RESEARCH ARTICLE

Draft genome of the bluefin tuna blood fluke, *Cardicola forsteri*

Lachlan Coff[1¤], **Andrew J. Guy**[1,2], **Bronwyn E. Campbell**[1], **Barbara F. Nowak**[1,3], **Paul A. Ramsland**[1,4,5], **Nathan J. Bott**[1] *

1 School of Science, STEM College, RMIT University, Bundoora, Victoria, Australia, 2 ZiP Diagnostics, Collingwood, Victoria, Australia, 3 Institute for Marine and Antarctic Studies, University of Tasmania, Launceston, Tasmania, Australia, 4 Department of Immunology, Monash University, Melbourne, Victoria, Australia, 5 Department of Surgery, Austin Health, University of Melbourne, Heidelberg, Victoria, Australia

¤ Current address: Australian Centre for Disease Preparedness, CSIRO, East Geelong, Victoria, Australia
* Nathan.Bott@rmit.edu.au

Abstract

The blood fluke *Cardicola forsteri* (Trematoda: Aporocotylidae) is a pathogen of ranched bluefin tuna in Japan and Australia. Genomics of *Cardicola* spp. have thus far been limited to molecular phylogenetics of select gene sequences. In this study, sequencing of the *C. forsteri* genome was performed using Illumina short-read and Oxford Nanopore long-read technologies. The sequences were assembled *de novo* using a hybrid of short and long reads, which produced a high-quality contig-level assembly (N50 > 430 kb and L50 = 138). The assembly was also relatively complete and unfragmented, comprising 66% and 7.2% complete and fragmented metazoan Benchmarking Universal Single-Copy Orthologs (BUSCOs), respectively. A large portion (> 55%) of the genome was made up of intergenic repetitive elements, primarily long interspersed nuclear elements (LINEs), while protein-coding regions cover > 6%. Gene prediction identified 8,564 hypothetical polypeptides, > 77% of which are homologous to published sequences of other species. The identification of select putative proteins, including cathepsins, calpains, tetraspanins, and glycosyltransferases is discussed. This is the first genome assembly of any aporocotylid, a major step toward understanding of the biology of this family of fish blood flukes and their interactions within hosts.

Introduction

Cardicola forsteri (Trematoda: Aporocotylidae) is a blood fluke that infects bluefin tunas (*Thunnus* spp.) [1, 2]. *C. forsteri* parasitises *Thunnus* spp. hearts in its adult life cycle stage, while adult *C. orientalis* are found in the branchial arteries of the gills [3, 4]. Additionally, the co-infecting species *C. opisthorchis* is found in the hearts of Pacific bluefin tuna (PBT, *Thunnus orientalis*) [5], but has not been found in southern bluefin tuna (SBT, *Thunnus maccoyii*). While less prevalent in wild populations, infections with *Cardicola* spp. represents the most significant health issue for ranched bluefin tuna, a multimillion-dollar industry operating in Japan, Australia, and the Mediterranean [6]. Treatment with praziquantel (PZQ) has significantly reduced mortalities of SBT since its implementation in 2013, and *C. forsteri* is now the

dominant *Cardicola* spp. infecting SBT [7]. Despite the success of PZQ in the treatment of *Cardicola* infections, the required holding period and the risk of resistance has prompted research into alternative control measures. A variety of animal husbandry practices can also reduce the burden of helminthiases in ranched bluefin tunas, but additional pharmacological and immunological interventions would benefit the industry [8–10]. However, very little is known about *Cardicola* spp. beyond morphological characterisations and identifications of their intermediate hosts [11–13]. A full genome sequence of *C. forsteri* could help to answer fundamental questions of its biology and facilitate transcriptomic and proteomic investigations.

The genomes of the human blood flukes *Schistosoma mansoni* and *S. japonicum* (Trematoda: Schistosomatidae), were first published in 2009 using chain-termination (Sanger) sequencing technology [14, 15]. The advent of Illumina sequencing dramatically reduced sequencing costs, so genomes for other digenean trematodes, including *S. haematobium*, *Fasciola hepatica*, and *F. gigantica*, were published in the 2010s [16–18]. However, despite the relatively high base call accuracy of Illumina sequencing, its short-read lengths alone cannot bridge the large repeat regions typical of these genomes, so these initial assemblies are relatively discontiguous and fragmented. The addition of Third-Generation sequencing technologies, namely PacBio single molecule real time sequencing (SMRT) and Oxford Nanopore sequencing, has greatly improved these assemblies. The *S. mansoni* genome is now in its ninth revision (GCA_000237925.5), > 95% of which is assembled into seven autosomes and two sex chromosomes [19].

Cardicola was first diagnosed by Robert B. Short in 1953 [20], and 34 novel species have since been added to the genus, primarily based on morphological characterisations, making *Cardicola* the most speciose genus within the Aporocotylidae. Until now, only ribosomal 28S, internal transcribed spacer 2 (ITS-2), and mitochondrial cytochrome c oxidase subunit I (*cox*1) sequences of *C. forsteri* have been published, and these were recently used to demonstrate polyphyly in the genus and to reclassify other species in the genus [21]. Whole genome sequencing (WGS) could further aid in molecular phylogenetics of Aporocotylidae and their evolutionary relationships to other trematodes. Additionally, putative functional information can be mined from hypothetical proteins, which would direct further studies into the structural biology and host–parasite interactions of *Cardicola* spp.

Vaccines have been implemented in aquaculture since the 1980s to control a variety of infections [22], and immunization of farmed bluefin tunas with immunodominant antigens could be a viable control measure for infections with *Cardicola* spp. [1]. While there are currently no approved vaccines against helminthiases for fishes, several targets are undergoing human clinical trials, and some vaccines are approved to control helminthiases in mammalian livestock [23]. These vaccines chiefly interfere with digestive enzymes or target the surface tegument, which is the primary interface of platyhelminths accessible by the host immune system. A functionally annotated genome assembly of *C. forsteri* would facilitate further research into host–parasite interactions and rational vaccine design by homology to known vaccine targets in other digeneans. In this study, we present the first draft of the *C. forsteri* genome, assembled *de novo* from a hybrid of short-read (Illumina) and long-read (Oxford Nanopore) sequences. Predicted genes were functionally annotated, and putative glycosyltransferases and vaccine targets are discussed.

Methods

Specimen collection

Whole adult *Cardicola forsteri* specimens were flushed from the hearts of wild-caught southern bluefin tuna (SBT, *T. maccoyii*), ranched in the lower Spencer Gulf, according to the protocol

described by Aiken et al. [24]. These SBT were from untreated pontoons collected during the 2019 harvest, as per Power et al. [25]. Sampling was performed during harvesting under animal ethics approval (RMIT Animal Ethics Committee #22802) from specimens after euthanasia, which was performed by commercial SBT companies using industry best-practice techniques. The *C. forsteri* specimens were stored in RNAlater® at -20 ˚C.

Genomic DNA extraction and sequencing

RNAlater® was completely washed from the specimens with Tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.6). For Illumina short-read sequencing, genomic (g)DNA from two adult specimens was extracted separately using the ISOLATE II Genomic DNA Kit (Bioline), according to the manufacturer's instructions. The yield of double-stranded (ds)DNA in each sample was measured by Qubit™ 4 Fluorometer (Thermo Fisher Scientific) using the dsDNA HS Assay Kit, according to the manufacturer's instructions. The purity of each sample was indicated with A260/280 and A260/230 ratios using the NanoDrop™ One microvolume spectrophotometer (Thermo Fisher Scientific), according to the manufacturer's instructions. These two samples were sent to the Ramaciotti Centre for Genomics for library preparation and sequencing. The short-read libraries from each sample were prepared using Nextera DNA Flex Library Prep Kit (Illumina®, Inc.), according to the manufacturer's instructions. Each sample was sequenced in paired ends of 150-bp lengths using the NextSeq® 500 System, which yields 300–550 bp insert size (Illumina®, Inc.).

For Nanopore long-read sequencing, high-molecular weight (HMW) gDNA was extracted from approximately 30 pooled adult specimens using the MagAttract® HMW DNA Kit (QIAGEN), according to the manufacturer's instructions. The yield of dsDNA in the sample was measured by Qubit™ 4 Fluorometer, according to the manufacturer's instructions. Purity was indicated with A_{260}/A_{280} and A_{260}/A_{230} ratios using the NanoDrop™, according to the manufacturer's instructions. Fragmentation was assessed using the TapeStation automated electrophoresis system (Agilent Technologies, Inc.), according to the manufacturer's instructions. This sample was also sent to the Ramaciotti Centre for Genomics for library preparation and sequencing. The HMW gDNA extracted from the pooled specimens was sequenced with GridION MK1 (Oxford Nanopore Technologies Ltd.), using the FLO-MIN106 flow cell and Ligation Sequencing Kit (SQK-LSK110). Long-read sequences were base called using Guppy v4.3.4 High-Accuracy model [26].

Trimming and removal of host contaminants from the short-read library

Base call quality scores were assessed using FastQC. To improve basecall accuracy, four nucleotides were trimmed from the 3′ end of each paired-end read in the short-read library using Trimmomatic v0.36 [27].

Each short-read sequence was queried against the genomes of Pacific bluefin tuna (*T. orientalis*, GCA_009176245.1) and *S. mansoni* (GCA_000237925.4) using pBLAT v2.5 with default parameters [28, 29]. These reference genomes were chosen as the closest relatives of the host (*T. maccoyii*) and parasite (*C. forsteri*), respectively, with high-quality published genomes [29]. Of the matching reads with high query-coverages (116–146 bp), those that matched only with the *T. orientalis* and not the *S. mansoni* genome were removed from the short-read library. Species identity was confirmed *in silico* using pBLAT, with complete sequence identity of *C. forsteri* ITS-2 (AB742428.1) and 28S (AB742426.1) nucleotide sequences to the short-read library [3, 21].

Short-read assembly and *k*-mer optimisation

Short reads for each specimen were assembled using the Hamiltonian de Bruijn graph assembler ABySS v2.1.5 [30], with *k*-mers from 50 to 102 and a minimum *k*-mer depth of four (kc parameter). The contiguity statistics for each assembly were then examined, and the *k*-mer that produced the highest N50 and *E*-size, and lowest L50 for each of the specimens was selected as the optimal *k*-mer for short-read assembly of that specimen.

