Check for updates

DATA NOTE

ERGA-BGE reference genome of *Placobdella costata*: an ectoparasitic glossiphoniid leech of freshwater turtles with a Palearctic distribution

[version 1; peer review: awaiting peer review]

Alejandro Manzano-Marín[1], Clemens Grosser[2], Astrid Böhne[3], Rita Monteiro[3], Thomas Marcussen[4,5], Torsten H. Struck[4], Rebekah A. Oomen[4-8],
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,
Caroline Howard[9], Kerstin Howe[9], Mark Blaxter[9], Shane McCarthy[9], Jonathan M.D. Wood[9], Fergal Martin[10], Leanne Haggerty[10], Chiara Bortoluzzi[11]

[1]Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Vienna, 1030, Austria
[2]Am Wasserturm 20, Independent researcher, Elstertrebnitz, 04523, Germany
[3]Museum Koenig Bonn, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, 53113, Germany
[4]Natural History Museum, University of Oslo, Oslo, Norway
[5]Centre for Ecological & Evolutionary Synthesis, University of Oslo, Oslo, Norway
[6]Department of Biological Sciences, University of New Brunswick Saint John, Saint John, New Brunswick, Canada
[7]Tjärnö Marine Laboratory, University of Gothenburg, Gothenburg, Sweden
[8]Centre for Coastal Research, University of Agder, Kristiansand, Norway
[9]Tree of Life, Wellcome Genome Campus, Wellcome Sanger Institute, Cambridge, England, CB10 1SA, UK
[10]Wellcome Genome Campus, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK
[11]Environmental Bioinformatics group, Amphipôle, Quartier UNIL-Sorge, SIB Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland

v1 **First published:** 06 Oct 2026, **6**:395
https://doi.org/10.12688/openreseurope.24522.1
Latest published: 06 Oct 2026, **6**:395
https://doi.org/10.12688/openreseurope.24522.1

Abstract

Placobdella costata, commonly known as the European pond turtle leech, is one of two Palearctic representatives of its genus, with most recognised diversity being North American. Current molecular evidence supports the hypothesis of a single dispersal event from North America to Europe. *Placobdella costata* is primarily a blood-feeding ectoparasite of freshwater turtles and has been identified as a

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

vector of the apicomplexan *Haemogregarina stepanowi.* Like other blood-feeding glossiphoniids, it hosts a nutritional symbiont in specialised organs called bacteriomes. We expect the genome of *P. costata* to not only contribute to the hitherto poor genomic resources available for this leech family, but to also serve as a starting point to explore the evolution of new organs (such as the bacteriome), body parts (such as a proboscis), and biological traits (such as parental care), as well as to disentangle the host-microbe interactions between this leech and its nutritional symbiont. The genome of *P. costata* was assembled into two haplotypes. Results are reported for the primary haplotype, here used as reference. A total of 18 contiguous chromosomal pseudomolecules were assembled from the genome sequence along with the mitochondrial genome. This chromosome-level assembly encompasses 0.48 Gb, composed of 688 contigs and 29 scaffolds, with contig and scaffold N50 values of 1.76 Mb and 31.5 Mb, respectively.

Keywords
Placobdella costata, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Glossiphoniidae, see turtle leech



This article is included in the Horizon Europe gateway.



This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Corresponding author: Alejandro Manzano-Marín (alejandro.manzano.marin@gmail.com)

Author roles: Manzano-Marín A: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Grosser C**: Investigation, Resources, Writing – Review & Editing; **Böhne A**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **H. Struck T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **A. Oomen R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Howard C**: Investigation, Supervision, Writing – Review & Editing; **Howe K**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Blaxter M**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **McCarthy S**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Wood JMD**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Bortoluzzi C**: Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Biodiversity Genomics Europe (Grant no. 101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/TAI1450724.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Manzano-Marín A *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Manzano-Marín A, Grosser C, Böhne A *et al.* **ERGA-BGE reference genome of *Placobdella costata*: an ectoparasitic glossiphoniid leech of freshwater turtles with a Palearctic distribution [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:395 https://doi.org/10.12688/openreseurope.24522.1

First published: 06 Oct 2026, **6**:395 https://doi.org/10.12688/openreseurope.24522.1

