DATA NOTE

The genome sequence of the liver fluke *Opisthorchis viverrini* (Poirier, 1886) Stiles & Hassall, 1896

[version 1; peer review: 2 approved]

Arporn Wangwiwatsin[1,2], Siriyakorn Kulwong[1,2], Chitsakul Phuyao[2], Attapol Titapun[2,3], Watcharin Loilome[1,2], Poramate Klanrit[1,2], Nisana Namwat[1,2], Paiboon Sithithaworn[2,4], Stephen R. Doyle[5], Matthew Berriman[5,6], Thomas Crellen[5,7,8],
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective

[1]Department of Systems Biosciences and Computational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
[2]Cholangiocarcinoma Research Institute, Khon Kaen University, Khon Kaen, Thailand
[3]Department of Surgery, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
[4]Department of Parasitology, Khon Kaen University, Khon Kaen, Thailand
[5]Wellcome Sanger Institute, Hinxton, England, UK
[6]School of Infection & Immunity, University of Glasgow, Glasgow, Scotland, UK
[7]School of Biodiversity One Health and Veterinary Medicine, University of Glasgow, Glasgow, Scotland, UK
[8]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, England, UK

v1 **First published:** 10 Jan 2025, **10**:1
https://doi.org/10.12688/wellcomeopenres.23535.1
Latest published: 10 Jan 2025, **10**:1
https://doi.org/10.12688/wellcomeopenres.23535.1

Abstract

We present a genome assembly from a specimen of *Opisthorchis viverrini* (liver fluke; Platyhelminthes; Trematoda; Opisthorchiida; Opisthorchiidae). The genome sequence has a total length of 627.20 megabases. Most of the assembly (97.89%) is scaffolded into 6 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 18.04 kilobases in length.

Keywords

Opisthorchis viverrini, liver fluke, genome sequence, chromosomal, Opisthorchiida, neglected tropical diseases, foodborne trematodiasis

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 10 Jan 2025	✓ view	✓ view

1. **Jose Tort**, Universidad de la República, Montevideo, Uruguay

2. **Ferdinand Marlétaz**, University College London, London, UK

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Tree of Life gateway.

Corresponding author: Thomas Crellen (thomas.crellen@glasgow.ac.uk)

Author roles: Wangwiwatsin A: Conceptualization, Funding Acquisition, Project Administration, Resources, Writing – Original Draft Preparation; **Kulwong S**: Data Curation, Investigation; **Phuyao C**: Data Curation, Investigation, Resources; **Titapun A**: Funding Acquisition, Investigation, Resources; **Loilome W**: Funding Acquisition, Project Administration, Resources, Supervision; **Klanrit P**: Project Administration; **Namwat N**: Funding Acquisition, Resources; **Sithithaworn P**: Investigation; **Doyle SR**: Supervision; **Berriman M**: Project Administration, Supervision; **Crellen T**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] as well as through a Sir Henry Wellcome fellowship to T.C. (reference: 215919/B/19/Z), and through National Science Research and Innovation Fund (NSRF) under the Basic Research Fund of Khon Kaen University (CARI-BRF64-48) through Cholangiocarcinoma Research Institute allocated to A.W.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Wangwiwatsin A, Kulwong S, Phuyao C *et al.* **The genome sequence of the liver fluke *Opisthorchis viverrini* (Poirier, 1886) Stiles & Hassall, 1896 [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:1 https://doi.org/10.12688/wellcomeopenres.23535.1

First published: 10 Jan 2025, **10**:1 https://doi.org/10.12688/wellcomeopenres.23535.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Platyhelminthes; Trematoda; Digenea; Opisthorchiida; Opisthorchiata; Opisthorchiidae; *Opisthorchis*; *Opisthorchis viverrini* (Poirier, 1886) Stiles & Hassall, 1896 (NCBI:txid6198)

