoid species (Monogenea, Diplozoidae), gill parasites of cyprinid fishes, Parasitol. Res. 108 (4) (2011) 935–941, https://doi.org/10.1007/s00436-010-2135-0 (Epub 2010/10/29. PubMed PMID: 20981442).
[83] M.A.L. Picard, C. Cosseau, S. Ferre, T. Quack, C.G. Grevelding, Y. Coute, et al., Evolution of gene dosage on the Z-chromosome of schistosome parasites, Elife 7 (2018), https://doi.org/10.7554/eLife.35684 (Epub 2018/07/26. PubMed PMID: 30044216; PubMed Central PMCID: PMCPMC6089595).
[84] A.I. Grossman, R.B. Short, G.D. Cain, Karyotype evolution and sex chromosome differentiation in Schistosomes (Trematoda, Schistosomatidae), Chromosoma. 84 (3) (1981) 413–430 (Epub 1981/01/01. PubMed PMID: 7327052).
[85] S. Lawton, H. Hirai, J. Ironside, D. Johnston, D. Rollinson, Genomes and geography: genomic insights into the evolution and phylogeography of the genus *Schistosoma*, Parasit. Vectors 4 (1) (2011) 131, https://doi.org/10.1186/1756-3305-4-131.
[86] S.A. Ramm, Exploring the sexual diversity of flatworms: ecology, evolution, and the molecular biology of reproduction, Mol. Reprod. Dev. 84 (2) (2017) 120–131, https://doi.org/10.1002/mrd.22669 (Epub 2016/06/14. PubMed PMID: 27292123).
[87] D.W. Hall, M.L. Wayne, Ohno's "peril of hemizygosity" revisited: gene loss, dosage compensation, and mutation, Genome Biol. Evol. 5 (1) (2013) 1–15, https://doi.org/10.1093/gbe/evs106 (Epub 2012/11/30. PubMed PMID: 23193178; PubMed Central PMCID: PMCPMC3607304).
[88] M.Y. Pakharukova, V.A. Vavilin, B. Sripa, T. Laha, P.J. Brindley, V.A. Mordvinov, Functional analysis of the unique cytochrome P450 of the liver fluke *Opisthorchis felineus*, PLoS Negl. Trop. Dis. 9 (12) (2015), e0004258, https://doi.org/10.1371/journal.pntd.0004258 (Epub 2015/12/02. PubMed PMID: 26625139; PubMed Central PMCID: PMCPMC4666407).
[89] X. Wang, W. Chen, Y. Tian, Y. Huang, X. Li, X. Yu, RNAi-mediated silencing of enolase confirms its biological importance in *Clonorchis sinensis*, Parasitol. Res. 113 (4) (2014) 1451–1458, https://doi.org/10.1007/s00436-014-3785-0 (Epub 2014/01/25. PubMed PMID: 24458653).
[90] J. Sripa, P. Pinlaor, P.J. Brindley, B. Sripa, S. Kaewkes, M.W. Robinson, et al., RNA interference targeting cathepsin B of the carcinogenic liver fluke, *Opisthorchis viverrini*, Parasitol. Int. 60 (3) (2011) 283–288, https://doi.org/10.1016/j.parint.2011.04.003 (Epub 2011/05/14. PubMed PMID: 21565281; PubMed Central PMCID: PMCPMC3682765).
[91] C. Machicado, M.P. Soto, O. Timoteo, A. Vaisberg, M. Pajuelo, P. Ortiz, et al., Screening the pathogen box for identification of new chemical agents with anti-*Fasciola hepatica* activity, Antimicrob. Agents Chemother. 63 (3) (2019), https://doi.org/10.1128/AAC.02373-18 (Epub 2019/01/04. PubMed PMID: 30602522; PubMed Central PMCID: PMCPMC6395923).
[92] T.L. Campos, P.K. Korhonen, R.B. Gasser, N.D. Young, An evaluation of machine learning approaches for the prediction of essential genes in eukaryotes using protein sequence-derived features, Comput. Struct. Biotechnol. J. 17 (2019) 785–796, https://doi.org/10.1016/j.csbj.2019.05.008 (Epub 2019/07/18. PubMed PMID: 31312416; PubMed Central PMCID: PMCPMC6607062).
[93] T.L. Campos, P.K. Korhonen, P.W. Sternberg, R.B. Gasser, N.D. Young, Predicting gene essentiality in *Caenorhabditis elegans* by feature engineering and machine-learning, Comput. Struct. Biotechnol. J. 18 (2020) 1093–1102, https://doi.org/10.1016/j.csbj.2020.05.008 (Epub 2020/06/04. PubMed PMID: 32489524; PubMed Central PMCID: PMCPMC7251299).
[94] J.T. Robinson, D. Turner, N.C. Durand, H. Thorvaldsdottir, J.P. Mesirov, E. L. Aiden, Juicebox.js provides a cloud-based visualization system for Hi-C data, Cell Syst. 6 (2) (2018), 256-8e1, https://doi.org/10.1016/j.cels.2018.01.001 (Epub 2018/02/13. PubMed PMID: 29428417; PubMed Central PMCID: PMCPMC6047755).