DATA NOTE

The chromosomal genome sequence of the marine leech, *Branchellion lobata* Moore, 1952 and its associated microbial metagenome sequences

[version 1; peer review: 1 approved, 2 approved with reservations]

Shana K. Goffredi[1], Ralph Appy[2], José M. Martín-Durán[3], Graeme Oatley[4], Elizabeth Sinclair[4], Eerik Aunin[4], Noah Gettle[4], Camilla Santos[4], Michael Paulini[4], Haoyu Niu[4], Victoria McKenna[4], Rebecca O'Brien[4], 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, EBI Aquatic Symbiosis Genomics Data Portal Team, Aquatic Symbiosis Genomics Project Leadership

[1]Occidental College, Los Angeles, California, USA
[2]Cabrillo Marine Aquarium, San Pedro, California, USA
[3]Queen Mary University of London School of Biological and Behavioural Sciences, London, England, UK
[4]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 02 Jun 2025, **10**:304
https://doi.org/10.12688/wellcomeopenres.24186.1
Latest published: 02 Jun 2025, **10**:304
https://doi.org/10.12688/wellcomeopenres.24186.1

Abstract

We present a genome assembly from an individual *Branchellion lobata* (the marine leech; Annelida; Clitellata; Hirudinida; Piscicolidae). The genome sequence is 174.1 megabases in span. Most of the assembly is scaffolded into 17 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 16.48 kilobases in length. The metagenome of *Branchellion lobata* was also assembled. Of ten binned metagenomes, three were classified as high-quality assembled genomes (MAGs).

Keywords

Branchellion lobata, marine leech, genome sequence, chromosomal, Hirudinida; microbial metagenome

Open Peer Review

Approval Status ? ? ✓

	1	2	3
version 1 02 Jun 2025	? view	? view	✓ view

1. **Christian Müller**, University of Greifswald, Greifswald, Germany
2. **Wanchang Zhang**, Nanchang University, Nanchang, China
3. **David A WEISBLAT**, University of California, Berkeley, USA

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: Aquatic Symbiosis Genomics Project Leadership (Mark.Blaxter@sanger.ac.uk)

Author roles: Goffredi SK: Conceptualization, Funding Acquisition, Investigation, Resources, Writing – Original Draft Preparation; **Appy R**: Resources; **Martín-Durán JM**: Conceptualization, Resources, Writing – Review & Editing; **Oatley G**: Investigation; **Sinclair E**: Investigation; **Aunin E**: Formal Analysis; **Gettle N**: Formal Analysis, Methodology; **Santos C**: Data Curation, Formal Analysis; **Paulini M**: Data Curation, Formal Analysis; **Niu H**: Project Administration; **McKenna V**: Project Administration; **O'Brien R**: Project Administration;

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the Gordon and Betty Moore Foundation through a grant (GBMF8897) to the Wellcome Sanger Institute to support the Aquatic Symbiosis Genomics Project and by Wellcome through core funding to the Wellcome Sanger Institute (220540). SKG receives a National Science Foundation grant (IOS-1947309).
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: Goffredi SK, Appy R, Martín-Durán JM *et al.* **The chromosomal genome sequence of the marine leech, *Branchellion lobata* Moore, 1952 and its associated microbial metagenome sequences [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2025, **10**:304 https://doi.org/10.12688/wellcomeopenres.24186.1

First published: 02 Jun 2025, **10**:304 https://doi.org/10.12688/wellcomeopenres.24186.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Annelida; Clitellata; Hirudinea; Hirudinida; Oceanobdelliformes; Piscicolidae; *Branchellion*; *Branchellion lobata* Moore, 1952 (NCBI: txid375517).

Background

Leeches are highly unusual annelids (class Clitellata; subclass Hirudinea) present in both terrestrial and aquatic habitats. They display a diversity of behaviours and nutritional strategies, ranging from parasitic blood-feeding to predatory non-blood-feeding. Marine leeches that prey on elasmobranchs and teleosts belong to the family Piscicolidae (order Rhynchobdellida), which comprises ~60 genera and 120 species (Burreson, 2020; Mann, 1962; Sawyer, 1986). Compared to terrestrial leeches, little is known about marine leeches, likely due to their inaccessibility, seasonality, and non-viability in captivity (Kearn, 2005). *Branchellion lobata* Moore 1952, is a species of marine leech found off the coasts of California, Chile, Costa Rica, Mexico, and Panama, that preys upon numerous elasmobranchs, including angel sharks, leopard sharks, bat rays, and butterfly rays (Burreson, 2006; Moser & Anderson, 1977; Ruiz-Escobar & Oceguera-Figueroa, 2019). Due to the difficulty of sampling their nomadic elasmobranch hosts, there are gaps in our understanding of *Branchellion* broadly.