Background

Opisthorchis viverrini, commonly known as the Southeast Asian liver fluke, is a species of trematode with a complex parasitic lifecycle which includes molluscan and piscine intermediate hosts and a mammalian definitive host. Since 1994 the parasite has been classified as carcinogenic to humans by the International Agency for Research on Cancer (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 1994). This parasite is a direct cause of bile duct cancer (cholangiocarcinoma; CCA), which is supported by both epidemiological evidence, as there is geographical overlap between the distribution of *O. viverrini* infection and an elevated incidence of CCA, and empirical laboratory studies where infection with the parasite induces CCA in hamster models (Elkins *et al.*, 1990; Haswell-Elkins *et al.*, 1994; Loilome *et al.*, 2006; Ohshima *et al.*, 1994; Pinlaor *et al.*, 2004a, Pinlaor *et al.*, 2004b; Yongvanit *et al.*, 2012). The parasite is prevalent throughout Southeast Asia, where consumption of dishes containing raw freshwater fish is a cultural tradition, and it is estimated to infect 12 million people across Thailand, Cambodia, Lao PDR and Vietnam (Zhao *et al.*, 2021). However, *O. viverini* also poses risks to travellers who may contract the parasite, either through direct consumption or via contamination of surfaces or cooking utensils (Grundy-Warr *et al.*, 2012; Shin *et al.*, 2010). The morbidity resulting from infection with *O. viverrini* is considered a Neglected Tropical Disease under the category of 'Foodborne Trematodiasis' (Borlase *et al.*, 2023). The incidence of bile duct cancer in humans is notably higher in *O. viverrini*-endemic areas (Florio *et al.*, 2020), and is a major burden in terms of the cost to both healthcare systems and socioeconomically, as it typically afflicts the highest earning family members (Khuntikeo *et al.*, 2018).

Humans contract *O. viverrini* by consuming raw freshwater fish of the family Cyprinidae (carp), which are encysted with the infective stage of the parasite (metacercariae). Adult *O. viverrini* reside in the bile ducts of mammalian definitive hosts, where they can survive for over a decade (Crellen *et al.*, 2024). As a hermaphroditic species, each adult worm possesses both male and female reproductive organs. Upon reaching reproductive maturity, eggs are released into the intestine and subsequently expelled into the environment with faeces. In areas with poor sanitation, these eggs contaminate natural water sources, where they are ingested by *Bithynia* snails. Within the snail host, the eggs hatch and develop into larval stages (miracidia, redia, and cercariae) and multiply asexually. Cercariae exit the snails as free-swimming cercariae before seeking out a piscine host and penetrating the skin, then encysting into metacercariae within fish muscle. Metacercariae are then inadvertently consumed by mammalian definitive hosts including humans and, potentially, domestic cats and dogs (Harinasuta & Harinasuta, 1984; Sota *et al.*, 2022). In cases where the tumour is resectable, surgeons in Thailand have encountered multiple instances where adult worms were present inside the bile ducts of cancer patients, and they were removed along with the tumour and adjacent margin areas. Such cases have become less common in Thailand due to parasite control programs (Sithithaworn *et al.*, 2014). However, they may still occur in other endemic areas lacking adequate public health measures.

Here, we present a chromosome-level reference genome for *O. viverrini* containing six autosomal chromosomes and a complete mitochondrial genome. We used long-read sequencing and Hi-C chromatin conformation on two adult worms obtained from human patients in Thailand. The *O. viverrini* worms were recovered from the bile ducts of cancer patients undergoing curative surgery during the course of treatment, by a team of Hepato-Pancreato-Biliary surgeons led by Prof. Narong Khuntikeo at Srinakarind Hospital in Thailand. The specimens used here were held by, and accessed from, the biobank of the Cholangiocarcinoma Research Institute at Khon Kaen University.

The availability of this new genome data represents a significant improvement over a previous draft genome (Young *et al.*, 2014), which relied solely on short-read sequencing and used pooled DNA from multiple adult worms which were collected as metacercariae from fish, rather than human patients. The circularised mitochondrial genome also replaces a previous incomplete version (Cai *et al.*, 2012). This, greatly improved, reference for *O. viverrini* will provide an essential foundation for functional genomics aimed at developing diagnostics and therapeutics against the parasite, as well as enabling genomic epidemiology to inform evidence-based disease control interventions across Southeast Asia (Crellen *et al.*, 2021).

Genome sequence report

The genome of *Opisthorchis viverrini* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 22.18 Gb (gigabases) from 2.90 million reads, providing an estimated 32-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 127.26 Gb from 842.81 million reads. Specimen and sequencing details are summarised in Table 1.

Assembly errors, including 54 missing joins or mis-joins and 6 haplotypic duplications, were corrected by manual curation. This reduced the scaffold number by 2.56% and increased the scaffold N50 by 17.89%. The final assembly has a total length of 627.20 Mb in 342 sequence scaffolds, with 917 gaps, and a scaffold N50 of 175.3 Mb (Table 2).