Introduction

Placobdella costata, a member of the Glossiphoniidae family (Annelida: Euhirudinea) also known as "Schildkrötenegel" (in German), is a freshwater leech found parasitising primarily freshwater turtles (*Emys* and *Mauremys* species) (Figure 1). Members of the genus *Placobdella* can be distinguished by the possession of two pairs of eyespots, a single pair of caecate bacteriomes attached to the base of the oesophagus, and bilobed ovaries (Siddall et al., 2005). This last characteristic is not shared with known Palearctic species of the genus, which display simple ovaries (Ben Ahmed et al., 2023). Like most Glossiphoniidae, *P. costata* is dorso-ventrally flattened and displays parental care, carrying developing eggs inside thin-walled, transparent cocoons on its belly at least until the first feeding of the young leeches (Kutschera and Peter, 2001). Recently, molecular analysis of Palearctic *Placobdella* specimens has revealed the presence of cryptic diversity (Kvist et al., 2022), with one novel species (*Placobdella nabeulensis*) recently described from within specimens identified as *Placobdella costata* (Ben Ahmed et al., 2023). Pending further taxonomic study of the *P. costata* species complex, Ben Ahmed et al. (2023) suggest preliminarily the use of the atrium shape to discriminate between *P. nabeulensis* (T-shaped atrium) and *P. costata* (U-shaped atrium with pronounced deepening between the two atrial cornuae). The specimen selected for sequencing belongs to the so-called C4 clade of *P. costata*, following its sampling location and *cox1* gene sequence (Kvist et al., 2022).

As a blood-feeding ectoparasite, *P. costata* has been identified as both a vector and definitive host of the apicomplexan *Haemogregarina stepanowi*, a parasite of the European pond turtle (*Emys orbicularis*) (Dvořáková et al., 2014). *Placobdella costata* can be found attached to the underside of rocks and other submerged substrates, as well as attached to turtle hosts. It has also been observed to feed on humans (Cichocka et al., 2021). *Placobdella costata* is known to host a strain of the symbiotic alphaproteobacterium "*Candidatus* Reichenowia" (*Hyphomicrobiales*: *Rhizobiaceae*) inside its oesophageal bacteriomes located at the base of the proboscis (Manglicmot et al., 2020), with a putative role as a B vitamin provider.

Developing a high-quality reference genome for *P. costata* will contribute to different areas of research. First, it will serve as a starting point to further unravel the Palaearctic cryptic diversity of specimens identified as *P. costata.* It will help advance research on the evolution of common glossiphoniid traits, such as parental care and a dorso-ventrally flattened body. Additionally, as a bacteriome-bearing leech, its genome is expected to advance research into the evolution of this novel organ. Moreover, analysis of its genome will advance the understanding of potential host-microbe interactions, moving forward symbiosis research in Euhirudinea. More broadly, as one of just a handful of leech reference genomes, it sets the stage for a deeper understanding of the evolutionary traits of this group of annelids.

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni et al., 2023).

Materials & methods

Sample and sampling information

On 31st July 2023, Clemens Grosser sampled 70 juvenile specimens (monoecious hermaphrodite) of *Placobdella costata* in Zschiesewitz, Saxony-Anhalt, Germany. The specimens were found under rocks in freshwater streams and were identified by Clemens Grosser who is an expert taxonomist. No sampling permits were required. Specimens were snap-frozen in liquid nitrogen and were stored in freezers at −75 °C inside cryo-tubes until DNA extraction.

Vouchering information

For vouchering, adults were relaxed in 10% ethanol and massaged to keep them stretched. Afterwards, 70% ethanol was increasingly added until fully replaced. A proxy-voucher (adult individual) has been deposited in the University of Vienna's Zoological collection with ID UVZC_EV3497.

