Check for updates

DATA NOTE

REVISED ERGA-BGE reference genome of *Hirudo verbana,* a once neglected freshwater haematophagous European medicinal leech

[version 2; peer review: 2 approved]

Alejandro Manzano-Marín[1], Astrid Böhne[2], Rita Monteiro[2], Thomas Marcussen[3,4], Torsten H. Struck[3], Rebekah A. Oomen[3-7], 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[8], Kerstin Howe[8], Mark Blaxter[8], Shane McCarthy[8], Jonathan M.D. Wood[8], Fergal Martin[9], Anna Lazar[9], Leanne Haggerty[9], Chiara Bortoluzzi[10]

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

v2 **First published:** 22 Dec 2025, **5**:395
https://doi.org/10.12688/openreseurope.21672.1
Latest published: 10 Feb 2026, **5**:395
https://doi.org/10.12688/openreseurope.21672.2

Abstract

Hirudo verbana Carena, 1820, commonly known as the southern medicinal leech, is one of several European medicinal leeches, whose full diversity has just recently started to be uncovered. Historically, it has been widely used as a medicinal leech and for centuries it was treated erroneously under the specific name of *Hirudo medicinalis* L. 1758. Recent molecular and taxonomic analyses have revealed subspecific diversity within the morphospecies *H. verbana*. *Hirudo verbana* is a blood-feeding species sucking blood from amphibians, fish, and mammals. It occupies freshwater habitats, typically shallow

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 10 Feb 2026		✓ view
version 1 22 Dec 2025	✓ view	✓ view

1. **Naim Saglam**, Firat University, Elazığ,

ponds and lakes. Studies show that this leech species has a "naturally limited microbiome", suggesting it may serve as a powerful model system for the study of gut microbiota. We expect this chromosome-level assembly of *H. verbana* to serve as a high-quality genomic resource for this most famous leech genus and to serve as a foundation to the study of the diversification and biodiversity of European medicinal leeches, as well as their gut-associated symbionts. The genome of *H. verbana* was assembled into two haplotypes through a phased assembly approach; however, only the primary haplotype was designated as the reference genome for annotation and downstream analyses. The entirety of the primary haplotype was assembled into 14 contiguous chromosomal pseudomolecules, including the mitogenome. This chromosome-level assembly encompasses 0.18 Gb, composed of 277 contigs and 27 scaffolds, with contig and scaffold N50 values of 1.3 Mb and 13.4 Mb, respectively.

Keywords
Hirudo verbana genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, 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.

Turkey

2. **Christian Mueller**, University of Greifswald, Greifswald, Germany

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

Corresponding author: Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: Manzano-Marín A: Investigation, Resources, Writing – Original Draft Preparation, 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; **Struck TH**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Oomen RA**: 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; **Lazar A**: 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. A.M.-M. has received funding from European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie Grant Agreement No. [840270] (LEECHSYMBIO).
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, Böhne A, Monteiro R *et al.* **ERGA-BGE reference genome of *Hirudo verbana,* a once neglected freshwater haematophagous European medicinal leech [version 2; peer review: 2 approved]** Open Research Europe 2026, **5**:395 https://doi.org/10.12688/openreseurope.21672.2

First published: 22 Dec 2025, **5**:395 https://doi.org/10.12688/openreseurope.21672.1

REVISED **Amendments from Version 1**

In this revised version of the Data note, we have addressed all comments from the two reviewers. Major changes include a photograph of the species *Hirudo verbana*, a few additional sentences specifying the sampling permits, and a few additional references to better contextualise our work.

Any further responses from the reviewers can be found at the end of the article

Introduction

Hirudo verbana Carena, 1820, a member of the *Hirudinidae* family (Annelida: Euhirudinea), inhabits freshwater habitats, most commonly shallow ponds and lakes, where it feeds on amphibians, fish, and mammals (Elliott & Kutschera, 2011; Vecchioni *et al.*, 2025) (Figure 1). While it mainly lives underwater, it emerges onto land to lay its spongy cocoons, a reproductive trait shared with other members of the *Hirudinidae* family. *Hirudo* are large cylindrical to dorsoventrally flattened leeches with their whole dorsal surface roughened by small papillae (Nesemann & Neubert, 1999). Their jaws are monostichodont, with salivary papillae on the jaws being absent. The caudal sucker is medium sized and never exceeds the maximum body diameter. There is no furrow on the upper lip of the cranial sucker. The crop has one pair of short caeca per somite.

Hirudo verbana has a characteristic colouration pattern with two broad, reddish, and diffuse paramedian dorsal stripes and a unicoloured greenish to yellow venter with a pair of black marginal stripes (Utevsky & Trontelj, 2005). In general, the external colouration pattern of European *Hirudo* species is a good discriminatory characteristic (Trontelj & Utevsky, 2005).