Blood feeders must contend with vitamin deficiencies (especially B vitamins), and red blood cells that are difficult to digest (Toh *et al.*, 2010). To overcome these dietary hurdles, they often partner with internal bacteria that are believed to play an important role in counteracting the low digestibility and vitamin B deficiency, specifically. A study by Goffredi *et al.* (2023) showed that the *Branchellion* microbiome consisted of three main bacteria: *Vibrio* (~50% of the community), and undescribed gammaproteobacterial, which is also present in hatchlings, and an undescribed betaproteobacteria, the latter two each comprising ~16% of

Genome sequence report for the host

The genome was sequenced from an adult *Branchellion lobata* (Figure 1) collected from Mugu Lagoon, Ventura County, California, USA (34.11, –119.13). A total of 201-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 176 missing joins or mis-joins and removed 43 haplotypic duplications, reducing the assembly length by 1.45% and the scaffold number by 60.83%, and increasing the scaffold N50 by 8.41%.

The final assembly has a total length of 174.1 Mb in 84 sequence scaffolds with a scaffold N50 of 10.4 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets

Figure 1. Photograph of a *Branchellion lobata* (not the specimen used for genome sequencing).

of scaffolds assigned to different phyla. Most (98.81%) of the assembly sequence was assigned to 17 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). 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.

The estimated Quality Value (QV) of the final assembly is 47.8 with *k*-mer completeness of 99.95%, and the assembly has a BUSCO v completeness of 72.5% (single = 72.3%, duplicated = 0.2%), using the metazoa_odb10 reference set (n = 954).

Metagenome report

Ten binned genomes were generated from the metagenome assembly (Figure 6), of which three were classified as high-quality metagenome assembled genomes (MAG) (see methods). The completeness values for these assemblies range from approximately 36% to 99% with contamination below 5%. For details on binned genomes see Table 3.

Methods

Sample acquisition

Branchellion lobata specimens were collected from the Pacific round ray (*Urobatis halleri*) from Mugu lagoon salt marsh (34°06'19.5"N 119°07'33.2"W; Ventura County, California) on Sept 23, 2021. The host animal was collected by beach seine under specific use permit ID: S-200810003-20163-001 to Ralph Appy). Leeches were removed and preserved on dry ice, before storing at –80 °C. Specimen ID QMOUL0000035 (ToLID wcBraLoba9) was used for long read PacBio sequencing, specimen ID QMOUL0000034 (ToLID wcBraLoba8) was used for Hi-C sequencing and specimen ID QMOUL0000031 (ToLID wcBraLoba5) was used for RNA sequencing.

Table 1. Genome data for *Branchellion lobata*, wcBraLoba9.1.

Project accession data	
Assembly identifier	wcBraLoba9.1
Species	*Branchellion lobata*
Specimen	wcBraLoba9
NCBI taxonomy ID	375517
BioProject	PRJEB57262
BioSample ID	Genome sequencing: SAMEA11604974 Hi-C scaffolding: SAMEA11604973 RNA sequencing: SAMEA11604970
Isolate information	wcBraLoba9: whole organism (genome sequence) wcBraLoba8: whole organism (Hi-C sequencing) wcBraLoba5: whole organism (RNA sequencing)
Assembly metrics	
Consensus quality (QV)	47.8
k-mer completeness	99.95%
BUSCO*	C:72.5%[S:72.3%,D:0.2%], F:7.4%,M:20.1%,n:954
Percentage of assembly mapped to chromosomes	98.81%
Organelles	Mitochondrial genome: 16.48 kb
PacificBiosciences Sequel IIe	ERR10462068, ERR10462069, ERR10462070
Hi-C Illumina	ERR10466804
PolyA RNA-Seq Illumina	ERR12708735
Genome assembly	
Assembly accession	GCA_947562095.1
Accession of alternate haplotype	GCA_947563305.1
Span (Mb)	174.1
Number of contigs	720
Contig N50 length (Mb)	0.5
Number of scaffolds	84
Scaffold N50 length (Mb)	10.4
Longest scaffold (Mb)	14.37

* BUSCO scores based on the metazoa_odb10 BUSCO set using version 5.3.2. 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/Branchellion lobata/dataset/CANOAO01/busco.

Nucleic acid extraction

Protocols developed by the Wellcome Sanger Institute (WSI) Tree of Life laboratory are publicly available on protocols.io (Howard *et al.*, 2025). The workflow for high molecular weight (HMW) DNA extraction includes a sequence of procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the wcBraLoba9 sample was weighed and dissected on dry ice

