Check for updates

DATA NOTE

ERGA-BGE genome of *Anisus vorticulus* (Troschel, 1834): the Lesser Ramshorn Snail, an endangered freshwater snail protected under the EU's Habitats and Species Directive

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

Bernhard Hausdorf[1], Elicio Tapia[1], Rita Monteiro[2], Astrid Böhne[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, Leanne Haggerty[8], Fergal Martin[8], Thomas Brown[9,10]

[1]Leibniz Institute for the Analysis of Biodiversity Change, Museum of Nature Hamburg, Hamburg, Germany
[2]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Bonn, 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]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK
[9]Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany
[10]Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Berlin, Germany

v1 **First published:** 23 Jul 2025, **5**:190
https://doi.org/10.12688/openreseurope.20499.1
Latest published: 23 Jul 2025, **5**:190
https://doi.org/10.12688/openreseurope.20499.1

Abstract

We have generated a high-quality reference genome for *Anisus vorticulus* to support monitoring of the genetic diversity of populations of this declining freshwater species. The chromosome-level genome assembly encompasses 1.04 Gb, composed of 835 contigs and 77 scaffolds, with contig and scaffold N50 values of 2.35 Mb and 59.2 Mb, respectively. The genome was assembled into 18 chromosomal pseudomolecules.

Keywords

Anisus vorticulus, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Planorbidae, Lesser Ramshorn Snail

Open Peer Review

Approval Status ✓ ?

	1	2
version 1 23 Jul 2025	✓ view	? view

1. **Takumi Saito**, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

2. **Maurine Neiman**, University of Iowa, Iowa City, USA

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

This article is included in the Horizon Europe gateway.

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

Corresponding author: Thomas Brown (brown@izw-berlin.de)

Author roles: Hausdorf B: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Tapia E**: Investigation, Resources, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Böhne A**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Struck TH**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Oomen RA**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Brown T**: Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project received funding from Horizon Europe under the Biodiversity, Circular Economy and Environment (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. Work at the Wellcome Sanger Institute is additionally supported by Wellcome Trust Grants 206194 and 218328.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Hausdorf B *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: Hausdorf B, Tapia E, Monteiro R *et al.* **ERGA-BGE genome of *Anisus vorticulus* (Troschel, 1834): the Lesser Ramshorn Snail, an endangered freshwater snail protected under the EU's Habitats and Species Directive [version 1; peer review: 1 approved, 1 approved with reservations]** Open Research Europe 2025, **5**:190
https://doi.org/10.12688/openreseurope.20499.1

First published: 23 Jul 2025, **5**:190 https://doi.org/10.12688/openreseurope.20499.1

Introduction

Anisus vorticulus, a member of the Planorbidae, is a freshwater snail, characterized by a very small, tightly coiled, flat, thin shell (maximum diameter 5 mm) with a peripheral blunt keel (Glöer, 2019). It is an annual habitat specialist, living only in shallow, sunlit, unpolluted, moderately well-vegetated, usually calcareous waters in floodplains and fens, such as drainage ditches, pools, streams, slow rivers and oxbow lakes (Glöer & Groh, 2007; Van Damme, 2012). It is spread through the Western Palaearctic (Europe, Turkey, Kazakhstan and western Siberia; Glöer & Groh, 2007; Van Damme, 2012). As a specialist for clear waters in sunny habitats (Van Damme, 2012), *A. vorticulus* is an indicator species for such pristine habitats.

Anisus vorticulus has been listed in the EU's Habitats and Species Directive (Council Directive 92/43/EEC on the Conservation of natural habitats and of wild fauna and flora) as a species of community interest, requiring special areas for conservation (Annex II) and strict protection (Annex IV). Furthermore, it is listed on Red Lists of several countries as critically endangered (e.g. Czech Republic; Germany) or endangered (e.g. Austria; Switzerland). In the global assessment of the IUCN it has been assessed as Data Deficient as population declines have been observed throughout its global range, but there is a lack of population data and knowledge of threats across the eastern extent of its distribution (Van Damme, 2012).