The 'snail plot' in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents

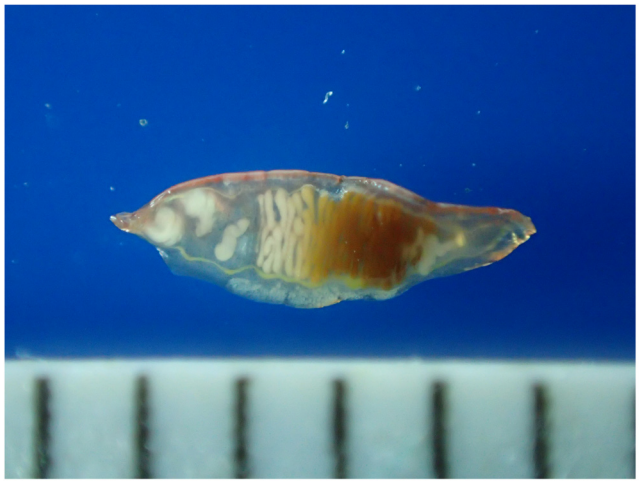

Figure 1. Adult *O. viverrini* acquired from a resected bile duct cancer case. (Not the specimen used for this genome sequencing). Photo taken by Mr. Chitsakul Phuyao, Cholangiocarcinoma Research Institute, Khon Kaen University.

a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (97.89%) was assigned to 6 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by the Hi-C data, are named in order of size (Figure 5; Table 3). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission, and as a separate fasta file.

The final primary assembly has a Quality Value (QV) of 56.9. The *k*-mer completeness for the combined primary and alternate assemblies is 98.55% (79.66% for the primary haplotype and 77.44% for the alternate haplotype). BUSCO (v5.4.3) analysis using the metazoa_odb10 reference set (n = 954) indicated a completeness score of 66.8% (single = 66.0%, duplicated = 0.8%).

Methods

Sample acquisition

One specimen of *Opisthorchis viverrini* (specimen ID SAN21000001, ToLID htOpiVive1) was used for genome sequencing, and another (specimen ID SAN21000002, ToLID htOpiVive2) was used for Hi-C sequencing. The specimens (adult worms) were collected from Khon Kaen, Thailand (latitude 16.468, longitude 102.8299) on 2019-06-30 during a curative surgery for a bile duct cancer treatment. The specimens were collected by Narong Kuntikeo and Attapol Titapun (Khon Kaen University) and identified by Paiboon Sithithaworn (Khon Kaen University). The adult worms were frozen at –80 °C immediately following collection and were subsequently shipped to the Wellcome Sanger Institute, Cambridgeshire, UK on 2022-09-28. The parasites were not cultured.

Table 1. Specimen and sequencing data for *Opisthorchis viverrini*.

Project information			
Study title	Opisthorchis viverrini (Asian liver fluke)		
Umbrella BioProject	PRJEB67413		
Species	*Opisthorchis viverrini*		
BioSample	SAMEA110137926		
NCBI taxonomy ID	6198		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	htOpiVive1	SAMEA110137936	Whole organism
Hi-C sequencing	htOpiVive2	SAMEA110137937	Whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR13361802	8.43e+08	127.26
PacBio Sequel IIe	ERR13363382	2.90e+06	22.18

Table 2. Genome assembly data for *Opisthorchis viverrini*, htOpiVive1.1.

Genome assembly		
Assembly name	htOpiVive1.1	
Assembly accession	GCA_964213165.1	
Accession of alternate haplotype	*GCA_964213135.1*	
Span (Mb)	627.20	
Number of contigs	1,260	
Number of scaffolds	342	
Longest scaffold (Mb)	233.41	
Assembly metrics*		***Benchmark***
Contig N50 length (Mb)	1.0	*≥ 1 Mb*
Scaffold N50 length (Mb)	175.3	*= chromosome N50*
Consensus quality (QV)	56.9	*≥ 40*
k-mer completeness	Primary: 79.66%, alternate: 77.44%, combined: 98.55%	*≥ 95%*
BUSCO v5.4.3 lineage: metazoa_odb10	C:66.8%[S:66.0%,D:0.8%], F:7.8%,M:25.4%,n:954	*S > 90%, D < 5%*
Percentage of assembly mapped to chromosomes	97.89%	*≥ 90%*
Organelles	Mitochondrial genome: 18.04 kb	*complete single alleles*

* Assembly metric benchmarks are adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024.

** BUSCO scores based on the metazoa_odb10 BUSCO set using version 5.4.3. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Opisthorchis_viverrini/dataset/GCA_964213165.1/busco.

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The htOpiVive1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Oatley *et al.*, 2023b). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Hi-C preparation

Tissue from the htOpiVive2 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. Tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation.