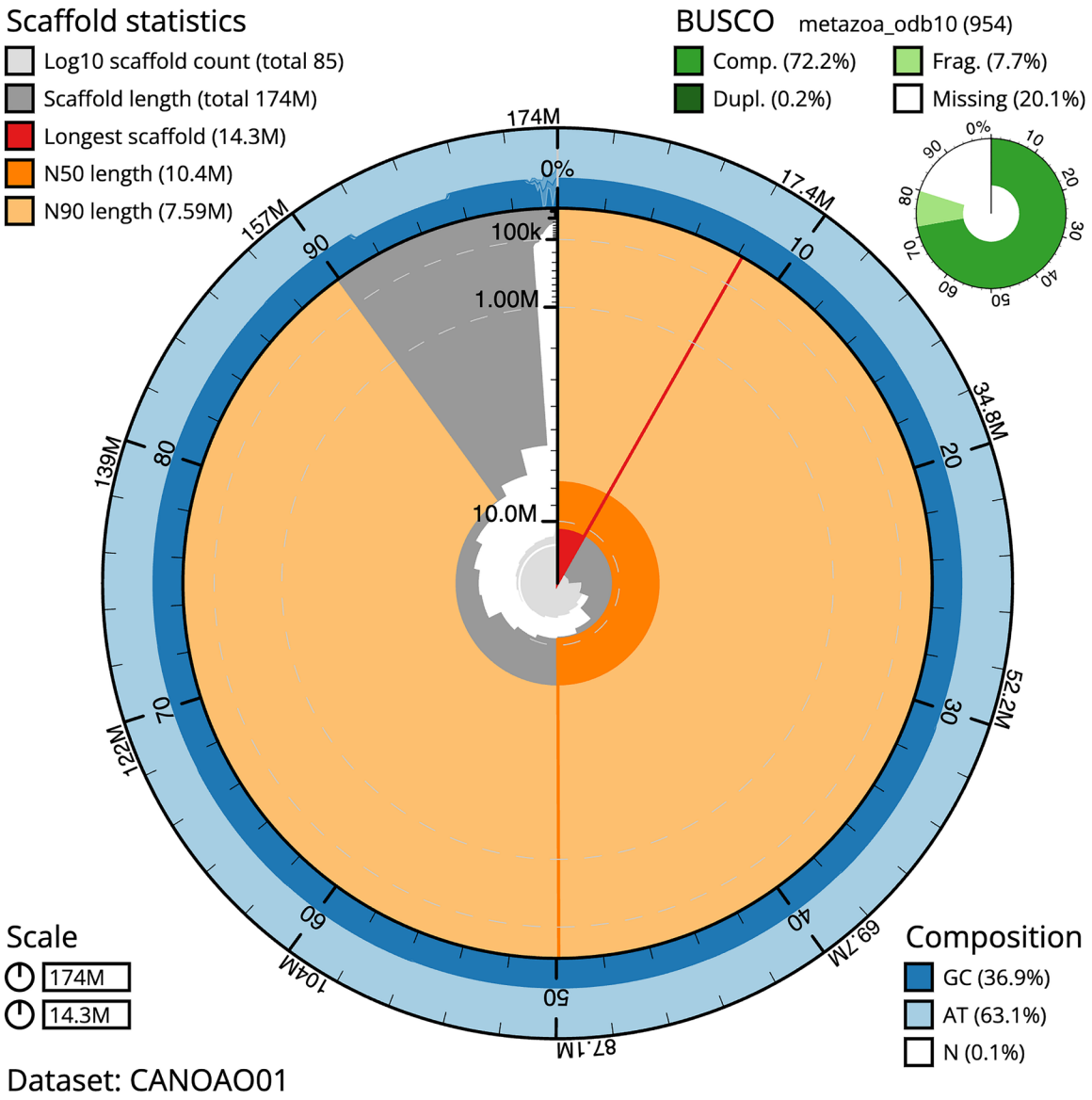

Figure 2. Genome assembly of *Branchellion lobata*, wcBraLoba9.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 174,144,220 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (14,345,173 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (10,374,467 and 7,593,766 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the metazoa_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Branchellion%20lobata/dataset/CANOAO01/snail.

(Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023). HMW DNA was extracted using the Manual MagAttract v1 protocol (Strickland *et al.*, 2023b). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023a), using AMPure PB beads to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