As *A. vorticulus* is listed in Annex II of EU's Habitats and Species Directive, special areas of conservation must be designated for the species. However, most populations throughout the global range of *A. vorticulus* are declining (Van Damme, 2012) and often have low densities between 1–50 individuals per m^2 (Glöer & Groh, 2007). Therefore, it is important to assess which populations are genetically diverse and have the best potential to survive as a basis for the decision making, on which populations conservation efforts should be focused, and which populations are most promising as sources for reintroduction. A reference genome is a valuable basis for monitoring the genetic diversity of populations.

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

Bernhard Hausdorf and Elicio Tapia sampled 20 specimens of hermaphroditic, monoecious *Anisus vorticulus*, identified by Bernhard Hausdorf based on morphology using Glöer (2019), from a ditch southeast of Gleisdreick, Kirchwerder, Bergedorf, Hamburg, Germany on 17 October 2023. Sampling was performed under a special permit under §§ 44, 45 des Gesetzes über Naturschutz und Landschaftspflege (BNatSchG) issued by the Amt für Naturschutz und Grünplanung, Behörde für Umwelt, Klima, Energie und Agrarwirtschaft, Hamburg, Germany to Bernhard Hausdorf on 26.05.2023 (valid until 31.12.2023). Sampling was performed using a hand sieve.

Specimens were euthanized as quickly as possible by submerging in liquid nitrogen and subsequently preserved using liquid nitrogen. Until DNA extraction, samples were preserved at −80°C.

Vouchering information

Physical proxy vouchers from the same locality as the here sequenced specimen have been deposited in the Zoological Museum Hamburg of the Leibniz Institute for the Leibniz Institute for the Analysis of Biodiversity Change under the accession number ZMH-MOL-0159875.

Frozen reference tissue material of whole proxy vouchers from the same locality as the here sequenced specimen is available at the Biobank Bonn of the Leibniz Institute for the Analysis of Biodiversity Change under the voucher IDs ZFMK-TIS-80688, ZFMK-TIS-80689 and ZFMK-TIS-80690.

Genetic information

The estimated genome size, based on ancestral taxa, is 1.31 Gb. This is a diploid genome with a haploid number of 18 chromosomes (2n=36). 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.*, 2023b). The xgAniVori1 sample was prepared for DNA extraction by weighing it on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Oatley *et al.*, 2023b). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from whole organism tissue of xgAniVori3 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana 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.

Tissue from the xgAniVori2 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. Tissue

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

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core.

At the minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (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 55kb BAC analysis kit.

Prepared libraries were normalised to 2nM and 15µL used for making complexes. For libraries below 2nM all 10uL 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, between 200 -300pM, and spiked with a Revio sequencing internal control. Samples were sequenced using the Revio system on Revio 25M SMRT cells (Pacific Biosciences, California, USA). The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

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

Genome assembly methods

The HiFi reads were first assembled using Hifiasm (version 0.19.8-r603) (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 (version 1.2a.2) (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 (Version 3) (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 and is included in the released assembly (GCA_964264155).

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 (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: SAMEA114640077 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 (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 (https://doi.org/10.48546/workflowhub.workflow.1096.1).

Results

Genome assembly

The genome assembly has a total length of 1,043,037,472 bp in 77 scaffolds, with 99.6% assigned to 18 chromosomes and the mitogenome (Figure 1 & Figure 2), with a GC content of 35.28%. The assembly has a contig N50 of 2.35 Mb and L50 of 130 and a scaffold N50 of 59.2 Mb and L50 of 8. The assembly has a total of 758 gaps, totaling 151.6 kb in cumulative size. The single-copy gene content analysis using the Mollusca database with BUSCO (Manni *et al.*, 2021) resulted in 94.0% completeness (93.2% single and 0.8% duplicated). 92.3% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 55.6 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 19,091 protein-coding genes with an associated 30,345 transcripts, in addition to 5,871 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Metazoa database with BUSCO resulted in 97.8% completeness. Using the OMAmer Lophotrochozoa database for OMArk (Nevers *et al.*, 2024) resulted in 94.75% completeness and 57.45% consistency (Table 2).