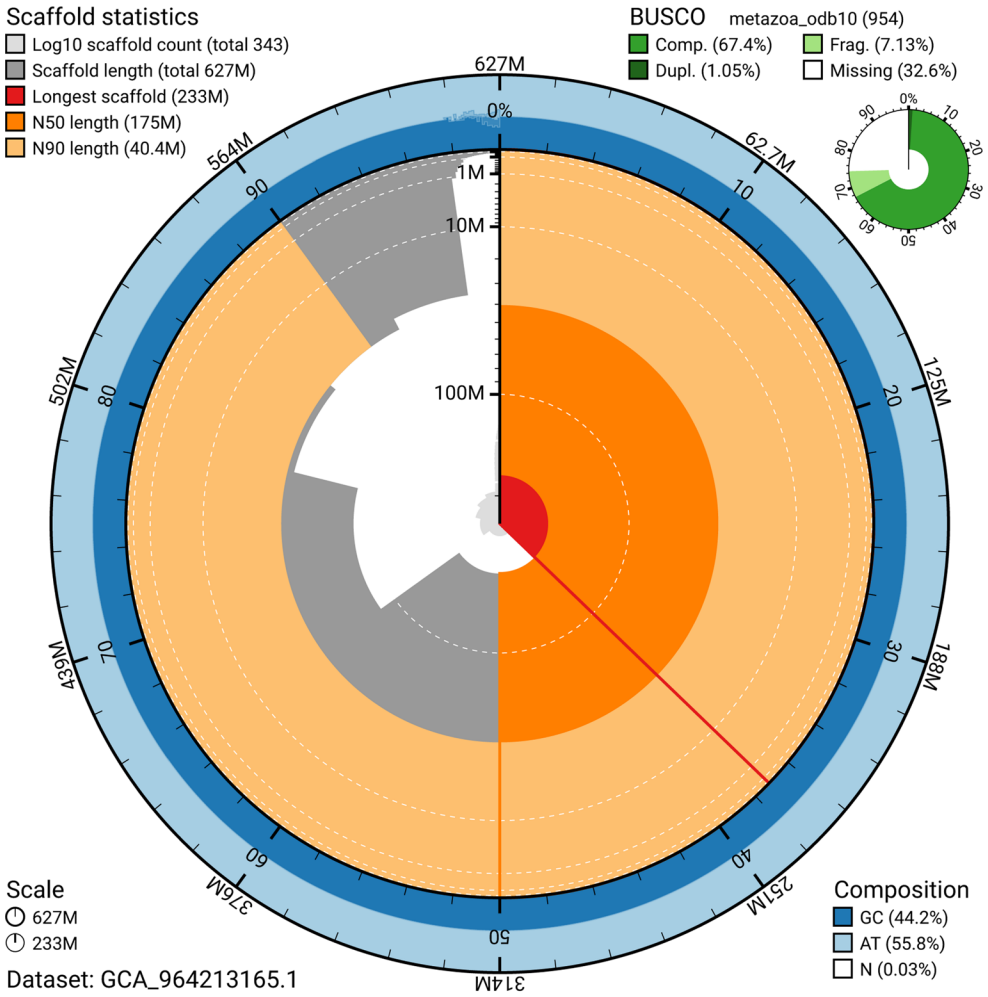

Figure 2. Genome assembly of *Opisthorchis viverrini*, htOpiVive1.1: metrics. The BlobToolKit 'snail plot' provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the metazoa_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964213165.1/dataset/GCA_964213165.1/snail.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Pacific Biosciences HiFi circular consensus DNA sequencing libraries were prepared using the PacBio Express Template Preparation Kit v2.0 (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for removal of single-strand overhangs, DNA damage repair, end repair/A-tailing, adapter ligation, and nuclease treatment. Library preparation also included a library purification step using AMPure PB beads (Pacific Biosciences, California, USA) and size selection step to remove templates shorter than 3 kb using AMPure PB modified SPRI. DNA concentration was quantified using the Qubit Fluorometer v2.0 and Qubit HS Assay Kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument and gDNA 165kb gDNA and 55kb BAC analysis kit. Samples were sequenced using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

For Hi-C library preparation, DNA was fragmented to a size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit following manufacturers'