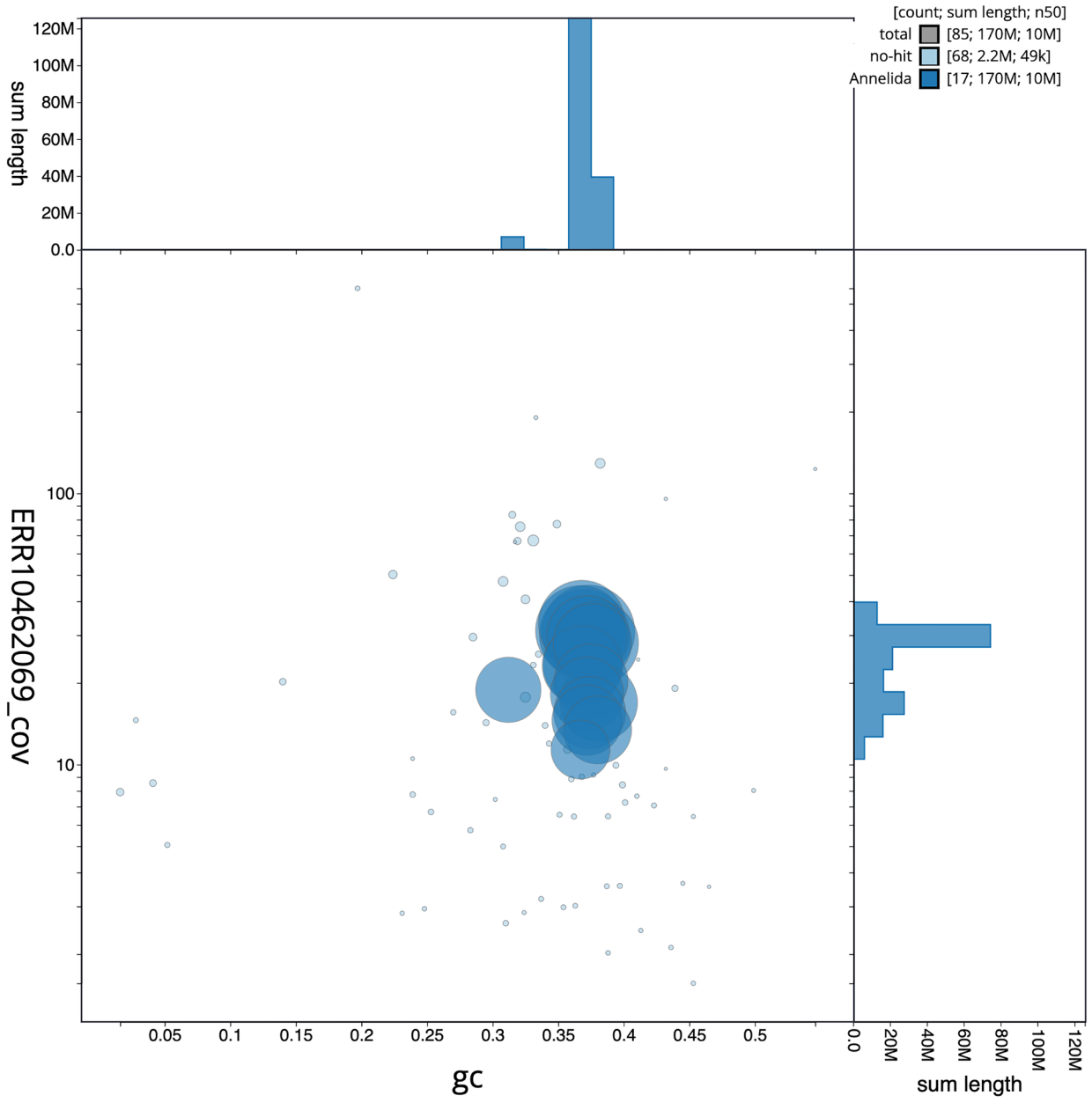

Figure 3. Genome assembly of *Branchellion lobata*, wcBraLoba9.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Branchellion%20lobata/dataset/CANOAO01/blob.

RNA was extracted from whole organism tissue of wcBraLoba5 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences Sequel IIe (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also