Historically, *H. verbana* (long treated as a colouration type/form of *H. medicinalis* L. 1758) has been, and still is, widely used and trafficked for medicinal and pseudo-medicinal purposes. The species has reliably been found in Switzerland, Austria, Italy, Slovenia, Croatia, Bosnia and Herzegovina, Serbia, Montenegro, Macedonia, Greece, Hungary, Moldova, Ukraine, Romania, Russian Federation, Slovakia, Spain, Turkey, and Uzbekistan (Arias *et al.*, 2021; Cséfalvay *et al.*, 2017; Popa *et al.*, 2024; Utevsky *et al.*, 2010) and it was first recognised as separate from *H. medicinalis* by Giacinto Carena (Carena, 1820) from specimens collected and sent to him from *Lacus Verbanus* (Italian: *Verbano* or *Lago Maggiore*). However, for over a century, his description of *H. verbana* was neglected and it was not until 1999 that Nesemann and Neubert re-established the species' status (Nesemann & Neubert, 1999; Trontelj & Utevsky, 2005).



Figure 1. Image of a contracted juvenile *Hirudo* verbana. Photographed specimen was taken from the same pupulation and sampling event as the individual sequenced. Photo credit: Alejandro Manzano-Marín.

Genetic studies aimed at uncovering the diversity of the European *Hirudo* spp. have revealed that, unlike other European congeners, this species shows significant intraspecific phylogenetic structure (Trontelj & Utevsky, 2012). Two clades were subsequently recognised: a Western and an Eastern phylogroup. Recently, two further phylogroups have been recognised: an Iberian "typical" and an Iberian "bilineated" (Arias *et al.*, 2021). The latter phylogroup has been described as *Hirudo verbana bilineata* Arias, Surugiu, Carballeira, Popa, Popa & Utevsky, 2021. Thus, it remains likely that unrecognised diversity occurs within the species.

Under laboratory conditions, a leech can feed on two to five times its own mass in one blood meal, which is then digested slowly over many months. Reaching well over 10 cm in length and a mass of several grams, these leeches are an important food source for predators, such as birds and fish. Experiments aimed at discerning differences in biological characteristics of European medicinal leeches have revealed that *H. verbana* has the highest fecundity and juvenile mortality, as well as small juvenile body size when compared to *H. medicialis* and *H. orientalis* (Petrauskiene *et al.*, 2011; Utevsky & Trontelj, 2005).

As other animals, *H. verbana* is host to diverse microbes in its digestive tract. Relevantly, it hosts what has been referred to as a "naturally limited microbiome" consisting of just over a dozen well-defined microbes with a specific localisation. The naturally low diversity microbiome housed by *H. verbana* has led to suggestions of this leech species being a powerful model for the study of microbe-host interactions (Nelson & Graf, 2012).

As of the 20th of June 2025, *H. verbana* is not listed on the IUCN Red List of Threatened Species. However, the species is listed on the Appendix II of CITES, establishing quotas for the export of live or frozen wild-taken individuals, on Annex V of the EU Council directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora, and on several national red lists.

Paulsen *et al.* (2020) have previously published a draft genome for a specimen identified as *H. verbana*. However, the highly fragmented assembly (61,282 contigs of >= 200 bp length, N50 of 8,638 bp) and lack of accompanying RNA-Seq data and a voucher specimen (complicating proper affiliation of the specimen) diminish the value of this genomic resource as reference. Therefore, a high-quality reference genome for *H. verbana* will be of great value to different areas of research. First, it will be used for the study of genomic evolution in *Hirudo*, for which other chromosome-level genome references are currently available or in progress (*H. tianjinensis*

Wang, Meng, Jin, Gao, Tong & Liu, 2022, *Hirudo nipponia* Whitman, 1886, and *H. medicinalis*). Second, it will serve as a starting point to understand and unravel cryptic diversity within the morphospecies *H. verbana*. Third, as with previous genomic investigations of *H. medicinalis* (Babenko *et al.*, 2020; Kvist *et al.*, 2020), this new genome will allow for the exploration of novel putative anticoagulants, which the leech excretes in its saliva to prevent clotting while feeding and which may have medical applications. Finally, as host of digestive-tract symbionts, and as a promising model system for digestive symbioses, a high-quality reference genome for *H. verbana* will contribute to the study of the establishment, maintenance, and crosstalk between the leech host and its symbionts.

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

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly and annotation.

Sample and sampling information

On 11 August 2023, Alejandro Manzano-Marín sampled 9 specimens of *Hirudo verbana* (hermaphrodite monoecious) from an unnamed pond in Vienna by using submerged legs and feet as bait (latitude: 48.23 - 48.25, longitude: 16.26 - 16.34). Species was determined based on the dorsal and ventral colouration patterns following (Trontelj & Utevsky, 2012). The specimens were identified by Alejandro Manzano-Marín in Vienna, Austria. In Austria, no specific regulations control the collection of *H. verbana*; thus permits were only required for the shipment of the specimens from Austria (CITES permits number AT-23-0027) to the United Kingdom (CITES permits number 24GBIMPD10R7G), where the sequencing took place. The specimen selected for sequencing belongs to the so-called Eastern phylogroup and its mitochondrial COI marker is identical to the sequence from a *H. verbana* individual collected from Severynivka in South-western Ukraine (GenBank: JN083798) (Trontelj & Utevsky, 2012). Once collected, the specimens were snap frozen and were kept at -80 °C until DNA extraction.