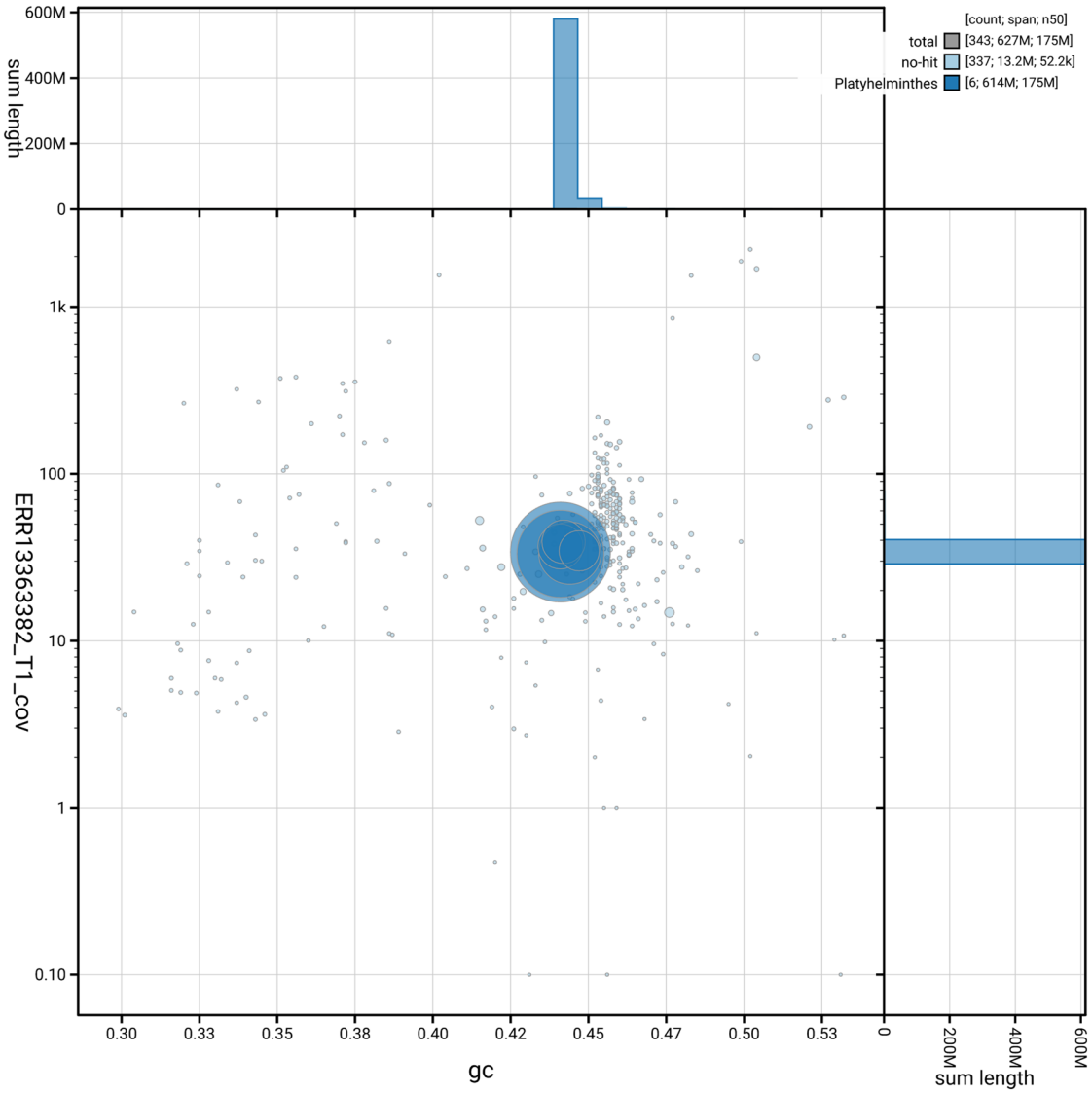

Figure 3. Genome assembly of *Opisthorchis viverrini*, htOpiVive1.1: BlobToolKit GC-coverage plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964213165.1/dataset/GCA_964213165.1/blob.

instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder

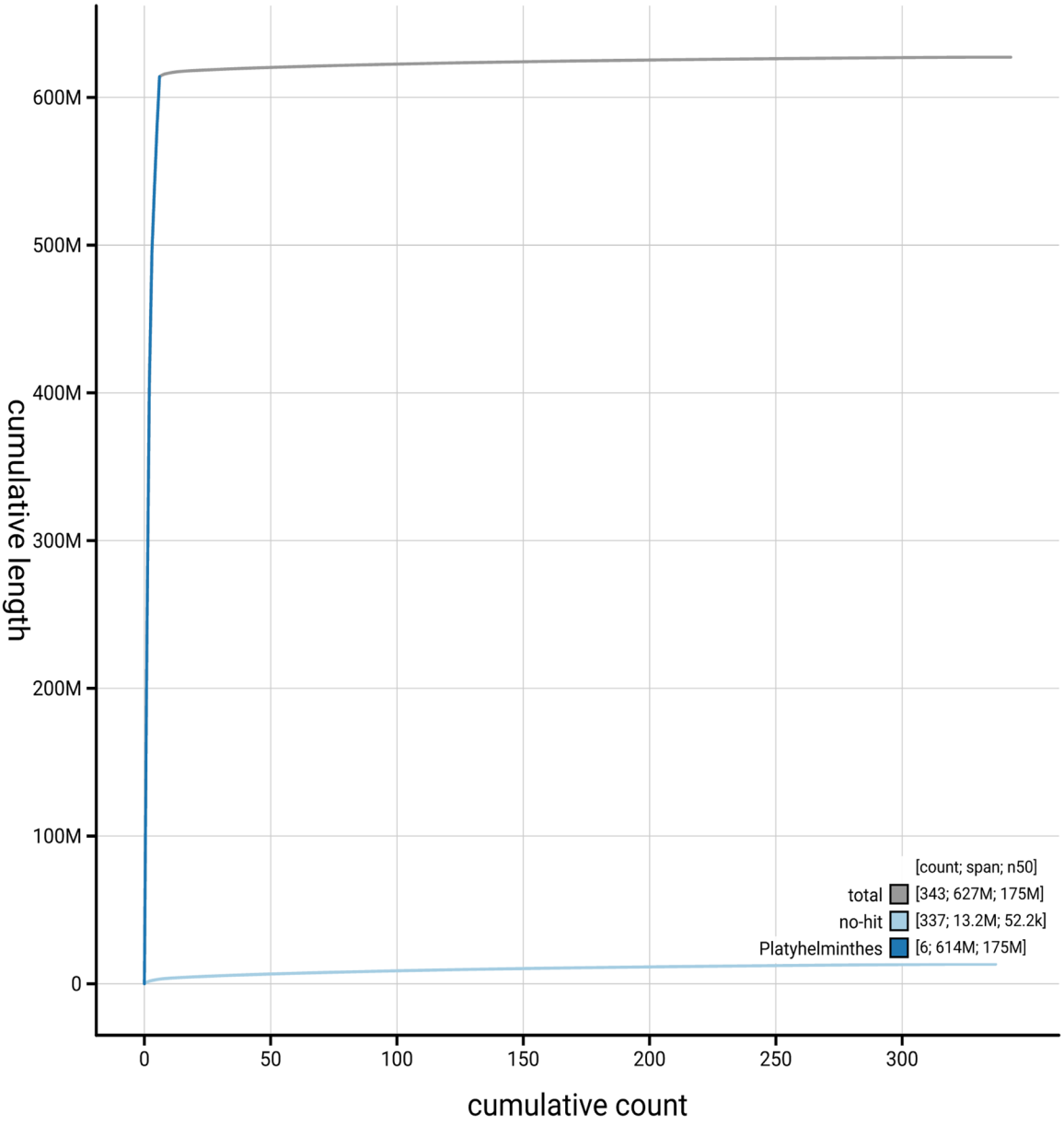

Figure 4. Genome assembly of *Opisthorchis viverrini* htOpiVive1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964213165.1/dataset/GCA_964213165.1/cumulative.

(Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

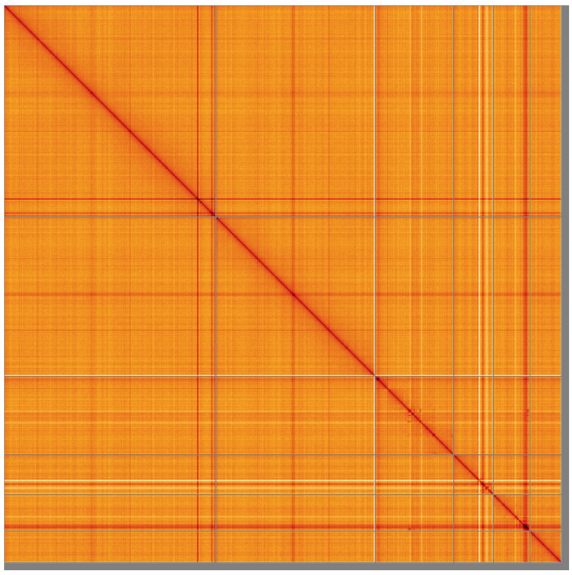

Figure 5. Genome assembly of *Opisthorchis viverrini* htOpiVive1.1: Hi-C contact map of the htOpiVive1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Ehd1atwxSLOgyWAPONN9Nw.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Opisthorchis viverrini*, htOpiVive1.