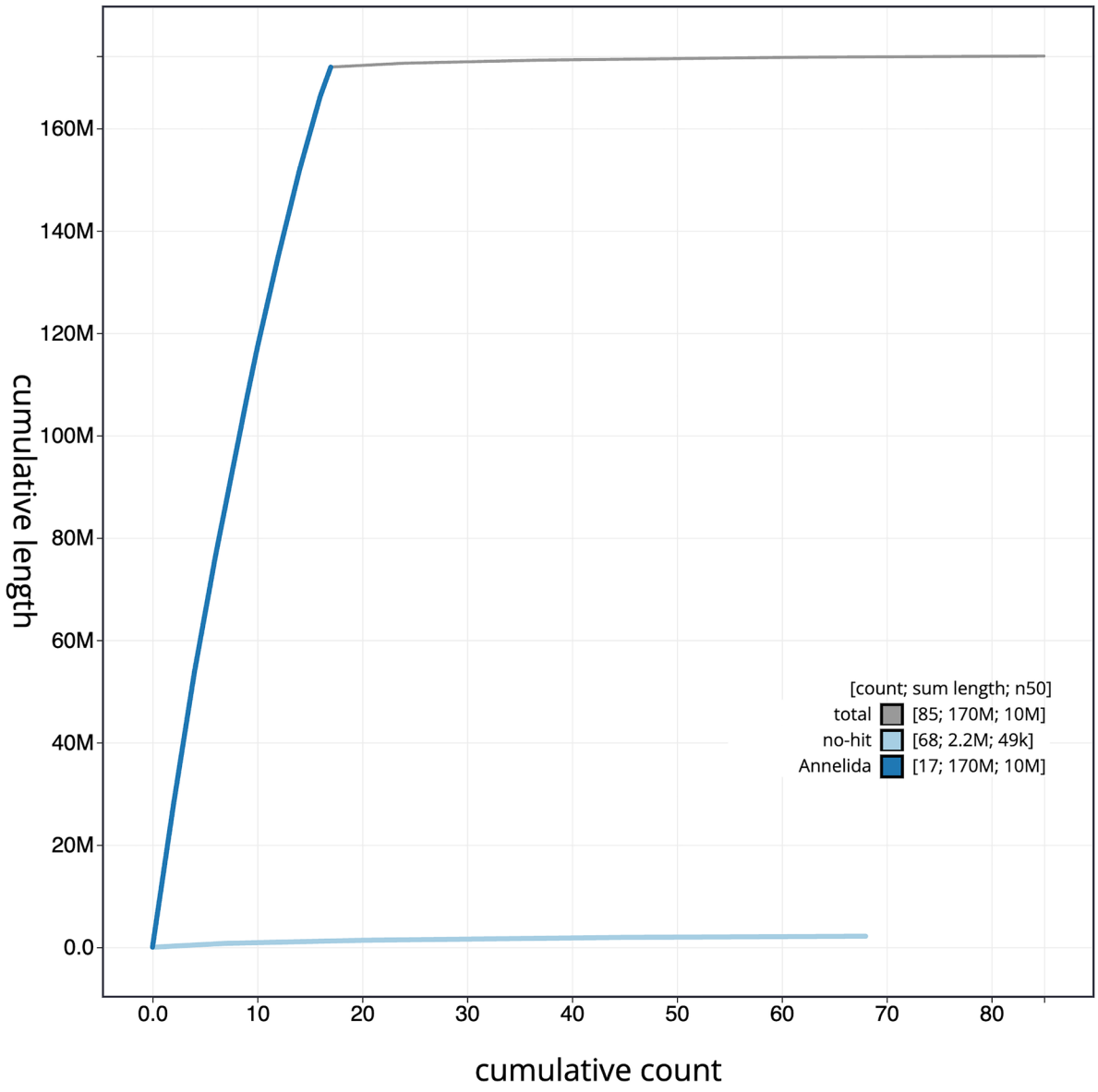

Figure 4. Genome assembly of *Branchellion lobata*, wcBraLoba9.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Branchellion%20lobata/dataset/CANOAO01/cumulative.

generated from whole organism tissue of wcBraLoba8 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly and curation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

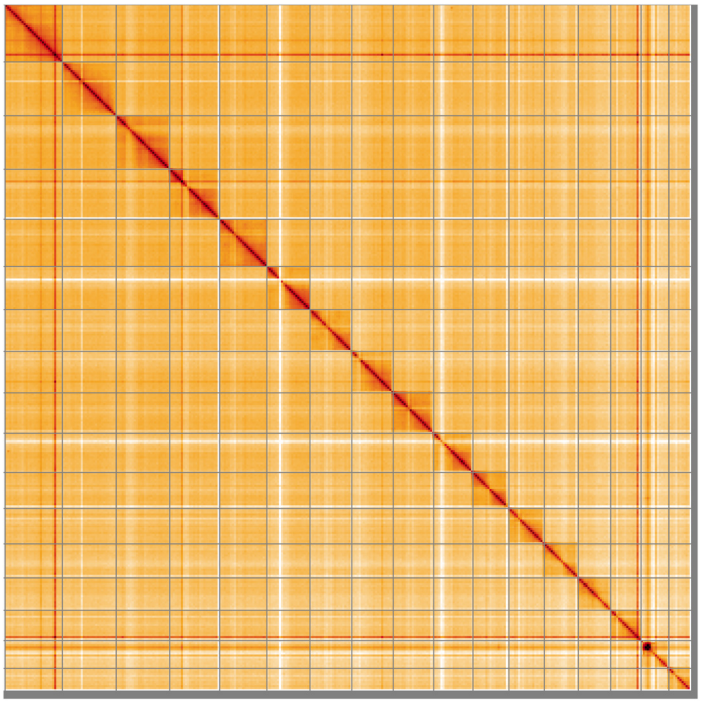

Figure 5. Genome assembly of *Branchellion lobata*, wcBraLoba9.1: Hi-C contact map of the wcBraLoba9.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=OLHyD15BQFuW1A4iz0VWpw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Branchellion lobata*, wcBraLoba9.

INSDC accession	Name	Length (Mb)	GC%
OX387246.1	1	14.35	37.5
OX387247.1	2	13.48	37.0
OX387248.1	3	13.42	36.5
OX387249.1	4	12.52	37.0
OX387250.1	5	11.81	38.0
OX387251.1	6	10.81	37.0
OX387252.1	7	10.49	37.0
OX387253.1	8	10.37	37.0
OX387254.1	9	10.13	37.5
OX387255.1	10	9.84	38.0
OX387256.1	11	9.01	37.5
OX387257.1	12	8.88	37.0
OX387258.1	13	8.52	37.5
OX387259.1	14	8.11	37.0
OX387260.1	15	7.59	38.0
OX387261.1	16	6.98	31.0
OX387262.1	17	5.67	36.5
OX387263.1	MT	0.02	20.0

The metagenome assembly was generated using Hifiasm (Cheng *et al.*, 2021) and binned using MetaBAT2 (Kang *et al.*, 2019), MaxBin (Wu *et al.*, 2014), bin3C (DeMaere & Darling, 2019), and MetaTOR. The resulting bin sets of each binning algorithm were individually optimised and then collectively refined using DAS Tool (Sieber *et al.*, 2018). PROKKA (Seemann, 2014) was used to identify tRNAs and rRNAs in each bin, CheckM (Parks *et al.*, 2015) (checkM_DB release 2015-01-16) was used to assess bin completeness/contamination, and GTDB-TK (Chaumeil *et al.*, 2022) (GTDB release 214) was used to taxonomically classify bins. Taxonomic replicate bins were identified using dRep (Olm *et al.*, 2017) with default settings (95% ANI threshold). The final bin set was filtered for bacteria and archaea. All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: contamination ≤ 5%, presence