Estimation of the size, ploidy, and heterozygosity of the *C. forsteri* genome

A *k*-mer depth plot for each specimen was constructed with Jellyfish v2.3.0 [31], and this was used to determine the peak *k*-mer depth (D') for estimation of the haploid genome size, according to the following equations:

$$D = \frac{D'l}{l - k + 1}$$

and

$$G = \frac{n_{read}(l - k + 1)}{D'} = \frac{n_{k-mer}}{D'} = \frac{n_{base}}{D}$$

where G is haploid genome size, D is read depth, l is average read length, and k is *k*-mer length [32].

The *k*-mer depth plot was also used as the input for GenomeScope [33], which estimated the haploid size, repeat length, and heterozygosity of the *C. forsteri* genome. Ploidy was estimated from this *k*-mer depth plot using Smudgeplot v0.2.5 [34].

De novo assembly of hybrid contigs

The genome was assembled *de novo* using the short-read contigs from each of the ABySS assemblies and long reads as inputs for Wengan v0.2 [35]. Briefly, Wengan corrects short-read contigs by splitting chimeric contigs that lack paired-end support. Then, a synthetic mate-pair library is constructed from the long reads, which are mapped to the corrected short-read contigs, spanning the gaps and repeat regions. This information, along with the full long reads, is used to construct a synthetic scaffolding graph and, ultimately, hybrid contigs. The Wengan input accounted for the size and coverage of the long reads: large genome size (> 500 Mb, -g parameter), insert sizes of synthetic mate-pair reads 0.5–20 kb (-i parameter), 3 long-reads required to keep a potentially erroneous mate-edge (-N parameter), and 5 kb as the minimum length of reduced paths to convert them to physical fragments (-P parameter).

Assessment of contiguity and completeness of the genome assembly

Contiguity statistics of the hybrid contigs for each assembly were calculated using QUAST v5.0.2 [36]. The completeness of each assembly was assessed using BUSCO v5.2.2 [37] against the metazoa_odb10 database, which comprises 954 Benchmarking Universal Single-Copy Orthologs (BUSCOs). These statistics were compared with those of published genome assemblies of other trematodes, using the same metazoan database for a like-for-like comparison (S1 Table).

Repeat masking and annotation of the draft genome

A *de novo* repeat library was compiled from the *C. forsteri* genome assembly using RepeatModeler v2.0.2a [38], which comprises RECON v1.08 [39], RepeatScout v1.0.6 [40], LTRharvest

v1.5.9 [41], and LTR_retriever v2.9.0 [42], and these repetitive elements were softmasked with RepeatMasker v4.1.2-p1 [43].

For consistency, the genome assembly with the greater contiguity and completeness was selected from the two for gene prediction and functional annotation. *Ab initio* and homology-based gene prediction were performed using the BRAKER2 pipeline [44]. This pipeline firstly executed the ProtHint pipeline, which generated a set of seed genes from the *C. forsteri* genome assembly using the self-training *ab initio* gene prediction tool GeneMark-ES [45]. These seed genes were then translated into seed proteins, which were queried against a database of reference proteins (*S. mansoni*, GCF_000237925.1) using DIAMOND [46]. The output from ProtHint was ultimately used to train GeneMark-EP+ [47], which generated a set of anchored genes to train AUGUSTUS v3.4.0 [48] for the final output of predicted genes. Hypothetical polypeptide sequences were translated from the predicted genes, and the completeness of the gene set was assessed using BUSCO in protein mode.

For functional annotation, the hypothetical polypeptide sequences were queried using BLASTp (*E*-value $\leq 10^{-5}$) against the curated Swiss-Prot database [49]. Select sequences were further characterised as putative proteases, glycosyltransferases, ribonucleases, calpains, cation channels, tetraspanins (TSPs), glutathione *S*-transferases (GSTs), TGF-β homologs, and fatty acid-binding proteins (FABPs) based on BLASTp matches to the National Center for Biotechnology Information's (NCBI's) non-redundant protein sequences (NR) database and motif identification using HMMER v3.2.2 [50] with the Pfam database v35.0 [51] and CAZy database [52]. Signal peptides were predicted using SignalP 6.0 (probability ≥ 0.95) [53], transmembrane domains were predicted using DeepTMHMM v0.0.47 (https://biolib.com/DTU/DeepTMHMM), and N-glycosylation sites were predicted using NetNGlyc v1.0 [54].

Phylogenetic analyses

Outgroup-rooted phylogenetic trees of *C. forsteri* cathepsins and fucosyltransferases were constructed using the maximum likelihood method, with homologous proteins from SBT set as the outgroup. The protein sequences were aligned with MUSCLE v3.8.1551 [55], and these alignments were constructed as phylogenetic trees using RAxML v8.2.12 [56] with the Gamma model of rate heterogeneity, the WAG amino acid substitution model, and the majority rule consensus tree criterion.

Results

Sequencing and assembly

Approximately 450 ng of dsDNA was extracted from each *C. forsteri* specimen selected for short-read sequencing, and Illumina sequencing yielded 86 million paired-end reads from Specimen 1 and 127 million from Specimen 2 (Table 1). Due to the higher read output and the contiguity of the resulting assemblies, only the Specimen 2 assembly is reported here, but more information on the Specimen 1 assembly is available in S1 Table. Approximately 0.2–0.3% of short reads were likely derived from the host (*T. maccoyii*) and were removed from the library prior to assembly. The optimal *k*-mer length was found to be 77, based on the N50, *E*-size, and L50 of the short-read assemblies (Fig 1A). Using this optimal *k*-mer, the peak *k*-mer depth (***D'***) is 15, followed by a secondary peak, which indicates a large number of repetitive bases (Fig 1B). The haploid genome size was estimated to be 530–600 Mb by calculating the number of *k*-mers over peak *k*-mer depth, whereas GenomeScope estimated 230–236 Mb, with 25–30 Mb (12–15%) of repetitive elements. Smudgeplot proposed *C. forsteri* to be diploid in its adult life cycle stage, based on the grouping of *k*-mer pairs into haplotypes. As the genome size and ploidy of *C. forsteri* have not been empirically validated, the GenomeScope

Table 1. Statistics from short-read Illumina paired-end sequencing of *Cardicola forsteri*.

Number of short reads	126,882,928
Number of short reads derived from the host	278,175 (0.22%)
Optimal *k*-mer	77
Estimated haploid genome size (Mb)	230
Estimated short-read coverage	80×
Estimated repeat length (Mb)	25 (12%)
Estimated heterozygosity (%)	0.14
Assembled genome size (bp) [a]	216,531,548 (94%)

[a] Total sequence length of the hybrid assembly.

estimate was selected for further calculations. Using the GenomeScope estimate of haploid genome size, short-read sequence coverage is 80×.

For long-read sequencing, approximately 19 μg of dsDNA was extracted from pooled *C. forsteri* specimens with a DNA integrity number (DIN) of 6.9, which indicates moderate degradation, and a modal fragment size of 15,640 bp (5–58 kb range). Nanopore sequencing yielded 4.97 million reads with an N50 of 5.5 kb and a total of 13.68 Gb (90.79% passed). Long-read coverage was estimated to be > 53×, using the haploid genome size estimate of 230–236 Mb. The short-read library was assembled with ABySS, and these short-read contigs were combined with the pooled long-read data and assembled with Wengan. The size of the final hybrid assembly is 217 Mb. The GC content of the *C. forsteri* genome is > 28%, which is lower than that of other platyhelminths (schistosome GC content > 34%) [57].

Contiguity and completeness of hybrid assembly

An important quality metric of genome assemblies is contiguity, with larger contigs indicating a less fragmented assembly. The assembly is highly contiguous, with 1,532 hybrid (assembled from short and long reads) contigs (N50 = 430,422 and L50 = 138) (Fig 2A), and the largest contig is > 3 Mb. These contigs are more contiguous than those of the latest *F. hepatica* assembly (S1 Table), which comprises only short-read sequences. The BUSCO (Benchmarking Universal Single-Copy Ortholog) analysis indicates a relatively complete assembly, with 66.0% of metazoan BUSCOs found complete in the assembly and only 7.2% fragmented (Fig 2B). This
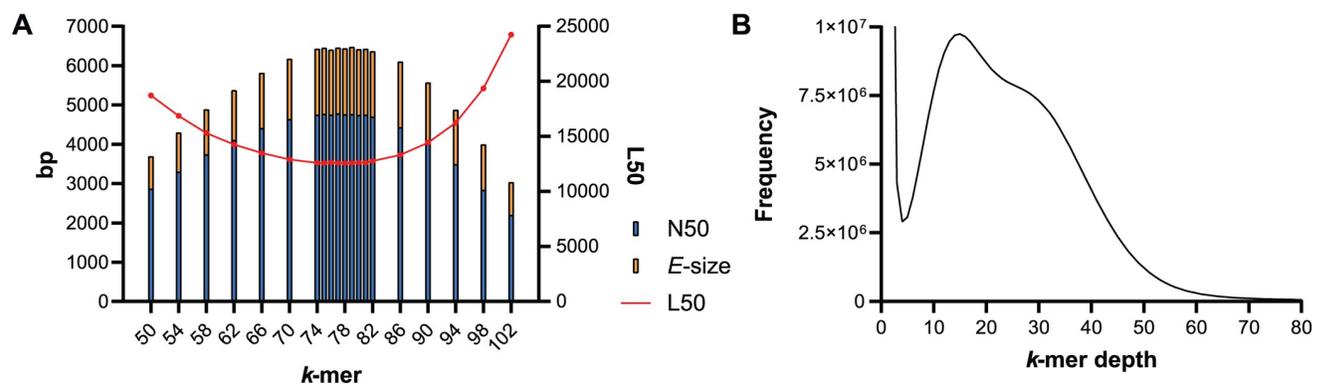