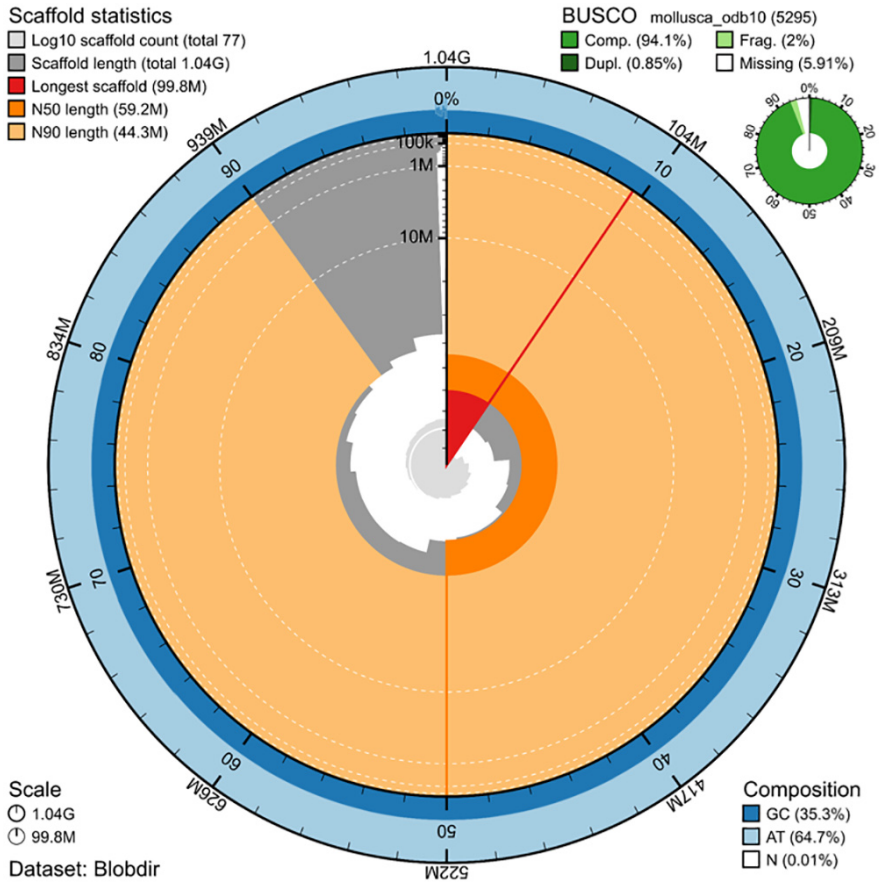

Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,043,037,472 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (99,839,100 bp, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (59,244,224 and 44,347,986 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 Mollusca database (odb10) is shown in the top right. The plot was generated using BlobToolKit (Challis *et al.*, 2020). An interactive version of this plot is available at https://blobtoolkit.genomehubs.org/view/Anisus%20vorticulus/dataset/GCA_964264155.1/snail.