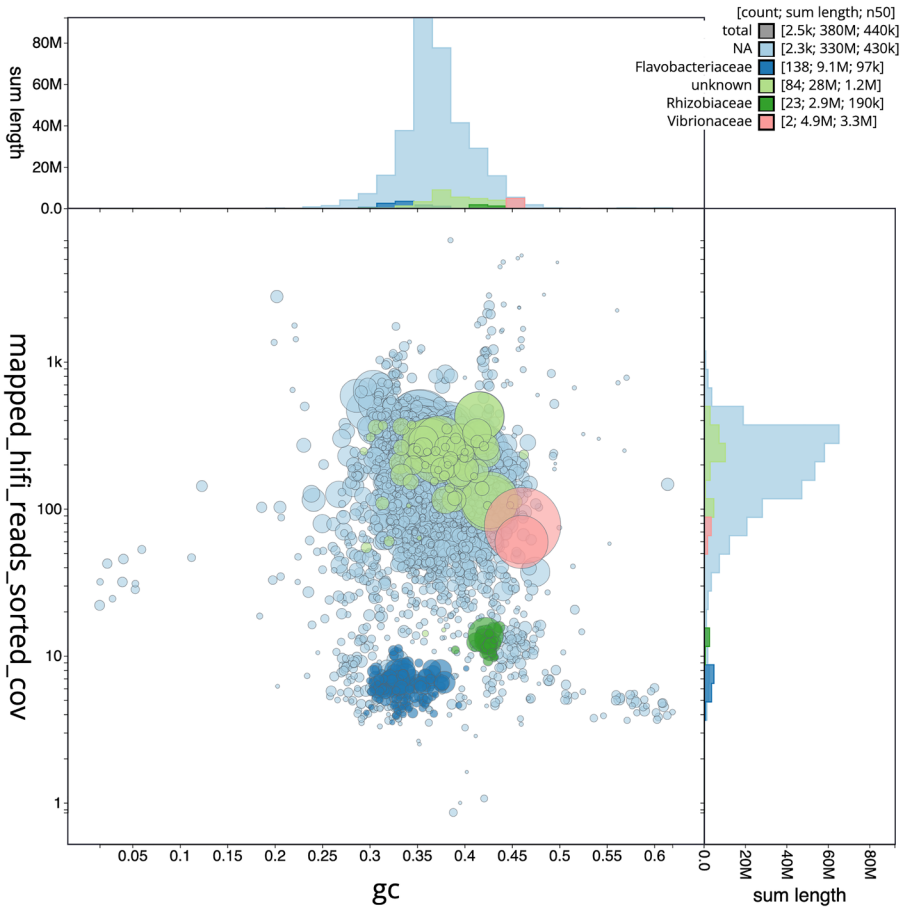

Figure 6. Blob plot of base coverage in mapped against GC proportion for sequences in the metagenome of the *Branchellion lobata* sample. Sequences are coloured by family. Circles are sized in proportion to sequence length on a square-root scale, ranging from 1,608 to 3,301,098. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure may be viewed here.

of 5S, 16S, and 23S rRNA genes, at least 18 unique tRNAs, and either ≥ 90% completeness or ≥ 50% completeness with fully circularised chromosomes. Bins that did not meet these thresholds, or were identified as taxonomic replicates of MAGs, were retained as 'binned metagenomes' provided they had ≥ 50% completeness and ≤ 10% contamination. Software tool versions and sources are given in Table 4.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), 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 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 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 genome was also analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021) were calculated.

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

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

Table 3. Quality metrics and taxonomic assignments of the MAGs.

NCBI taxon	Taxid	GTDB taxonomy	Quality	Size (bp)	Contigs	Circular	Mean coverage	Completeness (%)	Contamination (%)
Rhizobiaceae bacterium	271151	g__JAALLB01	High	2,997,465	23	No	83.83	91.08	4.32
uncultured *Mesonia* sp.	399731	g__Mesonia	High	2,787,421	29	No	93.98	92.54	0.19
Vibrio sinaloensis	379097	s__Vibrio sinaloensis	High	4,973,486	2	Partial	61.73	99.36	4.52
Gammaproteobacteria bacterium	86473	f__JAAOII01	Medium	1,857,420	1	No	83.39	86.44	3.27
Burkholderiales bacterium	208544	o__Burkholderiales	Medium	23,459,930	80	Partial	100.72	89.25	2
uncultured Aureibaculum sp.	2911477	g__Aureibaculum	Medium	3,158,360	29	No	97.02	84.94	0.13
Mesoflavibacter zeaxanthinifaciens	393060	s__Mesoflavibacter zeaxanthinifaciens	Medium	2,138,924	43	No	91.76	66.11	0
Burkholderiales bacterium	208544	o__Burkholderiales	Medium	1,393,589	1	Yes	94.27	96.34	0
Bizionia echini	649333	s__Algorimicrobium echini	Low	1,216,795	37	No	93.03	36.25	0
Gammaproteobacteria bacterium	86473	f__JAAOII01	Medium	1,843,069	2	No	88.25	86.44	1.23

Wellcome Open Research 2025, 10:304 Last updated: 05 AUG 2025

Table 4. Software tools: versions and sources.