Data availability

Placobdella costata and the related genomic study were assigned to Tree of Life ID (ToLID) 'wrPlaCost' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB76610. The sample information is available at the following BioSample accessions: SAMEA115369554, SAMEA115369555, and SAMEA115369556. The genome assembly is accessible from ENA under accession number GCA_964276725.1 and the annotated genome is available through the Ensembl website (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX13400259, ERX12864234, and ERX12635468. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Placobdella_costata/

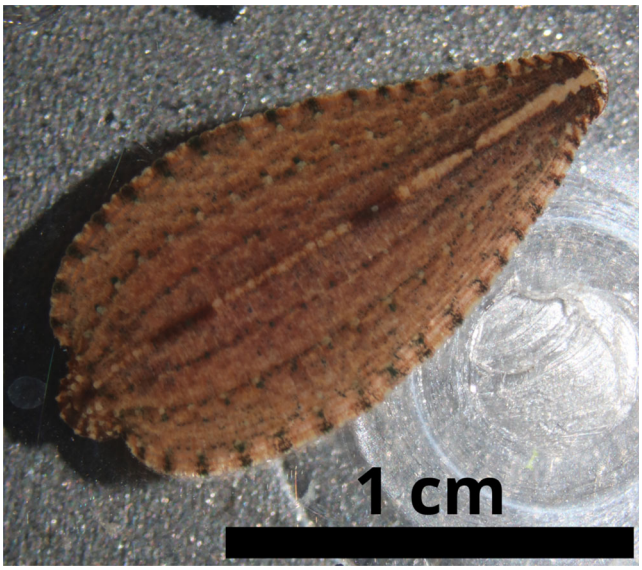

Figure 1. Image of a contracted adult *Placobdella costata*. Photographed specimen was taken from the same population and sampling event as the individual sequenced. Photo credit: Alejandro Manzano-Marín.

wrPlaCost1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/60953.

Genetic Information

The estimated genome size, based on ancestral taxa, is 0.59 Gb. This corresponds to a diploid genome with a haploid number of 9 chromosomes (2n = 18). All information for this species was retrieved from Genomes on a Tree (Challis et al., 2023).

DNA/RNA processing

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.*, 2023a). Sample wrPlaCost1 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.*, 2023b). HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2025a). 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.*, 2025b). 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.

RNA was extracted from whole organism tissue of wrPlaCost3 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.

To generate Hi-C chromatin conformation data, whole organism tissue of sample wrPlaCost2 was processed by 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.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. The sample had an average fragment of 13.4 kb and a total mass of 360 ng that was processed with the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and gDNA 55 kb BAC analysis kit. Prepared libraries were normalised to 2 nM and 15 µL used for making complexes. For libraries below 2 nM all 10 µL was used for making complexes. Primers were annealed and polymerases were hybridised to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean-up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration, in the range of 200–300 pM, and spiked with a Revio sequencing internal control. Samples were sequenced using the Revio system on a Revio 25 M SMRT cell (Pacific Biosciences, California, USA). 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' instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X instrument. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq X instrument. In total, 54x PacBio and 204x HiC data were sequenced to generate the assembly.

Genome assembly methods

HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. 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 (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence. 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 misjoins 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). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1104.1).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system (Aken et al., 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample: SAMEA115369554 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from UniProt (The UniProt Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known metazoan proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into two categories: protein-coding, and long non-coding. Models that did not overlap protein-coding genes, and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Griffiths-Jones, S., et al., 2006) against the genome, followed by RNAfold analysis. Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari et al., 2018) and passing the results through Infernal (Nawrocki & Eddy, 2013). Summary analysis of the released annotation was performed using the ERGA-BGE Genome Report ANNOT Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1096.1), incorporating tools such as AGAT v1.2, BUSCO v5.8.0, and OMArk v0.3.

Results

Genome assembly

The genome assembly has a total length of 488,196,438 bp in 18 nuclear chromosomal pseudomolecules and the mitochondrial genome (Figures 2 and 3), with a GC content of 31.8%. It features a contig N50 of 1,764,884 bp (L50 = 64) and a scaffold N50 of 31,539,400 bp (L50 = 6). There are 659 gaps, totalling 131.8 kb in cumulative size. The single-copy gene content analysis using the Metazoa (odb10) database with BUSCO (Manni et al., 2021) identified 94.0% completeness (92.5% single-copy and 1.6% duplicated). 81.1% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 49.5 as calculated by Merqury (Rhie et al., 2020).