Vouchering information

For vouchering, adults were relaxed in 10% ethanol and massaged for keeping them stretched. Afterwards, 70% ethanol was increasingly added until fully replaced. For genome and transcriptome sequencing, whole individuals were chopped and snap-frozen in liquid nitrogen. Physical reference materials for the here sequenced specimen have been deposited in the University of Vienna's Zoological collection https://zoologicalcollection.univie.ac.at/ under the accession number UVZC_EV3151.

Tissues (anterior and posterior ends) from the same individual have been deposited in the Biobank of the Leibnitz Institute https://leibniz-lib.de/en/research/research-centres/zmb/bonn-location/biobank.html under voucher IDs ZFMK-TIS-122183 for the anterior end and ZFMK-TIS-122184 for the posterior end.

Data availability

Hirudo verbana and the related genomic study were assigned to Tree of Life ID (ToLID) 'wcHirVerb1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB84141. The sample information is available at the following BioSample accession: SAMEA115178104. The genome assembly is accessible from ENA under accession number GCA_965178065.1 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: ERX13553724 and ERX13574799. Data used to generate the tables, figures and statistics in this report are available at the following repository: https://doi.org/10.5281/zenodo.17831908. 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/Hirudo_verbana/wcHirVerb1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/311461.

Genetic information

The estimated genome size, estimated by Genomes on a Tree (GoaT) (Challis *et al.*, 2023) by ancestral state reconstruction, is 0.22 Gb. This is a diploid genome with a haploid number of 13 chromosomes (2n=26). All information for this species was retrieved from Genomes on a Tree (Challis *et al.*, 2023).

DNA/RNA processing

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Denton *et al.*, 2023b; Howard *et al.*, 2025). The wcHirVerb1 sample was weighed and triaged (Jay *et al.*, 2023) to determine the appropriate extraction protocol. Tissue from the mid-body was homogenised by powermashing using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). 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 following the Megaruptor®3 for LI PacBio protocol (Bates *et al.*, 2023). Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation) (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.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), according to the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech). Prior to sequencing, libraries were normalised to 10 ng/µL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X. In total, 49x genome coverage in HiFi and 536x genome coverage in HiC data were sequenced to generate the assembly.

Genome assembly methods

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) in Hi-C phasing mode, where data were separated into two haplotypes. These haplotypes were then curated to generate a final assembly. The Hi-C reads were aligned to the contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and contigs were scaffolded with YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The resulting 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 oatk (Zhou *et al.*, 2024) as a single circular contig of 18,667 bp and it is included in the released assembly (GCA_965178065.1). 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). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (doi.org/10.48546/workflowhub.workflow.1104.1).

Cobionts

Bacterial cobionts in the raw data were identified using marker ribosomal RNA loci and then reads corresponding to each identified taxon were extracted and assembled using the MarkerScan pipeline (Vancaester & Blaxter, 2024).

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 BioSamples SAMN00113400, SAMN15803289, SAMN08595902, SAMN10389976, and SAMN10388009 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from 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 (Kozomara *et al.*, 2019) against the genome, followed by RNAfold analysis (Gruber *et al.*, 2008). 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 carried out using the ERGA-BGE Genome Report ANNOT Galaxy workflow (10.48546/workflowhub.workflow.1096.1).

Results

Genome assembly

The genome assembly has a total length of 177,769,257 bp in 27 scaffolds including the mitogenome (Figure 2 & Figure 3),

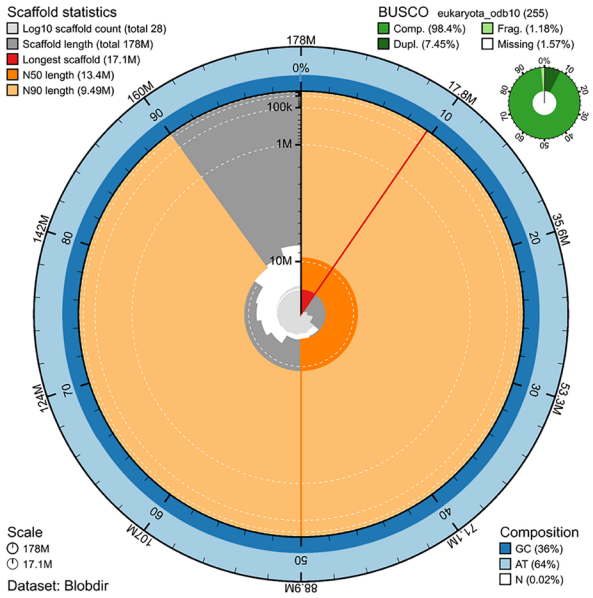

Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 177,769,257 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (17.1 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (13,422,010 and 9,491,130 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 Eukaryota database (odb10) is shown in the top right. The snailplot was generated using the BlobToolKit suite (Challis *et al.*, 2020).