Software tool	Version	Source
bin3C	version 0.3.3	https://github.com/cerebis/bin3C
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
CheckM	1.2.1	https://github.com/Ecogenomics/CheckM
DAS Tool	1.1.5	https://github.com/cmks/DAS_Tool
dRep	3.4.0	https://github.com/MrOlm/drep
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
gEVAL	N/A	https://geval.org.uk/
GTDB-TK	2.3.2	https://github.com/Ecogenomics/GTDBTk
Hifiasm	0.16.1	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
MaxBin	version 2.7	https://github.com/jtamames/SqueezeMeta/tree/master/bin/MaxBin
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MetaBAT2	version 2.15-15-gd6ea400	https://bitbucket.org/berkeleylab/metabat/src/master/
MetaTOR	-	https://github.com/koszullab/metaTOR
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
PROKKA	1.14.5	https://github.com/vdejager/prokka
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
YaHS	1.1a.2	https://github.com/c-zhou/yahs

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: *Branchellion lobata* (marine leech). Accession number PRJEB57262; https://identifiers.org/ena.embl/PRJEB57262. The genome sequence is released openly for reuse. The *Branchellion lobata* genome sequencing initiative is part of the Aquatics Symbiosis Genomics (ASG) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

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.10066175.

Members of the Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.14870789.

Members of the European Bioinformatics Institute ASG Data Portal team are listed here: https://doi.org//10.5281/zenodo.10076466.

Members of the Wellcome Sanger Institute/Aquatic Symbiosis Genomics Project Leadership are listed here: https://doi.org/10.5281/zenodo.10184833.

Acknowledgements

Special thanks to Martin Ruane, U.S. Navy, Point Mugu Air Station for providing access to round stingrays at the Mugu Lagoon.

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

Burreson EM: **Phylum Annelida: Hirudinea as vectors and disease agents.** In: *Fish diseases and disorders. Volume 1: protozoan and metazoan infections.* Wallingford, UK: CABI, 2006; 566–591.
Publisher Full Text

Burreson EM: **Marine and estuarine leeches (Hirudinida: Ozobranchidae and Piscicolidae) of Australia and New Zealand with a key to the species.** *Invertebr Syst.* 2020; **34**(3): 235–259.
Publisher 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

Chaumeil PA, Mussig AJ, Hugenholtz P, *et al.*: **GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database.** *Bioinformatics.* 2022; **38**(23): 5315–5316.
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

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
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

DeMaere MZ, Darling AE: **bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes.** *Genome Biol.* 2019; **20**(1): 46.
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.* 2023.
Publisher Full Text

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Goffredi SK, Appy RG, Hildreth R, *et al.*: **Marine vampires: persistent, internal associations between bacteria and blood-feeding marine annelids and crustaceans.** *Front Microbiol.* 2023; **13**: 1113237.
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

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

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *BioRxiv.* 2025.
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

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

Kang DD, Li F, Kirton E, *et al.*: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.** *PeerJ.* 2019; **7**: e7359.
PubMed Abstract | Publisher Full Text | Free Full Text

Kearn GC: **Leeches, lice and lampreys: a natural history of skin and gill parasites of fishes.** Springer Science & Business Media, 2005.
Reference Source

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

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

Mann KH: **Leeches (Hirudinea): their structure, physiology, ecology and embryology.** Oxford: Pergamon Press, 1962.
Reference Source

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

Moser M, Anderson S: **An intrauterine leech infection: *Branchellion lobata* Moore, 1952 (Piscicolidae) in the Pacific angel shark (*Squatina californica*) from California.** *Can J Zool.* 1977; **55**(4): 759–760.
Publisher Full Text

Olm MR, Brown CT, Brooks B, *et al.*: **dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication.** *ISME J.* 2017; **11**(12): 2864–2868.
PubMed Abstract | Publisher Full Text | Free Full Text

Parks DH, Imelfort M, Skennerton CT, *et al.*: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res.* 2015; **25**(7): 1043–55.
PubMed Abstract | Publisher Full Text | Free 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, 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

Ruiz-Escobar F, Oceguera-Figueroa A: **A new species of *Branchellion* Savigny, 1822 (Hirudinida: Piscicolidae), a marine leech parasitic on the giant electric ray *Narcine entemedor* Jordan & Starks (Batoidea: Narcinidae) off Oaxaca, Mexico.** *Syst Parasitol.* 2019; **96**(7): 575–584.
PubMed Abstract | Publisher Full Text

Sawyer RT: **Leech biology and behaviour. Vol. 2. Feeding biology, ecology, and systematics.** Oxford, UK: Clarendon Press, 1986.
Reference Source

Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics.* 2014; **30**(14): 2068–2069.
PubMed Abstract | Publisher Full Text

Sieber CMK, Probst AJ, Sharrar A, *et al.*: **Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy.** *Nat Microbiol.* 2018; **3**(7): 836–843.
PubMed Abstract | Publisher Full Text | Free Full Text

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023a.
Publisher Full Text

Strickland M, Moll R, Cornwell C, *et al.*: **Sanger Tree of Life HMW DNA extraction: manual MagAttract.** *protocols.io.* 2023b.
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
Publisher Full Text

Toh SQ, Glanfield A, Gobert GN, *et al.*: **Heme and blood-feeding parasites: friends or foes?** *Parasit Vectors.* 2010; **3**(1): 108.
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

Wu YW, Tang YH, Tringe SG, *et al.*: **MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm.** *Microbiome.* 2014; **2**(1): 26.
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 05 August 2025

https://doi.org/10.21956/wellcomeopenres.26679.r125791

✓ **David A WEISBLAT**
University of California, Berkeley, California, USA

Beyond the general interest in accumulating as inclusive a set of genome sequences as possible, sequencing genomes of leeches and other ciltellate annelids is of interest because of the extreme acceleration of genome rearrangements within this group of annelids. A dense sampling of clitellate genomes is needed to reconstruct the history of these genome rearrangements and eventually understand the underlying cause(s).