Fig 1. Analyses of *k*-mers in the *Cardicola forsteri* short-read library. (A) Contiguity statistics for the short-read assembly of different *k*-mers. N50 and *E*-size are shown as an overlaid column graph, while L50 is shown as an overlaid line graph with a secondary *y*-axis. (B) Plot of *k*-mer depth using the optimal *k*-mer 77.
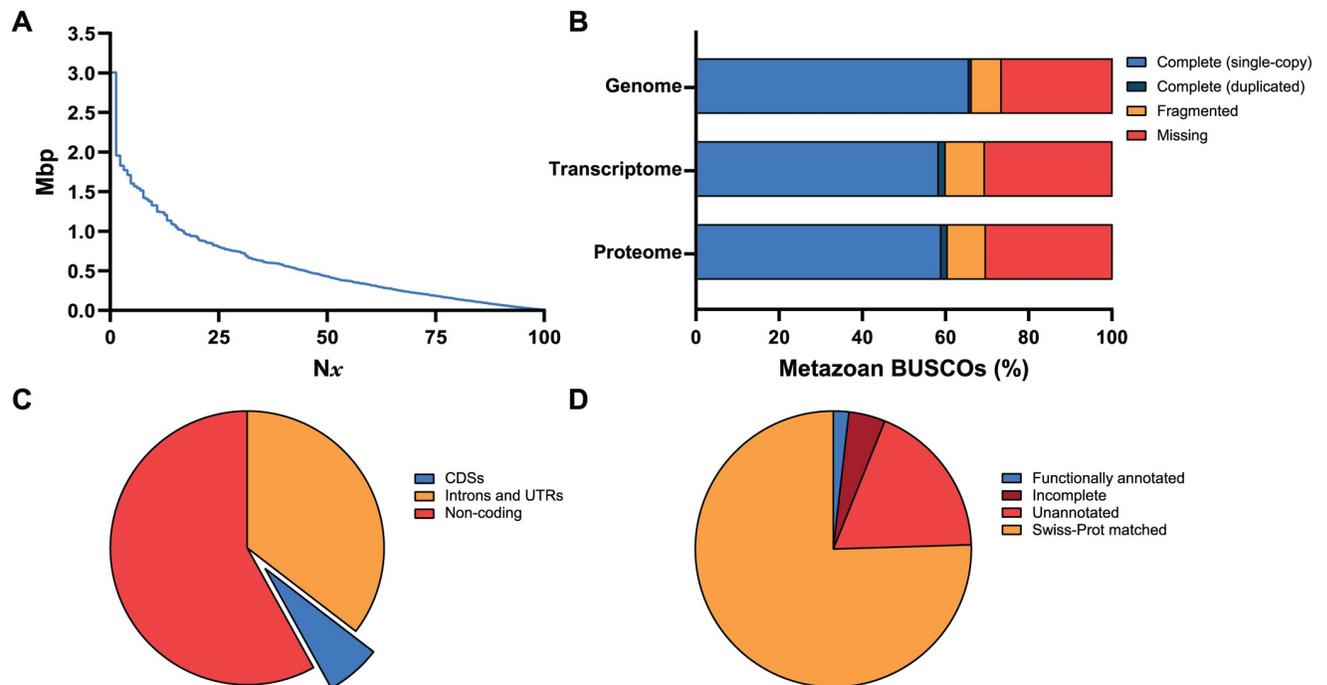
Fig 2. Contiguity, completeness, and gene prediction and annotation statistics of the *Cardicola forsteri* genome assemblies. (A) N*x* plot of the genome assembly, depicting the contig size distribution. (B) BUSCO statistics for the genome assembly, and the hypothetical transcripts and gene set (proteins), measured as a proportion of complete, fragmented, and missing BUSCOs from the metazoa_odb10 database. (C) Pie chart showing the sizes (bp) of predicted CDSs, introns, and UTRs as a proportion of the genome ($n = 216,531,548$). (D) Pie chart showing the number of hypothetical polypeptides ($n = 8,564$) that were functionally annotated, matched to the Swiss-Prot database or not (unannotated), and incomplete. BUSCOs, Benchmarking Universal Single-Copy Orthologs; CDS, coding sequence; UTR, untranslated region.

https://doi.org/10.1371/journal.pone.0276287.g002

also supports *C. forsteri* as a more complete assembly than that of *F. hepatica*, which has 65.4% and 10.5% complete and fragmented BUSCOs, respectively. For further comparison, *S. mansoni* BUSCOs are 71.5% complete and 5.0% fragmented.

Repetitive elements

Contrary to the estimates of GenomeScope, a large proportion of the *C. forsteri* genome was identified by RepeatModeler as repetitive and was masked prior to genome annotation. These repetitive elements comprise 124,057,400 bp (57.29%) of the assembly (Table 2). This

Table 2. Repetitive elements classified by repeat modeler and masked prior to the annotation of the *Cardicola forsteri* genome.

Repetitive elements	Number	Length (Mb)
LINEs	173,788	50.39 (23.27%)
LTRs	10,709	3.82 (1.76%)
DNA transposons	33,663	12.04 (5.56%)
Small RNA repeats	30,817	2.85 (1.32%)
Low-complexity repeats	137,547	8.29 (3.83%)
Unclassified repeats	269,189	46.67 (21.55%)
Total	655,713	124.06 (57.29%)

LINEs, long interspersed nuclear elements; LTRs, long terminal repeats.

https://doi.org/10.1371/journal.pone.0276287.t002

Table 3. Gene prediction statistics of the *Cardicola forsteri* genome assembly.

	Number	Total length (bp)	Average length (bp)
Genes	8,564	90,783,647 (41.93%)	10,600
CDSs	8,564	14,137,916 (6.53%)	1,650
Exons	57,052	14,747,359 (6.81%)	258
Introns	48,268	76,092,320 (35.14%)	1,576
5′ UTRs	203	405,498 (0.19%)	3
3′ UTRs	139	203,945 (0.09%)	3

CDS, coding sequence; UTR, untranslated region.

https://doi.org/10.1371/journal.pone.0276287.t003

proportion is higher than most other trematode genomes (40–54%), with the exception of *F. gigantica* (~ 70%) [57]. Long interspersed nuclear elements (LINEs) comprise > 20% of repetitive bases, with a smaller number of DNA transposons, simple repeats, and long terminal repeats (LTRs) identified in each assembly.

Genome annotation

The *C. forsteri* genome assembly was predicted to comprise a total of 8,564 protein-coding genes (covering 41.93% of the genome) and coding regions (CDSs), which cover 15.29% of the repeat-masked genome (Table 3). The vast majority (95.71%) of CDSs are complete, while the remaining are missing either a start or stop codon, or both. The average predicted gene and CDS length are 10,618 bp and 1,651 bp, respectively, with an average of 6.7 exons per gene. The predicted genes were translated into 8,564 hypothetical polypeptide sequences, 6,620 (77.30%) of which matched to the Swiss-Prot database, with 5,837 unique matches. The hypothetical polypeptide sequences (gene set) and their transcripts comprised 60.27% and 57.75% complete metazoan BUSCOs, respectively (Fig 2B). These results are mostly consistent with those of *Schistosoma* spp. assemblies, except that the average gene, intron, and exon length is smaller [58]. Furthermore, 70 putative proteases, 47 glycosyltransferases, 14 ribonucleases, 10 calpains, 6 cation channels, 6 tetraspanins (TSPs), 2 glutathione *S*-transferases (GSTs), 2 TGF-β homologs, and a fatty-acid binding protein (FABP) were functionally annotated, based on matches to protein sequences in the National Center for Biotechnology Information's (NCBI's) non-redundant protein sequences (NR) database (S2 Table). These protein families were selected as candidate immunogens and drug targets based on research into other digeneans. Matches were primarily to *S. japonicum*, *S. mansoni*, and *S. haematobium*, as well as some to *F. hepatica* and *Clonorchis sinensis*, which fits with known evolutionary relationships, as all are digeneans. Proteins of interest include calpains 1 and 2, secreted cathepsins B and L, and a CD63-like TSP with 4 predicted transmembrane domains (S2 Table).

Discussion

This is the first draft genome assembly of *C. forsteri* and the first of an aporocotylid. Most of the currently available genomes of trematodes belong to *Schistosoma* and *Fasciola* spp. as etiological agents of neglected tropical diseases (NTDs). *C. forsteri* is a significant pathogen of bluefin tuna and a member of the Schistosomatoidea but differs from its schistosome relatives in that it is monoecious and without suckers [11]. While schistosomes provide the closest point of genomic comparison, the species are expected to be sufficiently divergent to warrant a *de novo* genome assembly. Nanopore long reads were incorporated in this hybrid *de novo*

assembly, which aided in spanning the large repeat regions to greatly improve the contiguity and completeness of the hybrid assembly over short-read assemblies alone.

In silico methods for estimating genome size from *k*-mer depth varied greatly in their results, indicating that the *C. forsteri* genome is between 230 and 600 Mb. While *C. forsteri* was assumed to be diploid throughout these analyses based on the Smudgeplot analysis and the adult life cycle stage of schistosomes [59], this remains unknown, and polyploidy might have confounded these estimates. Ploidy varies among platyhelminths, sometimes even within species, so polyploidy would not be unusual for *C. forsteri* [60–62]. Additionally, the large number of repetitive elements typical of trematodes (> 55%), which is far higher than the estimates from GenomeScope (12–15%), may have interfered both with estimates of genome size and the assembly itself. Both genome size and ploidy would be more accurately estimated using laboratory techniques, such as flow cytometry coupled with fluorescence-activated cell sorting (FACS) [63, 64]. As the estimated heterozygosity for both specimens is relatively low (< 1%), this is unlikely to complicate their assemblies, so conventional assembly tools (e.g. ABySS) were used, rather than those developed for highly heterozygous genomes [65]. However, as with estimates of genome size, estimates of heterozygosity may not be accurate. Luo et al. [58] reported relatively high heterozygosity of *S. japonicum* (1.05%).

Although the hybrid contigs are highly contiguous due to Wengan's synthetic scaffolding approach [35], the contigs themselves are not scaffolded. However, improvements to the contiguity of the assembly could not be achieved by scaffolding with the available Nanopore sequence data using LongStitch [66], so future assemblies should incorporate mate-pair sequences, optical mapping, or Hi-C data [67]. The contiguity statistics of these hybrid contigs is comparable to scaffolds of early assemblies of other trematode genomes (S1 Table) [16], and the BUSCO analyses indicates a relatively complete and unfragmented assembly. In particular, the proportion of fragmented BUSCOs is low (< 10% for the predicted genome, transcriptome, and proteome) relative to other trematode assemblies.