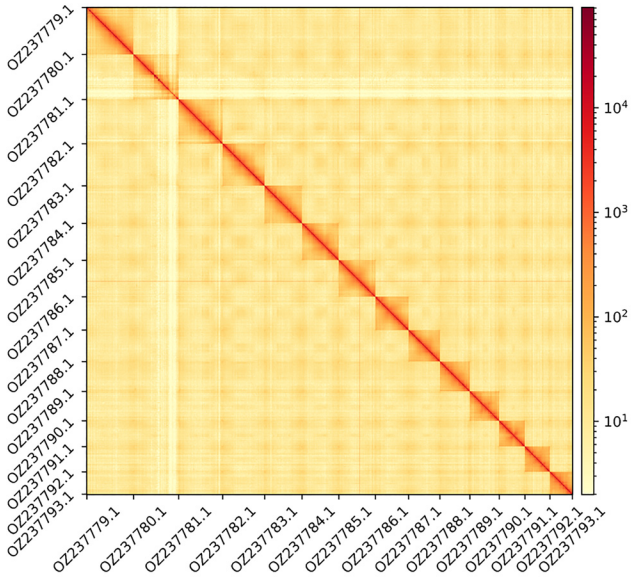

Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting. The x-axis and y-axis show the 14 chromosomes and the mitogenome (GenBank: OZ237793.1). The Hi-C contact map was generated using HiCExplorer (Wolff *et al.*, 2020).

with a GC content of 36%. The assembly has a contig N50 of 1,338,446 bp and L50 of 36 and a scaffold N50 of 13,422,010 bp and L50 of 6. The assembly has a total of 250 gaps, totaling 33.5 kb in cumulative size. The single-copy gene content analysis using the Eukaryota database with BUSCO (Manni *et al.*, 2021) resulted in 98.4% completeness (91.0% single and 7.5% duplicated). 77.1% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 55.1 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 15,195 protein-coding genes with associated 23,608 transcripts, in addition to 8,272 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Eukaryota database with BUSCO resulted in 97.6% completeness. Using the Metazoa-v2.0.0.h5 database for OMArk (Nevers *et al.*, 2025) resulted in 93.3% completeness and 64.1% consistency (Table 2).

Cobionts

Within the raw data for *Hirudo verbana* we identified sequences derived from seven distinct bacterial families. Because we do not, *a priori*, know the relationships between these bacteria and the leech, we use the term "cobionts": they could be mutualist symbionts, members of the microbiome, pathogens or chance associations with environmental organisms. After sorting prokaryote-derived reads into bins (containing from 0.2% to 8% of raw reads) with MarkerScan (Vancaester & Blaxter, 2024) we assembled genomes from these bacterial families (Table 3). Three of the bins had assembled spans (>10 Mb) that suggested that more than one strain or species had been co-assembled; these bins had high BUSCO duplication scores.

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	15,195	23,608	6,092	274	8.4
pseudogene	0.0	0.0	0.0	0.0	0.0
snoRNA	193	193	179	193	1.0
lncRNA	1,536	1,704	2,583	5	2.5
tRNA	5,709	5,709	80	5,709	1.0
snRNA	678	678	155	678	1.0
rRNA	155	155	226	155	1.0
scRNA	1	1	134	1	1.0
Other ncRNA	6,896	32,203	79 – 152	6,896	1.0 – 7.8

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (eukaryota_odb10) and OMArk (Metazoa-v2.0.0.h5).

	Complete	Singular	Duplicated	Fragmented	Missing
BUSCO	249 (97.6%)	0 (0.0%)	249 (97.6%)	3 (1.2%)	3 (1.2%)
OMArk	2,010 (93.3%)	352 (16.3%)	1,658 (77.0%)	-	143 (6.7%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	19,492 (64.1%)	3,024 (10.0%)	0 (0.00%)	7,874 (25.9%)	

Table 3. Cobionts. Information is taken directly from MarkerScan at https://tolqc.cog.sanger.ac.uk/erga-bge/annelids/Hirudo_verbana.