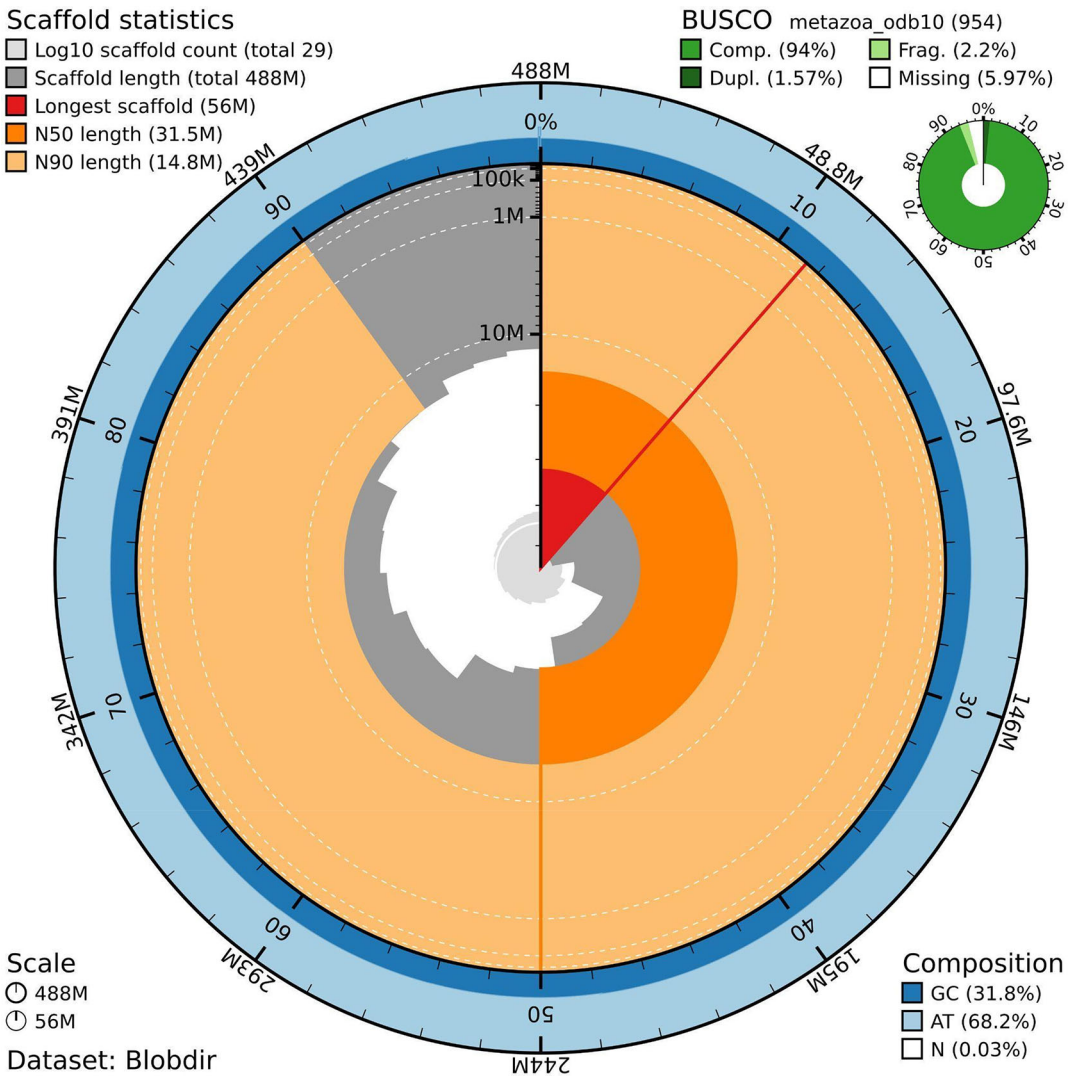

Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 488,196,438 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (56 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (31,539,400 and 14,824,695 bp), respectively. The pale grey spiral shows the cumulative sequence 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 found in the assembled genome from the Metazoa database (odb10) is shown on the top right. SnailPlots are described in Challis and Blaxter (2026).