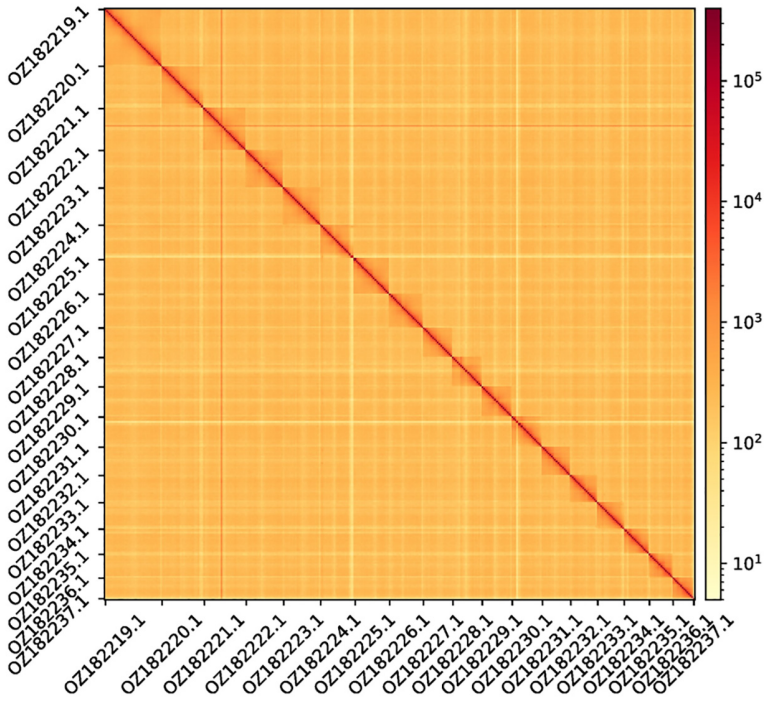

Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries with the pseudochromosomal molecules labelled on the axes. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 25 kb bin size for plotting.

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	19,091	30,345	21,916	374	9.6
lncRNA	3,745	4,370	7,096	2	2.7
scRNA	1	1	132	1	1
snoRNA	83	83	124	83	1
snRNA	194	194	141	194	1
rRNA	619	619	900	619	1
tRNA	410	410	75	410	1

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (metazoa_odb10) and OMArk (Lophotrochozoa).

	Complete	Singular	Duplicated	Fragmented	Missing
BUSCO	933 (97.8%)	919 (96.3%)	14 (1.5%)	13 (1.4%)	8 (0.8%)
OMArk	2,040 (94.75%)	1,930 (89.64%)	110 (5.11%)	-	113 (5.25%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	10,968 (57.45%)	1,981 (10.38%)	0 (0.00%)	6,142 (32.17%)	

Data availability

Anisus vorticulus and the related genomic study were assigned to Tree of Life ID (ToLID) 'xgAniVori1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB77487. The sample information is available at the following BioSample accessions: SAMEA114640075, and SAMEA114640076. The genome assembly is accessible from GenBank under accession number GCA_964264155.1 and the annotated genome is available through the Ensembl ERGA-BGE project page (https://projects.ensembl.org/erga-bge/) and Ensembl website (https://beta.ensembl.org). Sequencing data produced as part of this project are available from SRA at the following accessions: ERX12791210 and ERX12734311. Data used to generate the tables, figures and statistics in this report are available at the following repository: https://doi.org/10.5281/zenodo.15657686. 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/Anisus_vorticulus/xgAniVori1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/Anisus%20vorticulus.

Author contributions

BH and ET collected the sample, BH identified the species and provided metadata. *RM, AB, TM, TS and RAO* provided sampling and metadata support and management, *CH* extracted DNA, prepared libraries, and performed sequencing, *KH, JMW and SM* performed genome assembly and curation under the supervision of MB, LH performed genome annotation under the supervision of FM, TB 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.

Author information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.13960213.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.14870789.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.15440189.

Acknowledgements

We would like to acknowledge the assembly reviewer, Fernando Cruz, from the Centro Nacional de Análisis Genómico (CNAG). The authors 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.

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**

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

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

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**

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

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

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

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

Glöer P: **The freshwater gastropods of the west-palaearctis. Volume I–fresh-and brackish waters except spring and subterranean snails. Identification key, anatomy, ecology, distribution**. Hetlingen, Germany, 2019.
Reference Source

Glöer P, Groh K: **A contribution to the biology and ecology of the threatened species *Anisus vorticulus* (Troschel, 1834) (Gastropoda: Pulmonata: Planorbidae).** *Mollusca.* 2007; **25**: 33–40.
Reference Source

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.
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**. *protocols.io*. 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**

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**

Mazzoni C, Ciofi C, Waterhouse R: **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**

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

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

Pointon DL, Eagles W, Sims Y, *et al.*: **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**

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

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**

Van Damme D: **Anisus vorticulus**. The IUCN Red List of Threatened Species, 2012; **2012**: e.T155966A738056.
Publisher Full Text

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

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

Open Peer Review

Current Peer Review Status: ✔ ?