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
wcHirVerb1	Chitinophagaceae	4,553	0.38	C:99.2%[S:1.6%,D:97.6%],F:0.0%,M:0.8%,n:124	C:98.3%[S:92.7%,D:5.6%],F:0.8%,M:0.9%,n:124	3	3.33Mb	2,843	C:99.2%[S:62.1%,D:37.1%],F:0.0%,M:0.8%,n:124	24	4.53Mb	5,130
wcHirVerb1	Comamonadaceae	50,786	4.21	C:91.9%[S:0.9%,D:91.0%],F:1.6%,M:6.5%,n:688	C:32.1%[S:19.0%,D:13.1%],F:0.9%,M:67.0%,n:688	42	1.65Mb	4,219	C:91.8%[S:3.1%,D:88.7%],F:1.2%,M:7.0%,n:688	64	26.16Mb	116,895
wcHirVerb1	Myxococcaceae	3,696	0.31	C:93.5%[S:4.8%,D:88.7%],F:5.6%,M:0.9%,n:124	C:18.5%[S:15.3%,D:3.2%],F:3.2%,M:78.3%,n:124	14	0.99Mb	283	C:91.1%[S:41.9%,D:49.2%],F:7.3%,M:1.6%,n:124	6	4.99Mb	9,991
wcHirVerb1	Phyllobacteriaceae	11,363	0.94	C:100.0%[S:1.2%,D:98.8%],F:0.0%,M:0.0%,n:432	C:99.1%[S:22.5%,D:76.6%],F:0.2%,M:0.7%,n:432	50	6.96Mb	6,512	C:99.7%[S:23.1%,D:76.6%],F:0.0%,M:0.3%,n:432	72	12.59Mb	20,807
wcHirVerb1	Pseudobdellovibrionaceae	473	0.04	C:94.3%[S:0.8%,D:93.5%],F:2.4%,M:3.3%,n:124	C:0.8%[S:0.8%,D:0.0%],F:0.0%,M:99.2%,n:124	1	0.04Mb	6	C:92.7%[S:87.1%,D:5.6%],F:3.2%,M:4.1%,n:124	4	2.44Mb	5,111
wcHirVerb1	Rikenellaceae	1,152	0.1	C:94.8%[S:0.2%,D:94.6%],F:0.9%,M:4.3%,n:541	C:5.7%[S:5.7%,D:0.0%],F:0.4%,M:93.9%,n:541	3	0.18Mb	40	C:94.9%[S:47.9%,D:47.0%],F:0.7%,M:4.4%,n:541	13	5.03Mb	5,804
wcHirVerb1	Sphingobacteriaceae	15,660	1.3	C:92.4%[S:0.1%,D:92.3%],F:0.5%,M:7.1%,n:1068	C:87.9%[S:0.2%,D:87.7%],F:0.3%,M:11.8%,n:1068	42	11.49Mb	12,985	C:92.3%[S:4.1%,D:88.2%],F:0.5%,M:7.2%,n:1068	11	17.41Mb	30,309

Open Research Europe 2026, 5:395 Last updated: 05 OCT 2026

Data availability statement

The underlying data has been deposited in the European Nucleotide Archive (ENA), accession number PRJEB84141: https://www.ebi.ac.uk/ena/browser/view/PRJEB84141, and Ensembl, accession number GCA_965178065.1: https://ftp.ebi.ac.uk/pub/ensemblorganisms/Hirudo_verbana/GCA_965178065.1/. Data used to generate the tables, figures and statistics in this report are available at the following repository: https://doi.org/10.5281/zenodo.17831908 (Manzano-Marín *et al.*, 2025). All data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Author contributions

AM-M collected the species; AM-M identified the species; AM-M sampled and preserved biological material and provided metadata; AsB, RM, TM, THS, and RAO provided support in sampling, shipping of biological material, metadata collection, 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, AL, 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

The authors thank Peter Trontelj (University of Ljubljana) for valuable comments reviewing the manuscript. 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 reviewer, Lola Demirdjian from Genoscope.

References

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**; baw093.
PubMed Abstract | Publisher Full Text | Free Full Text

Arias A, Surugiu V, Carballeira R, *et al.*: **Unravelling the extent of diversity within the Iberian medicinal leeches (Hirudinea: *Hirudo*) using molecules and morphology.** *Biology (Basel).* 2021; **10**(4): 315.
PubMed Abstract | Publisher Full Text | Free Full Text

Babenko VV, Podgorny OV, Manuvera VA, *et al.*: **Draft genome sequences of *Hirudo medicinalis* and salivary transcriptome of three closely related medicinal leeches.** *BMC Genomics.* 2020; **21**(1): 331.
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

Carena H: **Monographie du genre *Hirudo*, ou description des espèces de sangsues qui se trouvent ou qui sont en usage en Piémont; avec des observations sur la génération, et sur d'autres points de l'histoire naturelle de quelques unes de ces espèces.** *Memorie della Reale Accademia delle Scienze di Torino.* 1820; **25**: 273–316.
Reference Source

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

Cséfalvay R, Janák M, Immerová B: **First reliable records of *Hirudo verbana* Carena, 1820 (Annelida: Hirudinida) from Slovakia and notes on its syntopy with *Hirudo medicinalis* Linnaeus, 1758.** *Folia Faun Slovaca.* 2017; **22**: 63–66.
Reference Source

Denton A, Cornwell C, Quail M, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash**. 2023a.
Publisher Full Text

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

Elliott JM, Kutschera U: **Medicinal leeches: historical use, ecology, genetics and conservation.** *Freshw Rev.* 2011; **4**(1): 21–41.
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

Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA websuite.** *Nucleic Acids Res.* 2008; **36**(Web Server issue): W70–W74.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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; 2025-04.
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**. 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

Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Res.* 2019; **47**(D1): D155–D162.
PubMed Abstract | Publisher Full Text | Free Full Text