I do not have the expertise to judge the technical aspects of this work, but it appears to be carefully done. It would be nice to see how the inferred number of 17 chromosomes compares to an cytologically determined karyotype.
My only substantive comment is in reference to a statement in the Background "Marine leeches that prey on elasmobranchs and teleosts belong to the family Piscicolidae (order Rhynchobdellida)....". My understanding is that Rhynchobdellida, though still commonly recognized, may not be a monophyletic grouping. Traditionally, rhynchobdellid leeches include both glossiphoniid and piscicolid leeches (proboscis-bearing species laying large yolky eggs), as distinct from a clade containing erpobdellid and hirudinid species (jawed instead of proboscis-bearing, and laying small embryos in albumen-filled cocoons). However, more recent molecular phylogenies based on extensive nuclear gene data (including a collaboration in which I am participating:
Schultz D, et al., 2024 [Ref 1]*) supports a tree in which piscicolids are sister to the group of erpobdellids and hirudinids, with glossiphoniids as an outgroup to all of these. This revised tree is significant for leech biologists because it supports the hypothesis that ancestral leeches had probosces and laid large yolky eggs and pushes the evolution of jawed leeches with small embryos to more recent times.*
I note that the sentence in question is also at odds with the species taxonomy given in the article as well.

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.

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 24 July 2025

https://doi.org/10.21956/wellcomeopenres.26679.r125785

?

Wanchang Zhang
Nanchang University, Nanchang, China

The authors present the genome assembly of a marine leech, *Branchellion lobata*, which feeds on teleost hosts. The organism is interesting. The genome assembly of *Branchellion lobata* is valuable for future research. The authors clearly declared the results and methods of the genome and metagenome assembly of the leech. The manuscript is clearly written; I believe the manuscript could be indexed after addressing some minor comments.
It seems that the authors didn't conduct genome annotation to the leech assembly, how many genes are annotated along the genome? Once a genome was published, the genome annotation is also important to perform downstream functional analyses.
A comparable chromosome Syteny analysis between this organism and other close leech organism could be carried out to assess the genome completeness.
Figure 1, a scale bar should be present.

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: Fish genetics

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, however I have significant reservations, as outlined above.

Reviewer Report 27 June 2025

https://doi.org/10.21956/wellcomeopenres.26679.r125053

?

Christian Müller
University of Greifswald, Greifswald, Germany

Manuskript-ID: 24186
Journal: Wellcome Open Research
Title: The chromosomal genome sequence of the marine leech, *Branchellion lobata* Moore, 1952 and its associated microbial metagenome sequences.
Authors: Goffredi SK, Appy R, Martín-Durán JM, Oatley G, Sinclair E, Aunin E, Gettle N, Santos C, Paulini M, Niu H, McKenna V, O'Brien R

Short Summary
The author(s) submitted a manuscript that describes the generation, assembly and quality analysis of the almost complete chromosomal genome and the complete mitochondrial genome of *Branchellion lobata*, a marine leech that feeds on cartilaginous fishes. In addition, the metagenome of *Branchellion lobata* was determined as well, including the genomes of the three most abundant bacterial species.

General Impression
The manuscript is part of the Tree of Life project in general and the Aquatic Symbiosis Genomics Project in detail. It addresses a topic that is in the first instance of interest mainly for biologist and ecologists, but beyond that also for a broad audience as it helps to shed light on the diversity of life. The manuscript is in line with a previous publication on the complete genome of *Piscicola geometra*, a limnic fish leech (**Doe et al. 2023**), and follows in style and content comparable genome papers that were published in Wellcome Open Research.
I fully appreciate the efforts of the authors, and I have no doubts that all experiments and analyses were performed with accuracy. The manuscript is well written; the results and data are clearly presented. In that context I strongly recommend its indexing. There are only a few minor remarks that need to be addressed.

Minor remarks:
1) The last sentence of the Background chapter is incomplete ("the latter two each comprising ~16% of "). Of what?
2) Figure 1: please indicate the source of the image
3) The manuscript lacks any information on the sources of the lab equipment, the reagents and kits that were used.
4) The manuscript lacks any further analyses like the identification of putative anticoagulants. At least a short comparison with the genome of *Piscicola geometra* would, in my opinion, significantly enhance the strength of the manuscript. Please consider.

References
1. Doe J, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, et al.: The genome sequence of the fish leech, Piscicola geometra (Linnaeus, 1761). *Wellcome Open Research*. 2023; **8**. Publisher Full Text

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?
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: animal physiology

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, however I have significant reservations, as outlined above.