Version 1

Reviewer Report 08 September 2025

https://doi.org/10.21956/openreseurope.22183.r57584

© **2025 Neiman M.** 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.

?

Maurine Neiman
University of Iowa, Iowa City, USA

This well-justified genome report provides a mostly clear summary of a high-quality genome assembly available for Anisus vorticulus, a planorbid snail native to Eurasia. This species is useful as an indicator for pristine aquatic habitats and is endangered across at least some of its range. The resources presented here will be of use both for conservation purposes (e.g., to help identify which populations harbor the most diversity and should thus be focused upon for protection) and as a new addition to the growing body of genomic resources from mollusks.

This report could be improved upon by additional clarity with respect to which of the 20 snails mentioned in the "Sample and sampling information" section in the methods were used for particular project components, from genome sequencing to RNA-based analyses. It is especially important to know how many snails were used to generate the genome assembly. This information is a necessary component of assessing assembly quality and utility with respect to assessing genetic diversity, but this information regarding sample size was not at all apparent in the report. I tried to find this information in the various links provided in the Data Availability statement, but the information were also not apparent here, including a completely empty ERGA portal page (https://portal.erga-biodiversity.eu/data_portal/Anisus%20vorticulus)

I also am concerned about the use of "ancestral" taxa to estimate genome size and haploid number. As documented in many animal taxa, including the snail species that I study (Neiman, Pilcher, Haase, Lamatsch, 2025, Royal Society Open Science, https://royalsocietypublishing.org/doi/10.1098/rsos.250171), nuclear genome size can vary quite a bit both between closely related taxa as well as within a species. Chromosome number can vary as well (McElroy, Neiman et al., 2021, Molecular Biology and Evolution, doi:10.1093/molbev/msab121; also see Wallace 1992; https://doi.org/10.1093/mollus/58.2.93).

In other words, when at all possible, scientists should avoid estimating genome size and chromosome counts from anything other than the focal taxa and should also use genome size estimates and chromosome counts from multiple individuals as a means to evaluate the potential for intraspecific variation.

Ideally, the authors would thus use flow cytometry to carefully characterize nuclear genome DNA content (and variation) within A. vorticulus, and then report their bioinformatic estimates of genome size with respect to the actual size as determined by flow cytometry. If they cannot do flow cytometry at the standard required to accurately measure nuclear genome DNA content or are not able to assess multiple conspecifics from across the species range as a means of checking for variation in nuclear genome DNA content, they should acknowledge this limitation of their study when it comes to describing estimated genome size.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: evolution of sex and genomes; genomics of Potamopyrgus antipodarum (New Zealand freshwater snail), and congeners

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 01 September 2025

https://doi.org/10.21956/openreseurope.22183.r57580

© 2025 Saito T. 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.

Takumi Saito
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Major comments
This paper presents the novel chromosomal genome assembly with high-quality methods for an endangered freshwater snail, and the assembled genome will serve as a foundation for future research as the reference data. All methodologies meet contemporary standards, and the assembled genome also has been appropriately evaluated through several statistics. There are

few points requiring revision, but adding some elements in the introduction/methods sections could make it more useful for readers and future research.
Minor comments
Introduction 2nd paragraph: It is recommended to cite the Red Lists of each country.
Introduction 3rd paragraph: Additional citations after "Therefore" are maybe needed for the robustness.
Materials & Methods, Vouchering information: If photographs of the shell were available here, they would be useful for readers and future research.
Materials & Methods, Genetic information: What is the ancestral taxa? Please explain it. I understand that this section is just preliminary information and therefore not particularly important, but I feel the current explanation is rather inadequate.
Materials & Methods, Genome assembly methods: The meaning of "article in preparation" is not so clear (Is the paper being prepared, or is the linked text currently under construction?). Please clarify them.

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: Phylogeny/Population Genetics/Taxonomy of freshwater snails

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.