Kvist S, Manzano-Marín A, de Carle D, *et al.*: **Draft genome of the European medicinal leech *Hirudo medicinalis* (Annelida, Clitellata, Hirudiniformes) with emphasis on anticoagulants.** *Sci Rep.* 2020; **10**(1): 9885.
PubMed Abstract | Publisher Full Text | Free 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**

Manzano-Marín A, Böhne A, Monteiro R, *et al.*: **ERGA-BGE reference genome of Hirudo verbana, a once neglected freshwater haematophagous European medicinal leech (Version v1).** [Data set], Zenodo. 2025.
http://www.doi.org/10.5281/zenodo.17831909

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: an atlas of European reference genomes.** *Nature.* 2023; **619**(7969): 252.
PubMed Abstract | **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**

Nelson MC, Graf J: **Bacterial symbioses of the medicinal leech Hirudo verbana.** *Gut Microbes.* 2012; **3**(4): 322–331.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nesemann H, Neubert E: **Bd. 6/2: Annelida, Clitellata: Branchiobdellida, Acanthobdellea, Hirudinea**. Spektrum Heidelberg, 1999; 187.
Reference Source

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

Paulsen RT, Agany DD, Petersen J, *et al.*: **A draft genome for *Hirudo verbana*, the Medicinal leech.** *BioRxiv.* 2020; 2020-12.
Publisher Full Text

Petrauskiene L, Utevska O, Utevsky S: **Reproductive biology and ecological strategies of three species of medicinal leeches (genus *Hirudo*).** *J Nat Hist.* 2011; **45**(11–12): 737–747.
Publisher Full Text

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

Popa OP, Ștefan A, Baltag EȘ, *et al.*: **High genetic diversity of *Hirudo verbana* Carena, 1820 (Annelida: Hirudinea: Hirudinidae) in Romania confirms that the balkans are refugia within refugium.** *Diversity.* 2024; **16**(12): 726.
Publisher 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**

Trontelj P, Utevsky SY: **Celebrity with a neglected taxonomy: molecular systematics of the medicinal leech (genus *Hirudo*).** *Mol Phylogenet Evol.* 2005; **34**(3): 616–624.
PubMed Abstract | **Publisher Full Text**

Trontelj P, Utevsky SY: **Phylogeny and phylogeography of medicinal leeches (genus *Hirudo*): fast dispersal and shallow genetic structure.** *Mol Phylogenet Evol.* 2012; **63**(2): 475–485.
PubMed Abstract | **Publisher Full Text**

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**

Utevsky SY, Trontelj P: **A new species of the medicinal leech (Oligochaeta, Hirudinida, *Hirudo*) from Transcaucasia and an identification key for the genus *Hirudo*.** *Parasitol Res.* 2005; **98**(1): 61–66.
PubMed Abstract | **Publisher Full Text**

Utevsky S, Zagmajster M, Atemasov A, *et al.*: **Distribution and status of medicinal leeches (genus *Hirudo*) in the Western Palaearctic: anthropogenic, ecological, or historical effects?** *Aquatic Conservation: Marine and Freshwater Ecosystems.* 2010; **20**(2): 198–210.
Publisher Full Text

Vancaester E, Blaxter ML: **MarkerScan: separation and assembly of cobionts sequenced alongside target species in biodiversity genomics projects [version 1; peer review: 1 approved, 3 approved with reservations].** *Wellcome Open Res.* 2024; **9**: 33.
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

Vecchioni L, Bellucci D, Novaga R, *et al.*: **Further evidence of the southern Mediterranean medicinal leech *Hirudo verbana* (Annelida, Hirudinea) feeding on fish, with a review of the use of fish hosts by *Hirudo* spp.: *Hirudo* verbana feeding on fish.** *Limnetica.* 2025; **44**(1): 101–111.
Publisher Full Text

Wang H, Meng FM, Jin SJ, *et al.*: **A new species of medicinal leech in the genus *Hirudo* Linnaeus, 1758 (*Hirudiniformes*, *Hirudinidae*) from Tianjin City, China.** *ZooKeys.* 2022; **1095**: 83–96.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Whitman CO: **The leeches of Japan.** *Q J Microsc Sci.* 1886; **26**(103): 317–416.

Wolff J, Rabbani L, Gilsbach R, *et al.*: **Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization.** *Nucleic Acids Res.* 2020; **48**(W1): W177–W184.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: a de novo assembly tool for complex plant organelle genomes.** *bioRxiv.* 2024; 2024–10.
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**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 11 February 2026

https://doi.org/10.21956/openreseurope.24798.r69390

© **2026 Mueller C.** This is an open access peer review report 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.

✔ **Christian Mueller**
University of Greifswald, Greifswald, Germany

I do not have any additional comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: animal physiology; molecular biology; leech biology and physiology; anticoagulants

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.

Version 1

Reviewer Report 20 January 2026

https://doi.org/10.21956/openreseurope.23444.r67473

© **2026 Mueller C.** This is an open access peer review report 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.

✔ **Christian Mueller**
[1] University of Greifswald, Greifswald, Germany
[2] University of Greifswald, Greifswald, Germany

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 *Hirudo verbana*, one of the European medicinal leech species. In addition the genomes of a total of seven different bacterial cobionts of *Hirudo verbana* were determined as well.