Ab initio and homology-based gene prediction produced 8,564 protein-coding genes spanning > 40% of the assembled *C. forsteri* genome. As with other trematodes, the ratio of intron-to-exon length is large (> 6:1). While gene length is relatively consistent with other digeneans, the proportion of repeats is generally larger (>55%). Longer repeat regions are being identified in more recent assemblies due to the sequencing of higher molecular weight long reads. This phenomenon was evidenced by the most recent *S. mansoni* assembly, for which Buddenborg et al. [19] found an increase of 11% repetitive bases over the previous assembly. Luo et al. [57] showed that the increase in the proportion of repetitive elements in *Fasciola* spp. largely occurs within intergenic regions, so repetitive elements could likewise be contained within intergenic regions of the *C. forsteri* genome. Nevertheless, average CDS length remains relatively consistent among digeneans (~ 1.5 kb). The predicted mRNA transcripts are not supported by RNA-seq data, but the vast majority (> 77%) of polypeptide sequences generated from the predicted gene set match to other trematodes, which fits into an evolutionary context.

As with schistosomiases in humans, the drug of choice for treating bluefin tuna infected with *Cardicola* spp. is PZQ, which is thought to induce paralysis of the parasites by interfering with voltage-gated Ca^{2+} channels (VGCCs) in adult platyhelminths [68, 69]. In this study, 6 putative cation channels were functionally annotated from this *C. forsteri* genome assembly (S2 Table). Park et al. [70] showed that PZQ activated a schistosomal transient receptor potential melastatin ion channel (TRPM, A0A5K4F0X5) *in vitro*, but a single missense mutation (Asn^{1388}→Thr) in the hydrophobic binding pocket confers resistance to PZQ, which occurs in *F. hepatica*. Interestingly, the first 54 residues (including Asn^{1388}) of the binding pocket do not align with the most homologous putative CfTRPM1 (Table 4, sequence identity = 84.54%), but

Table 4. Select *Cardicola forsteri* protein homologs and their predicted conserved structural domains.

Putative protein	Abbreviation	Pfam-A
Calpain 1	CfCalp1	PF00648, PF01067
Calpain 2	CfCalp2	PF00648, PF01067
Cathepsin B	CfCB	PF00112, PF08127
Cathepsin L	CfCL	PF00112, PF08246
25-kDa GST	Cf25GST	PF14497, PF02798
27-kDa GST	Cf27GST	PF13417, PF14497
TRPM	CfTRPM1	PF18139, PF00293, PF00520
TSP-2 (CD63-like)	Cf-TSP-2	PF00335

GST, glutathione *S*-transferase; TRPM, transient receptor potential melastatin; TSP, tetraspanin.

https://doi.org/10.1371/journal.pone.0276287.t004

all the remaining residues identified as essential to PZQ-sensitivity are conserved. Therefore, we conclude either that PZQ interacts differently with *C. forsteri* TRPM or that PZQ kills *C. forsteri* via alternative mechanisms. This may indicate that the PZQ mechanism of action is more complex than interactions with a single target.

Most of the putative *C. forsteri* proteins are predicted to be N-glycosylated. Correspondingly, 47 putative glycosyltransferases were identified in this study (S2 Table). As expected, many of these are likely responsible for the synthesis of N-glycans, as indicated by structural domains in the CAZy database [52]. In its adult life cycle stage, *C. forsteri* predominately synthesizes oligomannose N-glycans, as well as paucimannosidic and complex-type N-glycans carrying core fucose and xylose [71]. No α1–3-linked fucoses were found in adults, but 3 putative α1–3 fucosyltransferases were identified in this study (Table 5). In particular, CfFucTD is homologous (sequence identity > 46%) to schistosomal α1–3 fucosyltransferases D (E2EAI7), E (E2EAI8), and F (E2EAI9), which have been functionally characterised and were shown to synthesize Lewis X and fucosylated LacdiNAc motifs [72, 73]. Therefore, *C. forsteri* likely synthesizes these motifs in pre-adult life cycle stages, as postulated by Coff et al. [71]. Sustained IgG1 and IgG3 responses to cercarial multifucosylated LDN motifs have been associated with protective immunity against schistosomiases [74–77], so glycomics is crucial to a holistic

Table 5. Select putative *Cardicola forsteri* glycosyltransferases with their predicted structural domains.

Putative protein	Abbreviation	Pfam-A	CAZy
Fucosyltransferases			
Peptide–O-fucosyltransferase	CfFucTA	PF10250	–
Galactoside α1–3 fucosyltransferase	CfFucTB	PF00852, PF17039	GT10
	CfFucTC	PF00852, PF17039	GT10
	CfFucTD	PF00852, PF17039	GT10
α1–6-fucosyltransferase	CfFucTE	PF19745	NA
Xylosyltransferases			
Xylosyltransferase I	CfXylT1	PF02485, PF12529	GT13
Glycoprotein β1–2 xylosyltransferase	CfXylT2	PF04577	GT61
***N*-acetylgalactosyltransferases**			
β1–4 *N*-acetylgalactosaminyltransferase	CfGalNT1	PF13733, PF02709	GT7
Glucuronyltransferases			
Bifunctional β1–3 glucuronyltransferase	CfGlcAT1	PF03360	GT43
	CfGlcAT2	PF13896	GT49

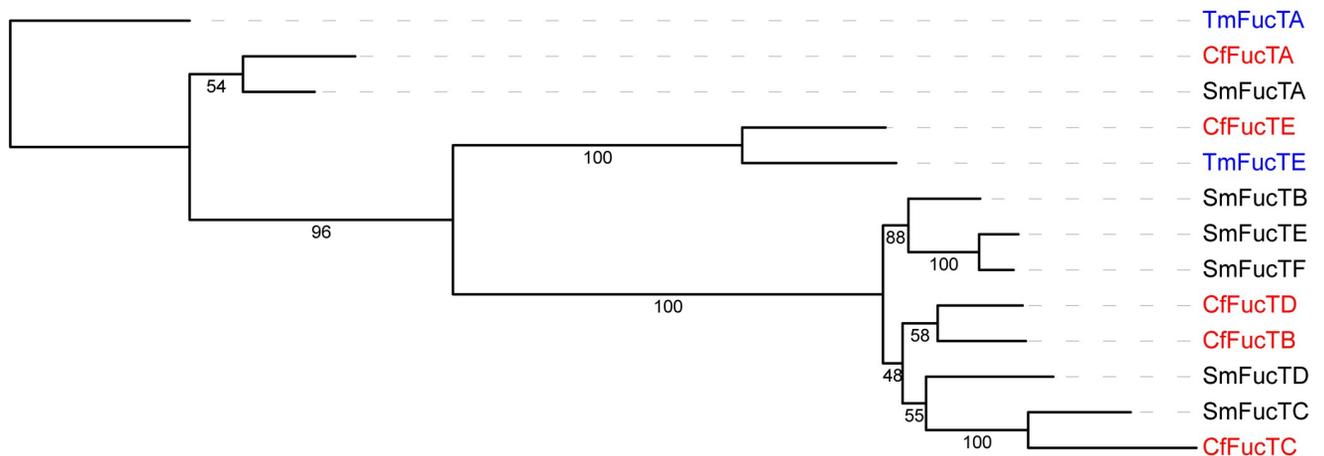https://doi.org/10.1371/journal.pone.0276287.t005

Fig 3. Phylogenetic tree of *Cardicola forsteri* fucosyltransferases. Outgroup-rooted maximum likelihood tree of the fucosyltransferases from *Cardicola forsteri* (shown in red) in Table 5 with fucosyltransferases from *Schistosoma mansoni* and *Thunnus maccoyii* (as the outgroup in blue): SmFucTA (G9HW08), SmFucTB (E2EAI5), SmFucTC (E2EAI6), SmFucTD (E2EAI7), SmFucTE (E2EAI8), SmFucTF (E2EAI9), TmFucTA (XP_042266362.1), TmFucTE (XP_042246885.1). Bootstrap values are shown on the branches.

https://doi.org/10.1371/journal.pone.0276287.g003

understanding of host–parasite interactions. Interestingly, CfFucTE is grouped with a hypothetical α1–6 fucosyltransferase from SBT (XP_042246885.1) in the phylogenetic analysis (Fig 3), although this node shares only 34% sequence identity. Additionally, several putative O-linked glycosyltransferases were identified, indicating that, like its schistosome relatives, *C. forsteri* may also synthesize O-glycans in its pre-adult life cycle stages [78, 79]. For example, CfFucTA shares homology (sequence identity = 51.51%) with *S. mansoni* protein–O-fucosyltransferase A (G9HW08), although these two proteins are grouped with low bootstrap support (54%). Two putative glucuronyltransferases were also identified, which could be involved in the synthesis of hexuronic acid-carrying N-glycans found in adult *C. forsteri*. However, glucuronyltransferases are not well characterised, so these require further functional studies. Unusually for a trematode, adult *C. forsteri* synthesizes core-xylose carrying N-glycans, and a putative β1–2 xylosyltransferase (CfXylT2) was also identified. Both β1–2-linked xyloses and glycosyltransferases elicit a humoral immune response in mammals following infection with *S. mansoni* [77, 80].