INSDC accession	Name	Length (Mb)	GC%
OZ171475.1	1	233.41	44.0
OZ171476.1	2	175.33	44.0
OZ171477.1	3	86.63	44.5
OZ171478.1	4	43.83	44.0
OZ171479.1	5	40.37	44.0
OZ171480.1	6	34.43	44.5
OZ171481.1	MT	0.02	40.5

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was used to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer databases (k = 31) that were pre-computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final, public version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these outputs into a blobdir for visualisation.

The genome assembly and evaluation pipelines were developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	http://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r587	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
Merqury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible.

The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material

- Legality of collection, transfer and use (national and international)

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Opisthorchis viverrini* (Asian liver fluke). Accession number PRJEB67413; https://identifiers.org/ena.embl/PRJEB67413. The genome sequence is released openly for reuse by the Wellcome Sanger Institute Tree of Life Programme (https://www.sanger.ac.uk/programme/tree-of-life/). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Author information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.12162482.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.12165051.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.12160324.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.12205391.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the universal protein knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Borlase A, Prada JM, Crellen T: **Modelling morbidity for neglected tropical diseases: the long and winding road from cumulative exposure to long-term pathology.** *Philos Trans R Soc Lond B Biol Sci.* 2023; **378**(1887): 20220279.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Cai XQ, Liu GH, Song HQ, *et al.*: **Sequences and gene organization of the mitochondrial genomes of the liver flukes *Opisthorchis viverrini* and *Clonorchis sinensis* (Trematoda).** *Parasitol Res.* 2012; **110**(1): 235–243.
PubMed Abstract | Publisher Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Crellen T, Sithithaworn P, Pitaksakulrat O, *et al.*: **Towards evidence-based control of *Opisthorchis viverrini*.** *Trends Parasitol.* 2021; **37**(5): 370–380.
PubMed Abstract | Publisher Full Text

Crellen T, Vita F, Braconi C, *et al.*: **Natural history of a parasite-induced biliary cancer.** *medRxiv.* 2024.
Publisher Full Text

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free Full Text

Elkins DB, Haswell-Elkins MR, Mairiang E, *et al.*: **A high frequency of hepatobiliary disease and suspected cholangiocarcinoma associated with heavy *Opisthorchis viverrini* infection in a small community in north-east Thailand.** *Trans R Soc Trop Med Hyg.* 1990; **84**(5): 715–719.
PubMed Abstract | Publisher Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Florio AA, Ferlay J, Znaor A, *et al.*: **Global trends in intrahepatic and extrahepatic cholangiocarcinoma incidence from 1993 to 2012.** *Cancer.* 2020; **126**(11): 2666–2678.
PubMed Abstract | Publisher Full Text | Free Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

Grundy-Warr C, Andrews RH, Sithithaworn P, *et al.*: **Raw attitudes, wetland cultures, life-cycles: socio-cultural dynamics relating to *Opisthorchis viverrini* in the Mekong basin.** *Parasitol Int.* 2012; **61**(1): 65–70.
PubMed Abstract | Publisher Full Text

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive

software distribution for the life sciences. *Nat Methods.* 2018; **15**(7): 475–476.
PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

Harinasuta C, Harinasuta T: ***Opisthorchis viverrini*: life cycle, intermediate hosts, transmission to man and geographical distribution in Thailand.** *Arzneimittelforschung.* 1984; **34**(9B): 1164–1167.
PubMed Abstract

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022.
Reference Source

Haswell-Elkins MR, Mairiang E, Mairiang P, *et al.*: **Cross-sectional study of *Opisthorchis viverrini* infection and cholangiocarcinoma in communities within a high-risk area in northeast Thailand.** *Int J Cancer.* 1994; **59**(4): 505–509.
PubMed Abstract | Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans: **Infection with liver flukes (*Opisthorchis viverrini, Opisthorchis felineus* and *Clonorchis sinensis*).** *IARC Monogr Eval Carcinog Risks Hum.* 1994; **61**: 121–75.
PubMed Abstract | Free Full Text

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Khuntikeo N, Thinkhamrop B, Bundhamcharoen K, *et al.*: **The Socioeconomic burden of cholangiocarcinoma associated with *Opisthorchis viverrini* sensu lato infection in Northeast Thailand: a preliminary analysis.** *Adv Parasitol.* 2018; **102**: 141–163.
PubMed Abstract | Publisher Full Text

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Loilome W, Yongvanit P, Wongkham C, *et al.*: **Altered gene expression in *Opisthorchis viverrini*-associated cholangiocarcinoma in hamster model.** *Mol Carcinog.* 2006; **45**(5): 279–287.
PubMed Abstract | Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
Reference Source