General Impression
The manuscript is part of the European Reference Genome Atlas (ERGA) initiative in general and the Biodiversity Genomics Europe (BGE) project in particular. It addresses a topic that is of great interest for biologist and ecologists on one hand and for physicians and non-medical practitioners on the other hand, 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 an increasing number of publications on complete genomes of different hematophagous and non-hematophagous leech species.
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 publication. There are only a few minor remarks that need to be addressed.

Minor remarks:
1) The statement on the distribution of *H. verbana* across Europe dates back to Utevsky et al. 2010, but in the meantime the species was reliably described also in Slovakia (Cséfalvay et al. 2017; Folia faunistica Slovaca 22:63-66) and in Romania (Popa et al. 2024; Diversity 16, 726. https://doi.org/10.3390/d16120726). Please include the new references.
On the other hand, the "reliable" descriptions of *H. verbana* in Germany are very likely artifacts and the result of accidental or intended releases of leeches to the wild, but the species does not belong to the natural fauna of Germany. The presence of stable reproductive populations of *H. verbana* in Germany is hence highly questionable.
2) Draft genomes of *H. medicinalis* were published in 2020 both by Kvist et al. and by Babenko et al. (Babenko et al. 2020; BMC Genomics 21:331; https://doi.org/10.1186/s12864-020-6748-0). Please acknowledge and cite also the contribution of the Russian colleagues. In addition, a draft genome of *H. verbana* was generated and a respective publication deposited in BioRxiv (Paulsen et al. 2020, https://doi.org/10.1101/2020.12.08.416024), but for unknown reasons never officially published. However, I recommend to also cite the contribution of the US colleagues.
3) I recommend to include a photograph of *H. verbana*, just as an eye-catcher.
4) The manuscript lacks any further analyses like the identification of putative anticoagulants. At least a short comparison with the genomes of Asian medicinal leeches would, in my opinion, significantly enhance the strength and the attractiveness of the manuscript. Please consider.

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: animal physiology; molecular biology; leech biology and physiology; anticoagulants

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.

Author Response 20 Jan 2026

Chiara Bortoluzzi

We thank the reviewer for their positive feedback and comments. All comments are addressed in the revised manuscript as follow:
1. You are absolutely right. Slovakia and Romania have been added to the list of countries and references are now added in the main text. “The species has reliably been found in Switzerland, Austria,Italy, Slovenia, Croatia, Bosnia and Herzegovina, Serbia, Montenegro, Macedonia, Greece, Hungary, Moldova, Ukraine, Romania, Russian Federation, Slovakia, Spain, Turkey, and Uzbekistan (Arias et al., 2021; Utevsky et al., 2010; Cséfalvay et al. 2017; Popa et al. 2024)” On the other hand, the "reliable" descriptions of H. verbana in Germany are very likely artifacts and the result of accidental or intended releases of leeches to the wild, but the species does not belong to the natural fauna of Germany. The presence of stable reproductive populations of H. verbana in Germany is hence highly questionable. With the term ‘reliable’, we wanted to convey rather the message of bona fide descriptions of H. verbana (taxonomically speaking), and did not intend to imply that there are in fact “reproductive populations” of H. verbana there. For example, even in Austria, H. medicinalis is what one would expect, but the sequenced population can reliably be traced to an artificial historical introduction that has developed into a very isolated stable reproductive population. We have however decided to remove Germany from the list, as we recognise the statement could in fact be read as such.
2. You are right. In its original text, we thought more of the highlight of the use of these genomes to identify putative novel anticoagulants (the focus of Kvist et. al 2020). Babenko et al. 2020 rather focused on draft transcriptomes to extract this information. This is why we originally omitted the citation. We have nevertheless added it now. We are aware of the previous draft genome pre-print of a specimen identified as H. verbana. We did not include this citation as we do not delve into previous genomic investigations of H. verbana. We have thus added a small mention to it.
3. We added a photograph of H. verbana.
4. We did not perform any further analyses on the identification of putative anticoagulants, as this paper is thought of as a genome report to make the reference genome and associated resources avaiable to the broader community. Thus, such analysis would go beyond the scope of this report. Other researchers interested in this area are welcome to use this genomic resource for this purpose.

Competing Interests: No competing interests were disclosed.

Reviewer Report 12 January 2026

https://doi.org/10.21956/openreseurope.23444.r66970

© 2026 Saglam N. This is an open access peer review report 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.