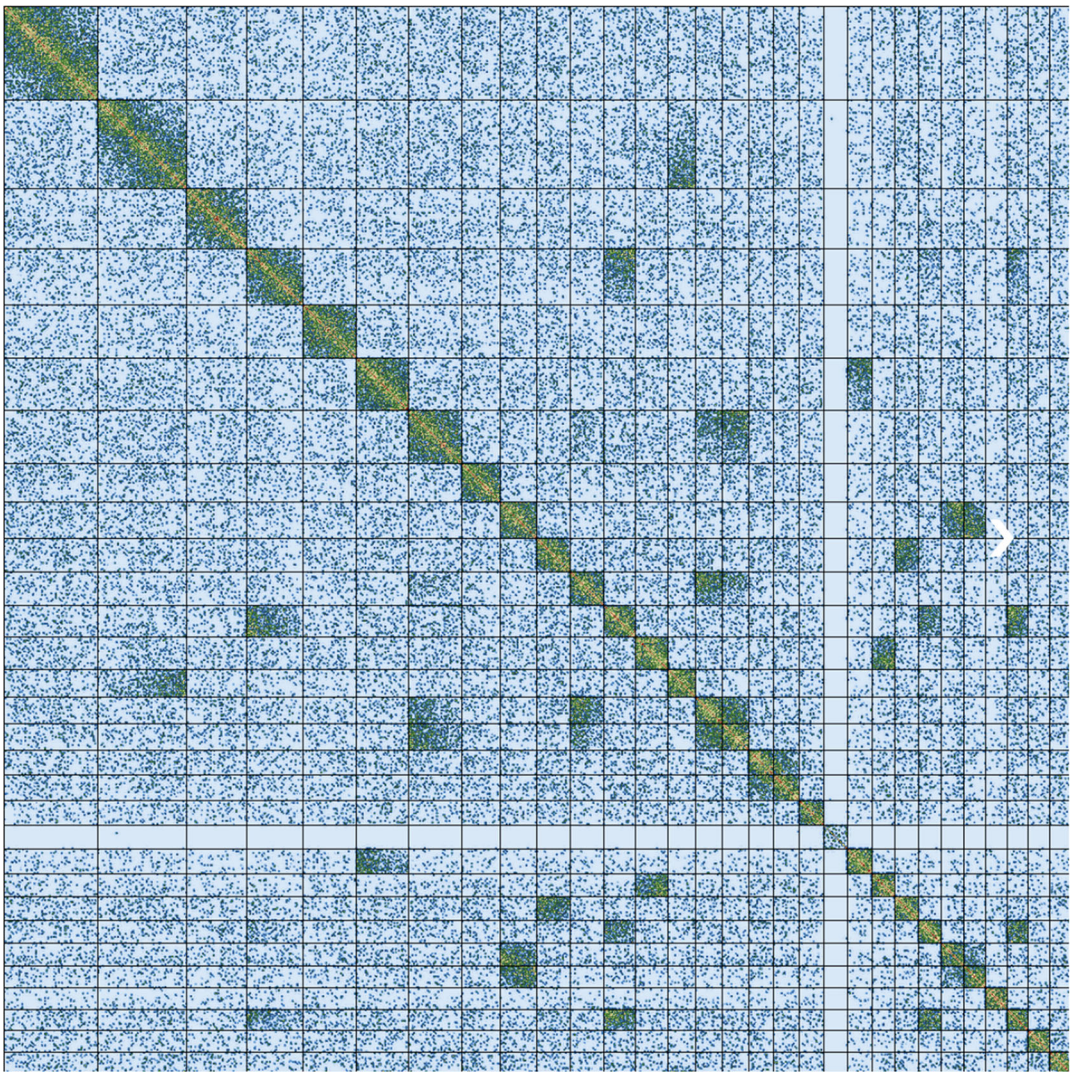

Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting.

Cobiont genome assembly

Three bacterial genomes were also assembled (Table 1), corresponding to an Enterobacteriaceae species (6.41 Mb span, 95% BUSCO complete), a Fulvivirgaceae species (3.5 Mb, 93%) and a Myxococcaceae species (3.96 Mb, 94.6%). A small amount of data deriving from a possible Microscillaceae species (only spanning 0.38 Mb and only 10% BUSCO complete) was also detected.