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023a.
Publisher Full Text

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life fragmented DNA clean up: automated SPRI.** *protocols.io.* 2023b.
Publisher Full Text

Ohshima H, Bandaletova TY, Brouet I, *et al.*: **Increased nitrosamine and nitrate biosynthesis mediated by nitric oxide synthase induced in hamsters infected with liver fluke (*Opisthorchis viverrini*).** *Carcinogenesis.* 1994; **15**(2): 271–275.
PubMed Abstract | Publisher Full Text

Pinlaor S, Ma N, Hiraku Y, *et al.*: **Repeated infection with *Opisthorchis viverrini* induces accumulation of 8-nitroguanine and 8-oxo-7,8-dihydro-2′-deoxyguanine in the bile duct of hamsters via inducible nitric oxide synthase.** *Carcinogenesis.* 2004a; **25**(8): 1535–1542.
PubMed Abstract | Publisher Full Text

Pinlaor S, Sripa B, Sithithaworn P, *et al.*: **Hepatobiliary changes, antibody response, and alteration of liver enzymes in hamsters re-infected with *Opisthorchis viverrini.*** *Exp Parasitol.* 2004b; **108**(1–2): 32–39.
PubMed Abstract | Publisher Full Text

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.
Publisher Full Text

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
PubMed Abstract | Publisher Full Text | Free Full Text

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

Shin HR, Oh JK, Masuyer E, *et al.*: **Epidemiology of cholangiocarcinoma: an update focusing on risk factors.** *Cancer Sci.* 2010; **101**(3): 579–585.
PubMed Abstract | Publisher Full Text | Free Full Text

Sithithaworn P, Yongvanit P, Duenngai K, *et al.*: **Roles of liver fluke infection as risk factor for cholangiocarcinoma.** *J Hepatobiliary Pancreat Sci.* 2014; **21**(5): 301–308.
PubMed Abstract | Publisher Full Text

Sota P, Suttiprapa S, Tangkawattana S, *et al.*: **Does *Opisthorchis viverrini* circulate between humans and domestic cats in an endemic area in Thailand?** *Parasitology.* 2022; **149**(10): 1334–1338.
PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Yongvanit P, Pinlaor S, Bartsch H: **Oxidative and nitrative DNA damage: key events in opisthorchiasis-induced carcinogenesis.** *Parasitol Int.* 2012; **61**(1): 130–135.
PubMed Abstract | Publisher Full Text

Young ND, Nagarajan N, Lin SJ, *et al.*: **The *Opisthorchis viverrini* genome provides insights into life in the bile duct.** *Nat Commun.* 2014; **5**(1): 4378.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhao TT, Feng YJ, Doanh PN, *et al.*: **Model-based spatial-temporal mapping of opisthorchiasis in endemic countries of Southeast Asia.** *eLife.* 2021; **10**: e59755.
PubMed Abstract | Publisher Full Text | Free Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 12 February 2025

https://doi.org/10.21956/wellcomeopenres.25958.r117152

✔ **Ferdinand Marlétaz**
Centre for Life's Origin & Evolution, Department of Ecology, Evolution & Environment, University College London, London, UK

The manuscript reports the genome of the liver fluke *Opisthorchis viverrini* sequenced at the Sanger institute assembled in 6 putative chromosomes (233Mb). Relevant and interesting background is provided. The assembly was performed using standard and established procedures and appears sufficiently described. As a comment that I made repeatedly for similar reports, the contact maps appears difficult to read regarding chromosomal identity and contacted and some improvement in the plotting would be desirable.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: comparative genomics, phylogeny of animals, evodevo

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 10 February 2025

https://doi.org/10.21956/wellcomeopenres.25958.r117156

Jose Tort
Genetics, Universidad de la República, Montevideo, Montevideo, 11800, Uruguay

The article describes the sequencing and assembly at chromosome level of the genome of the parasitic trematode *Opisthorchis viverrini*. The procedures are appropriate for the study, the quality of assembly obtained is very good according to provided quality parameters. The procedures and general characteristics of the assembly generated are clear.

There is no annotation associated with this assembly and is reported that this will be done using available RNA-seq data.

The data has been deposited (and are available) at the ENA. Is also reported that the sequences are released at the Tree of Life program but the link provided lead to general information of the program and not to the data. In fact after crawling within the site I could not find the data.

Rather than these, the natural place to locate the data is WormBase Parasite, a well established and widely used database of genomes from diverse helminths, This is the natural place for the research community to seek information on parasitic helminths genomes and associated data (transcriptomes, proteomes, etc).

Contributing to it would help to maintain a valuable resource much appreciated by the community.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Parasite molecular biologiy and genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.