Cathepsins are a family of proteases that are important virulence factors in trematodes, facilitating tissues invasion, digestion, and immune evasion [81]. Cathepsins B and L (cysteine proteases) are secreted in high concentrations and often elicit humoral immune responses. These cathepsins, particularly *F. hepatica* cathepsin L1 (Q7JNQ9), have been the target of vaccination of ruminants against fascioliases [82]. A putative *C. forsteri* protease was classified as CfCL (Table 4) from sequence identity (44.68%) and a conserved active site (Cys[134], His[281], and Asn[301]), which clusters with cathepsin L1 of schistosomes (Fig 4A). Buffoni et al. [83] postulated that 42 residues within FhCL1 are responsible for protective immunity in vaccinated sheep, however these are not well conserved in CfCL, which contains substantial non-conservative substitutions. The prime schistosomal cathepsin targeted in vaccine studies is the Sm31 antigen *S. mansoni* cathepsin B1 (P25792), which is an abundant digestive enzyme secreted into the gut [84–88]. However, CfCB shares closest homology (sequence identity = 69.65%) and a phylogenetic node (Fig 4B) with SmCB2 (Q95PM1) [89], which was reactive to IgG from mice both infected or vaccinated with *S. mansoni* cercariae [80]. Additionally, *S. japonicum* cathepsin B2 (Q7Z1I6) is thought to be involved in skin penetration by the cercariae and degradation of host immune proteins [90]. As essential secreted gut antigens, cathepsins are
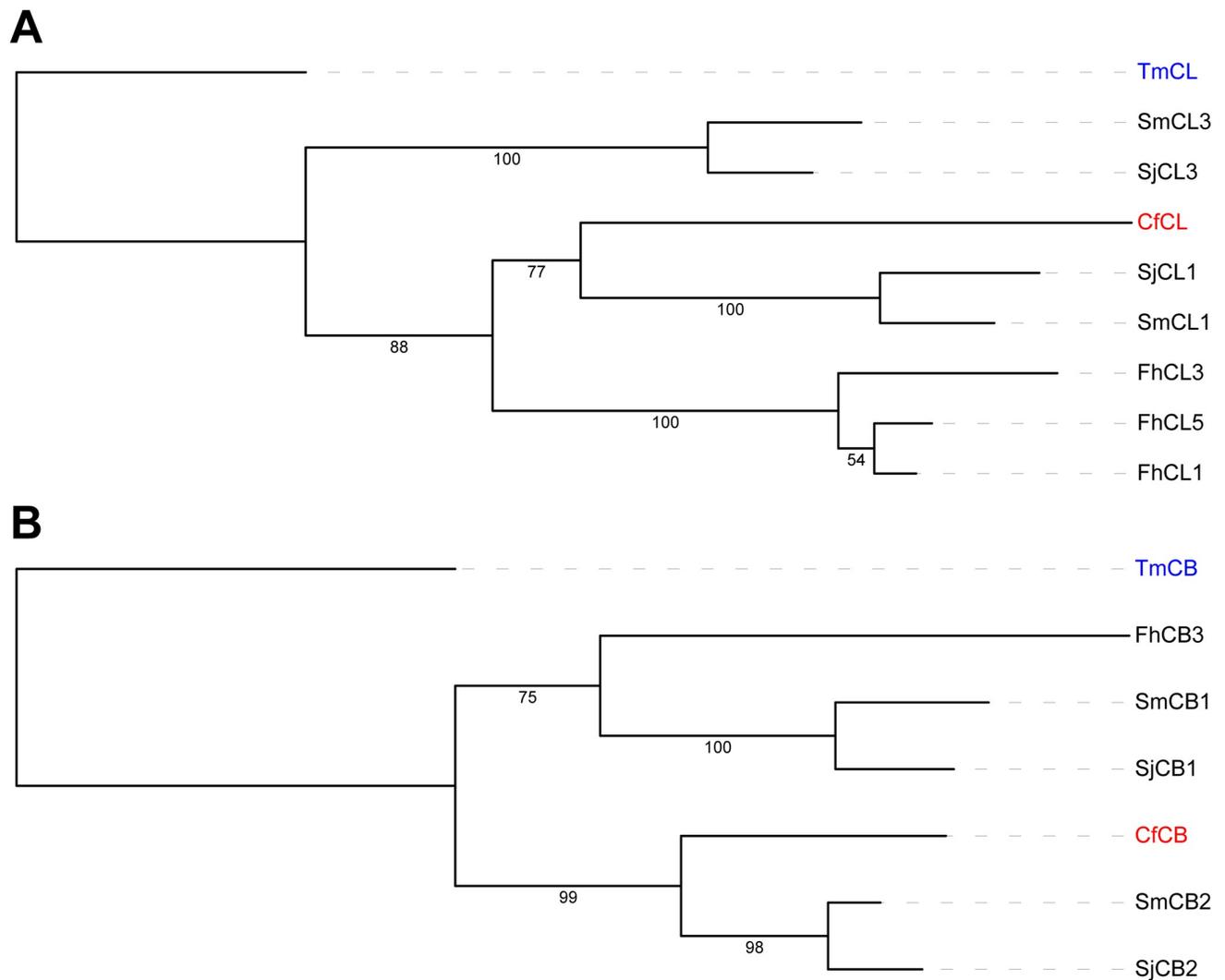
Fig 4. Phylogenetic tree of *Cardicola forsteri* cathepsins. Outgroup-rooted maximum likelihood tree of the cathepsins from *Cardicola forsteri* (shown in red) in Table 4 with cathepsins from *Schistosoma mansoni*, *Schistosoma japonicum*, *Fasciola hepatica*, and *Thunnus maccoyii* (as the outgroup in blue). Bootstrap values are shown on the branches. (A) Cathepsins L, including: SmCL1 (Q26564), SmCL3 (B4XC67), SjCL1 (C1LBR4), SjCL3 (C1LJC3), FhCL1 (Q7JNQ9), FhCL3 (B9VXS1), FhCL5 (A0A2H4PKA2), TmCL (XP_042278099.1). (B) Cathepsins B, including: SmCB1 (P25792), SmCB2 (Q95PM1), SjCB1 (P43157), SjCB2 (Q7Z1I6), FhB3 (A7UNB2), TmCB (XP_042261887.1).

https://doi.org/10.1371/journal.pone.0276287.g004

promising drug and vaccine targets. However, Farias et al. [80] found few epitopes within the active centre of schistosomal cathepsins, so direct inhibition is likely limited, which could explain the relative lack of success with cathepsin-based vaccines against schistosomiases compared with those against fascioliases. While immunogenicity is reportedly highest against gut antigens, tegumental proteins are also at the host–parasite interface and have been the subject of helminth vaccinology for decades.

Highly immunoreactive, calpains are Ca^{2+}-dependent cysteine proteases that participate in a variety of biological functions. Wang et al. [91] showed that two calpains, SmCalp1 (P27730.1) and SmCalp2 (ATN96084.1), are expressed on the tegument of adult *S. mansoni* and schistosomula, where they cleave host fibronectin, thereby preventing blood clotting. Two putative *C. forsteri* calpains were classified as CfCalp1 and CfCalp2 (Table 4), based on conserved structural domains and sequence identity (54.19% and 79.94%, respectively). Among

platyhelminths, amino acid sequences of calpains diverge greatly, however a Cys, His, and Asn residue are conserved in the active site of calpain 1. The His residue is replaced with Gln in calpain 2, and all these residues are conserved in CfCalp1 (Cys^{151}, His^{310}, and Asn^{334}) and CfCalp2 (Cys^{116}, Gln^{282}, and Asn^{310}). Since 1997, SmCalp1 (the Smp80 antigen) has been a target of vaccination against schistosomiasis [92], and promising results from pre-clinical trials have been published more recently [93, 94]. Other putative tegumental proteins of *C. forsteri* that could be viable vaccine candidates include GSTs, TSPs, and an FABP.

Similar to *S. mansoni* and *S. japonicum*, putative 25-kDa and 27-kDa GSTs were identified from the *C. forsteri* genome (Table 4), with homology to Class μ and Class ω GSTs, respectively [95]. As crucial detoxification enzymes, GSTs are found on the tegument of schistosomes and are frequently targeted by novel vaccines and drugs. The most advanced of these is Sh28GST (P30114), which is the target of the Bilhvax vaccine against urinary schistosomiasis [96, 97]. However, research into rSh28GST has stalled following Phase III trial results that reported a lack of efficacy in children [98]. While neither of the two putative *C. forsteri* GSTs shares close homology with Sh28GST (sequence identity < 27%), Tyr^7 and Arg^{18} are conserved in Cf25GST, and Tyr^{10} and Arg^{21} of Sh28GST were identified by Angelucci et al. [99] to form π-cation interactions with one another. Another tegumental target proceeding through vaccine trials is SmTSP-2 (Q8ITD7), a CD63-like tetraspanin that is thought to be involved in tegument formation and extracellular vesicle secretion [100, 101]. Of the 6 putative *C. forsteri* tetraspanins (S2 Table), Cf-TSP-2 was predicted to be CD63-like, with low homology to SmTSP-2 (sequence identity = 33.18%). Additionally, a putative 15-kDa FABP was identified, with homology (sequence identity = 56.49%) to the Sm14 FABP antigen (P29498), which is currently the target of Phase II clinical trials [102, 103].

Protective immunity to helminthiases is inhibited by immunomodulatory mechanisms deployed by the parasite, which polarise an anti-inflammatory T_H2 response and induce T cell anergy. These host–parasite interactions are not fully understood but involve parasite glycoconjugates interacting with host C-type lectin receptors (CLRs) and toll-like receptors (TLRs) [104]. The best described of these interactions is of schistosomal soluble egg antigens (SEA), namely the glycosylated T2 ribonuclease ω-1, interacting with mannose receptors (MRs) on monocyte-derived dendritic cells (Mo-DCs) [105]. No homologs of ω-1 were found in the *C. forsteri* genome, but two proteins contain a TGF-β-like domain (S2 Table), and TGF-β homologs are known to be expressed by other helminths to modulate host immunity [106, 107]. Nevertheless, potential mechanisms of evasion and modulation of the bluefin tuna immune responses by *C. forsteri* require further investigation.

The whole genome of *C. forsteri* was sequenced and assembled for the first time, into high-quality contigs. The highly repetitive genome was functionally annotated with putative glycosyltransferases and some potential vaccine targets, including calpains and an FABP. Future research should be directed toward functionally characterising these putative proteins using *in vitro* assays to further understand this parasite and its interactions with bluefin tuna. Additionally, this genome will provide a framework for genomic investigations into other *Cardicola* spp. to advance our understanding of their susceptibility to PZQ and the development of alternative control measures.