Genome annotation

The genome annotation consisted of 18,404 protein-coding genes with an associated 26,021 transcripts, in addition to 3,859 non-coding RNA genes of various types (Table 2). Using the longest isoform per transcript, the single-copy gene content analysis using the Metazoa (odb10) database with BUSCO resulted in 93.8% completeness. Using the OMAmer Metazoa-v2.0.0.h5 database for OMArk (Nevers et al., 2025) resulted in 95.7% completeness and 61.6% consistency (Table 3).

Table 1. Cobionts. Information is taken directly from MarkerScan at https://tolqc.cog.sanger.ac.uk/erga-bge/annelids/Placobdella_costata.

specimen	family	original classified reads			original assembly				re-assembly			
		count	(%)	BUSCO	BUSCO	contigs	contig length	number of reads	BUSCO	contigs	contig length	number of reads
wrPlaCost1	Enterobacteriaceae	12′405	0.01	C:97.8% [S:1.4%, D:96.4%], F:0.7%, M:1.5%, n:440	C:94.1% [S:52.7%, D:41.4%], F:1.8%, M:4.1%, n:440	51	5.96 Mb	10′214	C:95.0% [S:81.4%, D:13.6%], F:1.4%, M:3.6%, n:440	32	6.41 Mb	11′221
wrPlaCost1	Fulvivirgaceae	14′314	0.52	C:93.5% [S:0.3%, D:93.2%], F:0.4%, M:6.1%, n:768	C:92.8% [S:78.5%, D:14.3%], F:0.4%, M:6.8%, n:768	15	2.76 Mb	13′963	C:93.0% [S:87.4%, D:5.6%], F:0.4%, M:6.6%, n:768	6	2.47 Mb	37′095
wrPlaCost1	Microscillaceae	1′820	0.07	C:14.0% [S:2.0%, D:12.0%], F:0.7%, M:85.3%, n:768	C:1.6% [S:0.7%, D:0.9%], F:0.0%, M:98.4%, n:768	2	0.05 Mb	50	C:10.6% [S:9.0%, D:1.6%], F:0.3%, M:89.1%, n:768	9	0.38 Mb	2′014
wrPlaCost1	Myxococcaceae	1′040	0.04	C:95.2% [S:0.8%, D:94.4%], F:4.0%, M:0.8%, n:124	C:53.2% [S:53.2%, D:0.0%], F:5.6%, M:41.2%, n:124	3	1.23 Mb	394	C:94.3% [S:91.9%, D:2.4%], F:4.8%, M:0.9%, n:124	8	3.58 Mb	2′630

Open Research Europe 2026, 6:395 Last updated: 06 OCT 2026

Table 2. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
Protein-coding	18,404	26,021	12,799	693	8.4
lncRNA	3,158	3,469	6,921	2	2.4
snRNA	41	41	148	41	1.0
snoRNA	27	27	142	27	1.0
rRNA	423	423	140	423	1.0
tRNA	210	210	76	210	1.0
Other non-coding	2,339	2,339	90	2,339	1.0

Table 3. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (metazoa_odb10) and OMArk (Metazoa-v2.0.0.h5).

	Complete	Single-copy	Duplicated	Fragmented	Missing
BUSCO	895 (93.8%)	881 (92.3%)	14 (1.5%)	31 (3.2%)	28 (3.0%)
OMArk	2,059 (95.7%)	1,726 (80.2%)	333 (15.5%)	-	94 (4.4%)

	Consistent	Inconsistent	Contaminants	Unknown
OMArk	11,345 (61.6%)	2,982 (16.2%)	0.0 (0.0%)	4,077 (22.1%)

Conflict of interest

The authors declare no conflict of interest related to this study. The funding sources had no involvement in the study design, collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. All authors have participated sufficiently in the work to take public responsibility for the content and agree to the submission of this manuscript.

Funder information

Biodiversity Genomics Europe (Grant no. 101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/TAI1450724.

Author contributions

AM-M coordinated the project; CG collected the species; CG identified the species; AM-M sampled and preserved biological material and provided metadata; AsB, RM, TM, THS, and RAO provided sampling and metadata support and management; WSITOLL extracted DNA and RNA under the supervision of CH; WSISO prepared libraries, and performed sequencing; WSI ToL IT performed genome assembly and curation under the supervision of KH, MB, JMDW, and SMcC; FM and LH performed genome annotation; CB generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version.