Naim Saglam
[1] Firat University, Elazığ, Turkey
[2] Firat University, Elazığ, Turkey

This study follows the highest standards of modern genomics. The presentation of a "reference genome" for a species (*Hirudo verbana*) with taxonomic complexity and medicinal importance enhances the value of the article. However, it is important to consider the following points:

1. While the genome is stated as 0.18 Gb in the abstract, the estimated size is given as 0.22 Gb in the "Genetic Information" section. A brief explanation of whether this difference is due to repetitive sequences or regions lost during compilation would be helpful.
2. The study was conducted on samples collected from an unnamed pond in Vienna. Providing GPS coordinates for the "unnamed pond" in Vienna, if possible, would be beneficial for reproducibility.
3. *Hirudo* species are generally included in CITES Appendix II. Clarifying the exemption status regarding the collection of this species in Austria and providing an explanation of whether the necessary permits for this research were obtained would be helpful.
4. The inclusion of the genomes of seven different bacterial families, along with the leech genome, in this article enhances the biological value of the study. However, a brief discussion of the relationships between the bacterial families in Table 3 and other bacteria found in the leech's known gut microbiota would make the text more informative.
5. The article appears to lack a discussion section. Considering previous molecular studies conducted on *Hirudo* species, a discussion section should be included to enhance the article's clarity and informativeness.

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: Leeches, Molecular evolution, RN/DNA, Transcriptomics and Fish diseases.

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.

Author Response 20 Jan 2026

Chiara Bortoluzzi

We thank the reviewer for their positive feedback and additional comments. All comments have been addressed in the new revised manuscript as follow:
1. The difference in genome size between what reported in the abstract and what reported in the 'Genetic information' section is due to the fact the while the first value (0.18) is the actual size of the sequenced genome, the second value (0.22) is an estimate obtained from ancestral taxa. This estimate is used prior to sequencing to help calculate sequencing effort (e.g. library prep reagents and number of Revio flow cells) and it is known that the value might not be exact, but it is an educated guess. The assumption that a species might have its genome size within the range described for its taxonomic group is the basis for GoaT estimates: the median of genome sizes are computed from taxa with direct measurements and propagated to the closest common ancestor, then back to fill the missing data. More information on how GoaT calculates ancestral values to propagate can be found in https://wellcomeopenresearch.org/articles/8-24/v1. K-mer genome sizes are part of our pipelines and are used routinely to assess the need for top-up sequencing during the process of producing a genome. The final assessment of a sequence-based genome size is its assembly span, which, together with K-mer inferred genome size measurements, are not the definitive amount of DNA in cells, but a reference to the size of chromosome assemblies. We reformulated this part to make it more clear.
2. We now provide the general GPS coordinates in the main text. However, the exact GPS coordinates were omitted in order to protect the collection site from exploitation, as this is a problem when reporting collection sites for European medicinal leeches.
3. There are currently only two Hirudo species currently listed in CITES Appendix II: Hirudo medicinalis, and Hirudo verbana. Sometimes Hirudo spp. are treated as a whole by CITES agents as they cannot generally assess the specific species affiliation. The appendix II, as it relates to these two species, does not relate to the collection of the species in the wild, but rather it controls its trade, "Appendix II lists species that are not necessarily now threatened with extinction but that may become so unless trade is closely controlled. It also includes so-called "look-alike species", i.e. species whose specimens in trade look like those of species listed for conservation reasons (see Article II, paragraph 2 of the Convention). International trade in specimens of Appendix-II species may be authorized by the granting of an export permit or re-export certificate. No import permit is necessary for these species under CITES (although a permit is needed in some countries that have taken stricter measures than CITES requires). Permits or certificates should only be granted if the relevant authorities are satisfied that certain conditions are met, above all that trade will not be detrimental to the survival of the species in the wild. (See Article IV of the Convention)" (https://cites.org/eng/app/index.php). And "2. Appendix II shall include: (a) all species which although not necessarily now threatened with extinction may become so unless trade in specimens of such species is subject to strict regulation in order to avoid utilization incompatible with their survival; and (b) other species which must be subject to regulation

in order that trade in specimens of certain species referred to in sub-paragraph (a) of this paragraph may be brought under effective control. ” (https://cites.org/eng/disc/text.php#II). Therefore, its collection is not controlled, but rather its trade, establishing quotas per country (https://www.speciesplus.net/species#/taxon_concepts/6167/legal). See also Commission Reg. (EU) 2023/966 of 15 May 2023 (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023R0966). In fact, Hirudo verbana is included as it is considered a "look-alike species" to its congener H. medicinalis. In Austria, no specific regulations control its collection, as this was also not done in any recognised natural reserve and species is not recognised as one protected from its collection. Thus, permits were only required for its shipping from Austria and into the UK. These CITES permits were granted with numbers AT-23–0027 (Austria, export) and 24GBIMPD10R7G (UK, import). We added a few sentences about this in the main text in the section ‘Sample and sampling information’.
4. the bacterial MAGs reported were done so in the interest of reporting this finding, and given previous reports of the presence of some of these taxa as members of the microbiome of H. verbana. However, these have not been further analysed, as they were not the focus of this genome data report and are being currently fully analysed for a separate work focusing on the genomics, diversity, metabolic potential, and evolution of these entities.
5. We are aware that the article lacks a discussion section, but this is due to the fact that this is a Data Note, whose scope is to present data, in our case the genome assembly of Hirudo verbena. Unfortunately, we cannot add such a section due to a constraint in the article format.

Competing Interests: No competing interests were disclosed.