Supporting information

S1 Table. *Cardicola forsteri* assembly statistics. Comparison between the assembly statistics of *Cardicola forsteri* and related digeneans. Scaffold statistics are not applicable to the *C. forsteri* assembly, which is contig-level. Benchmarking Universal Single-Copy Orthologs (BUSCOs) were assessed using BUSCO v5.2.2 [37] against the metazoa_odb10 database. Both the

first and most recent assemblies of the related digeneans are included.
(XLSX)

S2 Table. Functional annotation of the *Cardicola forsteri* gene set. Functional annotation of select *Cardicola forsteri* genes. The closest matches to the National Center for Biotechnology Information's (NCBI's) non-redundant protein sequences (NR), Pfam v35.0 [51], and CAZy [52] databases are included. Predicted N-glycosylations sites, transmembrane domains, and secretion modes are also included, where relevant. Abbreviations are given, where relevant to genes discussed in the manuscript.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Lachlan Coff, Barbara F. Nowak, Paul A. Ramsland, Nathan J. Bott.

Data curation: Lachlan Coff.

Formal analysis: Lachlan Coff, Andrew J. Guy, Bronwyn E. Campbell.

Funding acquisition: Nathan J. Bott.

Investigation: Lachlan Coff.

Methodology: Lachlan Coff, Andrew J. Guy, Bronwyn E. Campbell, Barbara F. Nowak, Paul A. Ramsland.

Project administration: Nathan J. Bott.

Supervision: Paul A. Ramsland, Nathan J. Bott.

Validation: Paul A. Ramsland.

Visualization: Barbara F. Nowak, Paul A. Ramsland, Nathan J. Bott.

Writing – original draft: Lachlan Coff.

Writing – review & editing: Andrew J. Guy, Bronwyn E. Campbell, Barbara F. Nowak, Paul A. Ramsland, Nathan J. Bott.

References

1. Cribb T, Daintith M, Munday B. A new blood-fluke, *Cardicola forsteri*, (Digenea: Sanguinicolidae) of southern blue-fin tuna (*Thunnus maccoyii*) in aquaculture. T Roy Soc South Aust. 2000; 124: 117–120.

2. Bullard S, Goldstein R, Goodwin R, Overstreet R. *Cardicola forsteri* (Digenea: Sanguinicolidae) from the Heart of a Northern Bluefin Tuna, *Thunnus thynnus* (Scombridae), in the Northwest Atlantic Ocean. Comp Parasitol. 2004; 71: 245–246. https://doi.org/10.1654/4135

3. Shirakashi S, Tsunemoto K, Webber C, Rough K, Ellis D, Ogawa K. Two Species of *Cardicola* (Trematoda: Aporocotylidae) Found in Southern Bluefin Tuna *Thunnus maccoyii* Ranched in South Australia. Fish Pathol. 2013; 48: 1–4. https://doi.org/10.3147/jsfp.48.1

4. Ogawa K, Tanaka S, Sugihara Y, Takami I. A new blood fluke of the genus *Cardicola* (Trematoda: Sanguinicolidae) from Pacific bluefin tuna *Thunnus orientalis* (Temminck & Schlegel, 1844) cultured in Japan. Parasitol Int. 2010; 59: 44–48. https://doi.org/10.1016/j.parint.2009.10.003 PMID: 19835980

5. Ogawa K, Ishimaru K, Shirakashi S, Takami I, Grabner D. *Cardicola* opisthorchis n. sp. (Trematoda: Aporocotylidae) from the Pacific bluefin tuna, *Thunnus orientalis* (Temminck & Schlegel, 1844), cultured in Japan. Parasitol Int. 2011; 60: 307–312. https://doi.org/10.1016/j.parint.2011.05.002 PMID: 21616163

6. Power C, Nowak BF, Cribb TH, Bott NJ. Bloody flukes: a review of aporocotylids as parasites of cultured marine fishes. Int J Parasitol. 2020; 50: 743–753. https://doi.org/10.1016/j.ijpara.2020.04.008 PMID: 32619429

7. Hardy-Smith P, Ellis D, Humphrey J, Evans M, Evans D, Rough K, et al. *In vitro* and *in vivo* efficacy of anthelmintic compounds against blood fluke (*Cardicola forsteri*). Aquaculture. 2012; 334–337: 39–44. https://doi.org/10.1016/j.aquaculture.2011.12.037

8. Norbury L, Shirakashi S, Power C, Nowak BF, Bott NJ. Praziquantel use in aquaculture–Current status and emerging issues. Int J Parasitol. 2022; 18: 87–102. https://doi.org/10.1016/j.ijpddr.2022.02.001 PMID: 35220160

9. Pennacchi Y, Shirakashi S, Nowak BF, Bridle AR. Immune reactivity in early life stages of sea-cage cultured Pacific bluefin tuna naturally infected with blood flukes from genus *Cardicola* (Trematoda: Aporocotylidae). Fish Shellfish Immunol. 2016; 58: 490–499. https://doi.org/10.1016/j.fsi.2016.09.060 PMID: 27702677

10. Kirchhoff NT, Rough KM, Nowak BF. Moving Cages Further Offshore: Effects on Southern Bluefin Tuna, *T. maccoyii*, Parasites, Health and Performance. PLOS ONE. 2011; 6: e23705. https://doi.org/10.1371/journal.pone.0023705 PMID: 21901129

11. Cribb TH, Adlard RD, Hayward CJ, Bott NJ, Ellis D, Evans D, et al. The life cycle of *Cardicola forsteri* (Trematoda: Aporocotylidae), a pathogen of ranched southern bluefin tuna, *Thunnus maccoyi*. Int J Parasitol. 2011; 41: 861–870. https://doi.org/10.1016/j.ijpara.2011.03.011 PMID: 21569775

12. Shirakashi S, Tani K, Ishimaru K, Shin SP, Honryo T, Uchida H, et al. Discovery of intermediate hosts for two species of blood flukes *Cardicola orientalis* and *Cardicola forsteri* (Trematoda: Aporocotylidae) infecting Pacific bluefin tuna in Japan. Parasitol Int. 2016; 65: 128–136. https://doi.org/10.1016/j.parint.2015.11.003 PMID: 26571413

13. Shirakashi S, Matsuda T, Asai N, Honryo T, Ogawa K. *In vivo* cultivation of tuna blood fluke *Cardicola orientalis* in terebellid intermediate hosts. Int J Parasitol. 2020; 50: 851–857. https://doi.org/10.1016/j.ijpara.2020.04.006 PMID: 32592808

14. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of the blood fluke *Schistosoma mansoni*. Nature. 2009; 460: 352–358. https://doi.org/10.1038/nature08160 PMID: 19606141

15. Zhou Y, Zheng H, Chen Y, Zhang L, Wang K, Guo J, et al. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. Nature. 2009; 460: 345–351. https://doi.org/10.1038/nature08140 PMID: 19606140

16. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, et al. Whole-genome sequence of *Schistosoma haematobium*. Nat Genet. 2012; 44: 221–225. https://doi.org/10.1038/ng.1065 PMID: 22246508

17. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, et al. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. Genome Biol. 2015; 16: 71. https://doi.org/10.1186/s13059-015-0632-2 PMID: 25887684

18. Pandey T, Ghosh A, Todur VN, Rajendran V, Kalita P, Kalita J, et al. Draft Genome of the Liver Fluke *Fasciola gigantica*. ACS Omega. 2020; 5: 11084–11091. https://doi.org/10.1021/acsomega.0c00980 PMID: 32455229

19. Buddenborg SK, Tracey A, Berger DJ, Lu Z, Doyle SR, Fu B, et al. Assembled chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the evolution of its ZW sex-determination system. bioRxiv; 2021. p. 2021.08.13.456314. https://doi.org/10.1101/2021.08.13.456314

20. Short RB. A New Blood Fluke, *Cardicola laruei* n. g., n. sp., (Aporocotylidae) from Marine Fishes. J Parasitol. 1953; 39: 304–309. https://doi.org/10.2307/3273955

21. Yong RQ-Y, Cribb TH, Cutmore SC. Molecular phylogenetic analysis of the problematic genus *Cardicola* (Digenea: Aporocotylidae) indicates massive polyphyly, dramatic morphological radiation and host-switching. Mol Phylogenet Evol. 2021; 164: 107290. https://doi.org/10.1016/j.ympev.2021.107290 PMID: 34371186

22. Adams A. Progress, challenges and opportunities in fish vaccine development. Fish Shellfish Immunol. 2019; 90: 210–214. https://doi.org/10.1016/j.fsi.2019.04.066 PMID: 31039441

23. Perera DJ, Ndao M. Promising Technologies in the Field of Helminth Vaccines. Frontiers in Immunology. 2021; 12. https://www.frontiersin.org/articles/10.3389/fimmu.2021.711650 PMID: 34489961

24. Aiken HM, Hayward CJ, Nowak BF. An epizootic and its decline of a blood fluke, *Cardicola forsteri*, in farmed southern bluefin tuna, *Thunnus maccoyii*. Aquaculture. 2006; 254: 40–45. https://doi.org/10.1016/j.aquaculture.2005.10.013

25. Power C, Evenden S, Rough K, Webber C, Widdicombe M, Nowak BF, et al. Prevalence and Intensity of *Cardicola* spp. Infection in Ranched Southern Bluefin Tuna and a Comparison of Diagnostic Methods. Pathogens. 2021; 10: 1248. https://doi.org/10.3390/pathogens10101248 PMID: 34684197

26. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019; 20: 129. https://doi.org/10.1186/s13059-019-1727-y PMID: 31234903

27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

28. Wang M, Kong L. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinform. 2019; 20: 28. https://doi.org/10.1186/s12859-019-2597-8 PMID: 30646844

29. Suda A, Nishiki I, Iwasaki Y, Matsuura A, Akita T, Suzuki N, et al. Improvement of the Pacific bluefin tuna (*Thunnus orientalis*) reference genome and development of male-specific DNA markers. Sci Rep. 2019; 9: 14450. https://doi.org/10.1038/s41598-019-50978-4 PMID: 31595011

30. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 2017; 27: 768–777. https://doi.org/10.1101/gr.214346.116 PMID: 28232478

31. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics. 2011; 27: 764–770. https://doi.org/10.1093/bioinformatics/btr011 PMID: 21217122

32. Sohn J, Nam J-W. The present and future of *de novo* whole-genome assembly. Brief Bioinform. 2018; 19: 23–40. https://doi.org/10.1093/bib/bbw096 PMID: 27742661

33. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017; 33: 2202–2204. https://doi.org/10.1093/bioinformatics/btx153 PMID: 28369201

34. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020; 11: 1432. https://doi.org/10.1038/s41467-020-14998-3 PMID: 32188846

35. Di Genova A, Buena-Atienza E, Ossowski S, Sagot M-F. Efficient hybrid *de novo* assembly of human genomes with WENGAN. Nat Biotechnol. 2021; 39: 422–430. https://doi.org/10.1038/s41587-020-00747-w PMID: 33318652

36. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018; 34: i142–i150. https://doi.org/10.1093/bioinformatics/bty266 PMID: 29949969

37. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution. 2021; 38: 4647–4654. https://doi.org/10.1093/molbev/msab199 PMID: 34320186

38. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020; 117: 9451–9457. https://doi.org/10.1073/pnas.1921046117 PMID: 32300014

39. Bao Z, Eddy SR. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res. 2002; 12: 1269–1276. https://doi.org/10.1101/gr.88502 PMID: 12176934

40. Price AL, Jones NC, Pevzner PA. *De novo* identification of repeat families in large genomes. Bioinformatics. 2005; 21: i351–i358. https://doi.org/10.1093/bioinformatics/bti1018 PMID: 15961478

41. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinform. 2008; 9: 18. https://doi.org/10.1186/1471-2105-9-18 PMID: 18194517

42. Ou S, Jiang N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. 2018; 176: 1410–1422. https://doi.org/10.1104/pp.17.01310 PMID: 29233850

43. Smit A, Hubley R, Green P. RepeatMasker. 2022 [cited 25 Sep 2022]. http://repeatmasker.org

44. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021; 3: lqaa108. https://doi.org/10.1093/nargab/lqaa108 PMID: 33575650

45. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005; 33: 6494–6506. https://doi.org/10.1093/nar/gki937 PMID: 16314312

46. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021; 18: 366–368. https://doi.org/10.1038/s41592-021-01101-x PMID: 33828273

47. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2020; 2: lqaa026. https://doi.org/10.1093/nargab/lqaa026 PMID: 32440658

48. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics. 2008; 24: 637–644. https://doi.org/10.1093/bioinformatics/btn013 PMID: 18218656

49. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021; 49: D480–D489. https://doi.org/10.1093/nar/gkaa1100 PMID: 33237286

50. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39: W29–W37. https://doi.org/10.1093/nar/gkr367 PMID: 21593126

51. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021; 49: D412–D419. https://doi.org/10.1093/nar/gkaa913 PMID: 33125078

52. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022; 50: D571–D577. https://doi.org/10.1093/nar/gkab1045 PMID: 34850161

53. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022; 40: 1023–1025. https://doi.org/10.1038/s41587-021-01156-3 PMID: 34980915

54. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002; 310–322. PMID: 11928486

55. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792–1797. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

56. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

57. Luo X, Cui K, Wang Z, Li Z, Wu Z, Huang W, et al. High-quality reference genome of *Fasciola gigantica*: Insights into the genomic signatures of transposon-mediated evolution and specific parasitic adaption in tropical regions. PLOS Negl Trop Dis. 2021; 15: e0009750. https://doi.org/10.1371/journal.pntd.0009750 PMID: 34610021

58. Luo F, Yin M, Mo X, Sun C, Wu Q, Zhu B, et al. An improved genome assembly of the fluke *Schistosoma japonicum*. PLOS Negl Trop Dis. 2019; 13: e0007612. https://doi.org/10.1371/journal.pntd.0007612 PMID: 31390359

59. Xu X, Sun J, Zhang J, Wellems D, Qing X, McCutchan T, et al. Having a pair: the key to immune evasion for the diploid pathogen *Schistosoma japonicum*. Sci Rep. 2012; 2: 346. https://doi.org/10.1038/srep00346 PMID: 22468230

60. D'Souza TG, Storhas M, Michiels NK. The effect of ploidy level on fitness in parthenogenetic flatworms. Biol J Linn Soc. 2005; 85: 191–198. https://doi.org/10.1111/j.1095-8312.2005.00482.x

61. Hirai H, LoVerde PT. Triploid Cells Found in Intramolluscan Stages of *Schistosoma mansoni*. J Parasitol. 1989; 75: 800–802. https://doi.org/10.2307/3283067

62. Itagaki T, Sakaguchi K, Terasaki K, Sasaki O, Yoshihara S, Van Dung T. Occurrence of spermic diploid and aspermic triploid forms of *Fasciola* in Vietnam and their molecular characterization based on nuclear and mitochondrial DNA. Parasitol Int. 2009; 58: 81–85. https://doi.org/10.1016/j.parint.2008.11.003 PMID: 19087891

63. Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. Measuring Genome Sizes Using Read-Depth, k-mers, and Flow Cytometry: Methodological Comparisons in Beetles (Coleoptera). G3: Genes Genomes Genet. 2020; 10: 3047–3060. https://doi.org/10.1534/g3.120.401028 PMID: 32601059

64. Garavello M, Cuenca J, Dreissig S, Fuchs J, Houben A, Aleza P. Assessing Ploidy Level Analysis and Single Pollen Genotyping of Diploid and Euploid Citrus Genotypes by Fluorescence-Activated Cell Sorting and Whole-Genome Amplification. Front Plant Sci. 2019; 10. https://doi.org/10.3389/fpls.2019.01174 PMID: 31611896

65. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014; 24: 1384–1395. https://doi.org/10.1101/gr.170720.113 PMID: 24755901

66. Coombe L, Li JX, Lo T, Wong J, Nikolic V, Warren RL, et al. LongStitch: high-quality genome assembly correction and scaffolding using long reads. BMC Bioinform. 2021; 22: 534. https://doi.org/10.1186/s12859-021-04451-7 PMID: 34717540

67. Yamaguchi K, Kadota M, Nishimura O, Ohishi Y, Naito Y, Kuraku S. Technical considerations in Hi-C scaffolding and evaluation of chromosome-scale genome assemblies. Mol Ecol. 2021; 30: 5923–5934. https://doi.org/10.1111/mec.16146 PMID: 34432923

68. Vale N, Gouveia MJ, Rinaldi G, Brindley PJ, Gärtner F, Correia da Costa JM. Praziquantel for Schisto-somiasis: Single-Drug Metabolism Revisited, Mode of Action, and Resistance. Antimicrob Agents Chemother. 2017; 61: e02582–16. https://doi.org/10.1128/AAC.02582-16 PMID: 28264841

69. Thomas CM, Timson DJ. The Mechanism of Action of Praziquantel: Can New Drugs Exploit Similar Mechanisms? Curr Med Chem. 2020; 27: 676–696. https://doi.org/10.2174/0929867325666180926145537 PMID: 30259811

70. Park S-K, Friedrich L, Yahya NA, Rohr CM, Chulkov EG, Maillard D, et al. Mechanism of praziquantel action at a parasitic flatworm ion channel. Sci Transl Med. 2021; 13: eabj5832. https://doi.org/10.1126/scitranslmed.abj5832 PMID: 34936384

71. Coff L, Abrahams JL, Collett S, Power C, Nowak BF, Kolarich D, et al. Profiling the glycome of *Cardicola forsteri*, a blood fluke parasitic to bluefin tuna. Int J Parasitol. 2022; 52: 1–12. https://doi.org/10.1016/j.ijpara.2021.06.004 PMID: 34391752

72. Mickum ML, Rojsajjakul T, Yu Y, Cummings RD. *Schistosoma mansoni* α1,3-fucosyltransferase-F generates the Lewis X antigen. Glycobiology. 2016; 26: 270–285. https://doi.org/10.1093/glycob/cwv103 PMID: 26582608

73. van Noort K, Nguyen D-L, Kriechbaumer V, Hawes C, Hokke CH, Schots A, et al. Functional charac-terization of *Schistosoma mansoni* fucosyltransferases in *Nicotiana benthamiana* plants. Sci Rep. 2020; 10: 18528. https://doi.org/10.1038/s41598-020-74485-z PMID: 33116178

74. Mickum ML, Prasanphanich NS, Song X, Dorabawila N, Mandalasi M, Lasanajak Y, et al. Identification of Antigenic Glycans from *Schistosoma mansoni* by Using a Shotgun Egg Glycan Microarray. Infect Immun. 2016; 84: 1371–1386. https://doi.org/10.1128/IAI.01349-15 PMID: 26883596

75. Yang YYM, Li XH, Brzezicka K, Reichardt N-C, Wilson RA, van Diepen A, et al. Specific anti-glycan antibodies are sustained during and after parasite clearance in *Schistosoma japonicum*-infected rhe-sus macaques. PLOS Negl Trop Dis. 2017; 11: e0005339. https://doi.org/10.1371/journal.pntd.0005339 PMID: 28151933

76. Yang YYM, van Diepen A, Brzezicka K, Reichardt N-C, Hokke CH. Glycan Microarray-Assisted Identi-fication of IgG Subclass Targets in Schistosomiasis. Front Immunol. 2018; 9. https://doi.org/10.3389/fimmu.2018.02331 PMID: 30356796

77. Nkurunungi G, van Diepen A, Nassuuna J, Sanya RE, Nampijja M, Nambuya I, et al. Microarray assessment of N-glycan-specific IgE and IgG profiles associated with *Schistosoma mansoni* infection in rural and urban Uganda. Sci Rep. 2019; 9: 3522. https://doi.org/10.1038/s41598-019-40009-7 PMID: 30837526

78. Smit CH, van Diepen A, Nguyen DL, Wuhrer M, Hoffmann KF, Deelder AM, et al. Glycomic Analysis of Life Stages of the Human Parasite *Schistosoma mansoni* Reveals Developmental Expression Profiles of Functional and Antigenic Glycan Motifs. Mol Cell Proteomics. 2015; 14: 1750–1769. https://doi.org/10.1074/mcp.M115.048280 PMID: 25883177

79. Smit CH, Homann A, van Hensbergen VP, Schramm G, Haas H, van Diepen A, et al. Surface expres-sion patterns of defined glycan antigens change during *Schistosoma mansoni* cercarial transformation and development of schistosomula. Glycobiology. 2015; 25: 1465–1479. https://doi.org/10.1093/glycob/cwv066 PMID: 26347524

80. Farias LP, Vance GM, Coulson PS, Vitoriano-Souza J, Neto AP da S, Wangwiwatsin A, et al. Epitope Mapping of Exposed Tegument and Alimentary Tract Proteins Identifies Putative Antigenic Targets of the Attenuated Schistosome Vaccine. Front Immunol. 2021; 11. https://doi.org/10.3389/fimmu.2020.624613 PMID: 33763055

81. Cwiklinski K, Donnelly S, Drysdale O, Jewhurst H, Smith D, De Marco Verissimo C, et al. The cathep-sin-like cysteine peptidases of trematodes of the genus *Fasciola*. In: Rollinson D, Stothard JR, editors. Advances in Parasitology. Academic Press; 2019. pp. 113–164.