Acknowledgements

We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de. NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de. NBI Freiburg. We would like to acknowledge the assembly report reviewer Benjamin Istace from Genoscope.

Literature Cited

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database.* 2016; **2016**, **2016**: baw093.
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**: 892–905.
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.** 2023.
Publisher Full Text

Ben Ahmed R, Gajda Ł, Utevsky S, *et al.*: ***Placobdella nabeulensis* sp. nov. (Hirudinea: Glossiphoniidae), a new glossiphoniiform leech from Palearctic North Africa.** *Mol. Biol. Rep.* 2023; **50**(8): 6753–6767.
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.** *Wellcome Open Research.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Cichocka JM, Bielecki A, Jabłońska-Barna I, *et al.*: **Sucking of Human blood by *Placobdella costata* (O. F. Müller, 1846) (Hirudinida: Glossiphoniidae): Case study with notes on body form.** *Ecol. Evol.* 2021; **11**(24): 17593–17603.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Blaxter M: **Snail plots are badges of genome assembly quality.** *G3 Genes|Genomes|Genetics.* 2026; **16**(6): jkag074.
PubMed Abstract | Publisher 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

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V1.** 2023a.
Publisher Full Text

Denton A, Cornwell C, Quail M, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** 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

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA Extraction: Automated MagMax™ mirVana.** *protocols.io.* 2023.
Reference Source

Dvořáková N, Kvičerová J, Papoušek I, *et al.*: **Haemogregarines from western Palaearctic freshwater turtles (genera *Emys*, *Mauremys*) are conspecific with *Haemogregarina stepanowi* Danilewsky, 1885.** *Parasitology.* 2014; **141**(4): 522–530.
PubMed Abstract | Publisher 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

Griffiths-Jones S, Grocock RJ, van Dongen S, *et al.*: **Enright AJ: miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res.* 2006; **34**(suppl_1): D140–D144.
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

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.** 2023.
Publisher Full Text

Kalvari I, Nawrocki EP, Argasinska J, *et al.*: **Non-coding RNA analysis using the Rfam database.** *Curr. Protoc. Bioinformatics.* 2018; **62**(1): e51.
PubMed Abstract | Publisher Full Text | Free 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

Kutschera U, Peter W: **The evolution of parental care in freshwater leeches.** *Theory Biosci.* 2001; **120**(2): 115–137.
Publisher Full Text

Kvist S, Utevsky S, Marrone F, *et al.*: **Extensive sampling sheds light on species-level diversity in Palearctic *Placobdella* (Annelida: Clitellata: Glossiphoniiformes).** *Hydrobiologia.* 2022; **849**(5): 1239–1259.
Publisher Full Text

Manglicmot C, Oceguera-Figueroa A, Kvist S: **Bacterial endosymbionts of *Placobdella* (Annelida: Hirudinea: Glossiphoniidae): phylogeny, genetic distance, and vertical transmission.** *Hydrobiologia.* 2020; **847**(4): 1177–1194.
Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO: Assessing genomic data quality and beyond.** *Current Protocols.* 2021; **1**(12): e323.
PubMed Abstract | Publisher Full Text

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: An atlas of European reference genomes.** *Nature.* 2023; **619**: 252.
Publisher Full Text

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2933–2935.
PubMed Abstract | Publisher Full Text | Free Full Text

Nevers Y, Warwick Vesztrocy A, Rossier V, *et al.*: **Quality assessment of gene repertoire annotations with OMArk.** *Nat. Biotechnol.* 2025; **43**(1): 124–133.
PubMed Abstract | Publisher Full Text | Free Full Text

Oatley G, Denton A, Kovalevskaia A, *et al.*: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.3.** 2025a.
Publisher Full Text

Oatley G, Sampaio F, Denton A, *et al.*: **Sanger Tree of Life fragmented DNA clean up: automated SPRI V.2.** 2025b.
Publisher Full Text

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

Siddall ME, Budinoff RB, Borda E: **Phylogenetic evaluation of systematics and biogeography of the leech family Glossiphoniidae.** *Invertebr. Syst.* 2005; **19**(2): 105–112.
Publisher Full Text

The UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
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**: 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.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* 2019; 314–324.
Publisher 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