82. Cwiklinski K, Drysdale O, López Corrales J, Corripio-Miyar Y, De Marco Verissimo C, Jewhurst H, et al. Targeting Secreted Protease/Anti-Protease Balance as a Vaccine Strategy against the Helminth *Fasciola hepatica*. Vaccines. 2022; 10: 155. https://doi.org/10.3390/vaccines10020155 PMID: 35214614

83. Buffoni L, Garza-Cuartero L, Pérez-Caballero R, Zafra R, Javier Martínez-Moreno F, Molina-Hernán-dez V, et al. Identification of protective peptides of *Fasciola hepatica*-derived cathepsin L1 (FhCL1) in

vaccinated sheep by a linear B-cell epitope mapping approach. Parasites Vectors. 2020; 13: 390. https://doi.org/10.1186/s13071-020-04260-6 PMID: 32736582

84. Abdulla M-H, Lim K-C, Sajid M, McKerrow JH, Caffrey CR. Schistosomiasis Mansoni: Novel Chemo-therapy Using a Cysteine Protease Inhibitor. PLOS Med. 2007; 4: e14. https://doi.org/10.1371/journal.pmed.0040014 PMID: 17214506

85. Jílková A, Rubešová P, Fanfrlík J, Fajtová P, Řezáčová P, Brynda J, et al. Druggable Hot Spots in the Schistosomiasis Cathepsin B1 Target Identified by Functional and Binding Mode Analysis of Potent Vinyl Sulfone Inhibitors. ACS Infect Dis. 2021; 7: 1077–1088. https://doi.org/10.1021/acsinfecdis.0c00501 PMID: 33175511

86. Ricciardi A, Visitsunthorn K, Dalton JP, Ndao M. A vaccine consisting of *Schistosoma mansoni* cathepsin B formulated in Montanide ISA 720 VG induces high level protection against murine schistosomiasis. BMC Infect Dis. 2016; 16: 112. https://doi.org/10.1186/s12879-016-1444-z PMID: 26945988

87. Ricciardi A, Dalton JP, Ndao M. Evaluation of the immune response and protective efficacy of *Schistosoma mansoni* Cathepsin B in mice using CpG dinucleotides as adjuvant. Vaccine. 2015; 33: 346–353. https://doi.org/10.1016/j.vaccine.2014.11.016 PMID: 25448114

88. Perera DJ, Hassan AS, Jia Y, Ricciardi A, McCluskie MJ, Weeratna RD, et al. Adjuvanted *Schistosoma mansoni*-Cathepsin B With Sulfated Lactosyl Archaeol Archaeosomes or AddaVax™ Provides Protection in a Pre-Clinical Schistosomiasis Model. Front Immunol. 2020; 11. https://doi.org/10.3389/fimmu.2020.605288 PMID: 33304354

89. Caffrey CR, Salter JP, Lucas KD, Khiem D, Hsieh I, Lim K-C, et al. SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. Mol Biochem Parasitol. 2002; 121: 49–61. https://doi.org/10.1016/S0166-6851(02)00022-1

90. Zhu B, Luo F, Shen Y, Yang W, Sun C, Wang J, et al. *Schistosoma japonicum* cathepsin B2 (SjCB2) facilitates parasite invasion through the skin. PLOS Negl Trop Dis. 2020; 14: e0008810. https://doi.org/10.1371/journal.pntd.0008810 PMID: 33104723

91. Wang Q, Da'dara AA, Skelly PJ. The human blood parasite *Schistosoma mansoni* expresses extracellular tegumental calpains that cleave the blood clotting protein fibronectin. Sci Rep. 2017; 7: 12912. https://doi.org/10.1038/s41598-017-13141-5 PMID: 29018227

92. Hota-Mitchell S, Siddiqui AA, Dekaban GA, Smith J, Tognon C, Podesta RB. Protection against *Schistosoma mansoni* infection with a recombinant baculovirus-expressed subunit of calpain. Vaccine. 1997; 15: 1631–1640. https://doi.org/10.1016/S0264-410X(97)00081-9

93. Siddiqui AJ, Molehin AJ, Zhang W, Ganapathy PK, Kim E, Rojo JU, et al. Sm-p80-based vaccine trial in baboons: efficacy when mimicking natural conditions of chronic disease, praziquantel therapy, immunization, and *Schistosoma mansoni* re-encounter. Ann N Y Acad Sci. 2018; 1425: 19–37. https://doi.org/10.1111/nyas.13866 PMID: 29888790

94. Zhang W, Molehin AJ, Rojo JU, Sudduth J, Ganapathy PK, Kim E, et al. Sm-p80-based schistosomiasis vaccine: double-blind preclinical trial in baboons demonstrates comprehensive prophylactic and parasite transmission-blocking efficacy. Ann N Y Acad Sci. 2018; 1425: 38–51. https://doi.org/10.1111/nyas.13942 PMID: 30133707

95. Wright MD, Davern KM, Mitchell GF. The functional and immunological significance of some schistosome surface molecules. Parasitol Today. 1991; 7: 56–58. https://doi.org/10.1016/0169-4758(91)90191-p PMID: 15463423

96. Johnson KA, Angelucci F, Bellelli A, Hervé M, Fontaine J, Tsernoglou D, et al. Crystal Structure of the 28 kDa Glutathione *S*-Transferase from *Schistosoma haematobium*. Biochemistry. 2003; 42: 10084–10094. https://doi.org/10.1021/bi034449r PMID: 12939136

97. Riveau G, Deplanque D, Remoué F, Schacht A-M, Vodougnon H, Capron M, et al. Safety and Immunogenicity of rSh28GST Antigen in Humans: Phase 1 Randomized Clinical Study of a Vaccine Candidate against Urinary Schistosomiasis. PLOS Negl Trop Dis. 2012; 6: e1704. https://doi.org/10.1371/journal.pntd.0001704 PMID: 22802974

98. Riveau G, Schacht A-M, Dompnier J-P, Deplanque D, Seck M, Waucquier N, et al. Safety and efficacy of the rSh28GST urinary schistosomiasis vaccine: A phase 3 randomized, controlled trial in Senegalese children. PLOS Negl Trop Dis. 2018; 12: e0006968. https://doi.org/10.1371/journal.pntd.0006968 PMID: 30532268

99. Angelucci F, Baiocco P, Brunori M, Gourlay L, Morea V, Bellelli A. Insights into the Catalytic Mechanism of Glutathione S-Transferase: The Lesson from *Schistosoma haematobium*. Structure. 2005; 13: 1241–1246. https://doi.org/10.1016/j.str.2005.06.007 PMID: 16154081

100. Pinheiro CS, Ribeiro APD, Cardoso FC, Martins VP, Figueiredo BCP, Assis NRG, et al. A multivalent chimeric vaccine composed of *Schistosoma mansoni* SmTSP-2 and Sm29 was able to induce protection against infection in mice. Parasite Immunol. 2014; 36: 303–312. https://doi.org/10.1111/pim.12118 PMID: 24749785

101. Sotillo J, Pearson M, Potriquet J, Becker L, Pickering D, Mulvenna J, et al. Extracellular vesicles secreted by *Schistosoma mansoni* contain protein vaccine candidates. Int J Parasitol. 2016; 46: 1–5. https://doi.org/10.1016/j.ijpara.2015.09.002 PMID: 26460238

102. Tendler M, Almeida MS, Vilar MM, Pinto PM, Limaverde-Sousa G. Current Status of the Sm14/GLA-SE Schistosomiasis Vaccine: Overcoming Barriers and Paradigms towards the First Anti-Parasitic Human(itarian) Vaccine. Trop Med Infect Dis. 2018; 3: 121. https://doi.org/10.3390/tropicalmed3040121 PMID: 30469320

103. Santini-Oliveira M, Coler RN, Parra J, Veloso V, Jayashankar L, Pinto PM, et al. Schistosomiasis vaccine candidate Sm14/GLA-SE: Phase 1 safety and immunogenicity clinical trial in healthy, male adults. Vaccine. 2016; 34: 586–594. https://doi.org/10.1016/j.vaccine.2015.10.027 PMID: 26571311

104. Angeles JMaM, Mercado VJP, Rivera PT. Behind Enemy Lines: Immunomodulatory Armamentarium of the Schistosome Parasite. Front Immunol. 2020; 11. https://doi.org/10.3389/fimmu.2020.01018 PMID: 32582161

105. Everts B, Hussaarts L, Driessen NN, Meevissen MHJ, Schramm G, van der Ham AJ, et al. Schistosome-derived omega-1 drives Th2 polarization by suppressing protein synthesis following internalization by the mannose receptor. J Exp Med. 2012; 209: 1753–1767. https://doi.org/10.1084/jem.20111381 PMID: 22966004

106. Musah-Eroje M, Flynn RJ. *Fasciola hepatica*, TGF-β and host mimicry: the enemy within. Curr Opin Microbiol. 2018; 46: 80–85. https://doi.org/10.1016/j.mib.2018.09.002 PMID: 30317150

107. Johnston CJC, Smyth DJ, Kodali RB, White MPJ, Harcus Y, Filbey KJ, et al. A structurally distinct TGF-β mimic from an intestinal helminth parasite potently induces regulatory T cells. Nat Commun. 2017; 8: 1741. https://doi.org/10.1038/s41467-017-01886-6 PMID: 29170498