



DATA NOTE

The genome sequence of the pond louse, *Asellus aquaticus* (Linnaeus, 1758)

[version 1; peer review: 2 approved]

Jessica Thomas¹,
University of Oxford and Wytham Woods Genome Acquisition Lab,
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 First published: 02 Jun 2025, 10:291
<https://doi.org/10.12688/wellcomeopenres.24299.1>
Latest published: 02 Jun 2025, 10:291
<https://doi.org/10.12688/wellcomeopenres.24299.1>

Abstract

We present a genome assembly from a specimen of *Asellus aquaticus* (pond louse; Arthropoda; Malacostraca; Isopoda; Asellidae). The genome sequence has a total length of 1,977.80 megabases. Most of the assembly (93.57%) is scaffolded into 8 chromosomal pseudomolecules. The mitochondrial genome has a length of 16.53 kilobases.

Keywords

Asellus aquaticus, pond louse, genome sequence, chromosomal, Isopoda



This article is included in the [Tree of Life](#) gateway.

Open Peer Review

Approval Status

	1	2
version 1		
02 Jun 2025	view	view

1. **Tara Cronin** , Clemson university, Clemson, USA
2. **Martin Schwentner**, Natural History Museum Vienna, Vienna, Austria

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

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Thomas J: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, <https://doi.org/10.35802/218328>].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Thomas J *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: Thomas J, University of Oxford and Wytham Woods Genome Acquisition Lab, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the pond louse, *Asellus aquaticus* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:291 <https://doi.org/10.12688/wellcomeopenres.24299.1>

First published: 02 Jun 2025, 10:291 <https://doi.org/10.12688/wellcomeopenres.24299.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Crustacea; Multicrustacea; Malacostraca; Eumalacostraca; Peracarida; Isopoda; Asellota; Aselloidea; Asellidae; *Asellus*; *Asellus aquaticus* (Linnaeus, 1758) (NCBI:txid92525)

Background

The pond louse, *Asellus aquaticus*, is a freshwater isopod belonging to the family Asellidae. It is also known by several other common names, including pond slater, water louse, aquatic sowbug and water hoglouse. *Asellus aquaticus* can grow up to 15 mm in length and has a mottled brown body, resembling its terrestrial relative, the woodlouse. This species can be distinguished from the less common UK species, *Proasellus meridionalis*, by the presence of two white spots on its head, compared to a single white spot characteristic of *P. meridionalis* (Gregory, 2009). *A. aquaticus* is a detritivore, and feeds on organic debris at the bottom of ponds and streams. It is widely distributed in temperate freshwater habitats across the western Palearctic: including the British and Irish Isles, northern Europe, and Russia.

Recently, *Asellus aquaticus* has emerged as a model taxon for studying parallel adaptation to subterranean habitats (Balázs *et al.*, 2021). *Asellus aquaticus* has independently colonised several different cave sites across eastern Europe (Pérez-Moreno *et al.*, 2017; Verovnik & Konec, 2019). Isolated subterranean populations reveal strikingly similar phenotypes, with parallel evolution of cave-associated traits such as depigmentation and eye reduction (Protas *et al.*, 2011; Verovnik & Konec, 2019). A complete genome assembly of *A. aquaticus* provides a valuable resource for studying the genetic basis of parallel adaptation to subterranean environments and may help to characterise the genomic changes driving the evolution of subspecies of *A. aquaticus* in these habitats (Balázs *et al.*, 2021).

We present a chromosomally complete genome sequence for *Asellus aquaticus* based on an adult specimen (Figure 1) collected from a pond in Wytham Woods, Oxfordshire, UK as part of the Darwin Tree of Life Project. This project is a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland (Blaxter *et al.*, 2022).

Genome sequence report

Sequencing data

The genome of a specimen of *Asellus aquaticus* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 131.32 Gb from 12.64 million reads, which were used to assemble the genome. GenomeScope analysis estimated the haploid genome size at 1,920.22 Mb, with a heterozygosity of 1.96% and repeat content of 63.17%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 56 coverage. Hi-C sequencing produced 245.60 Gb from 1,626.49 million reads, and was used to scaffold the assembly. RNA sequencing data were also generated and are available in



Figure 1. Photograph of the *Asellus aquaticus* (qmAseAqua29) specimen used for genome sequencing.

public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The assembly was improved by manual curation, which corrected 1384 misjoins or missing joins and removed 691 haplotypic duplications. These interventions reduced the total assembly length by 24.83%, decreased the scaffold count by 18.37%, and increased the scaffold N50 by 48.32%. The final assembly has a total length of 1,977.80 Mb in 1,243 scaffolds, with 1,826 gaps, and a scaffold N50 of 253.56 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (93.57%) was assigned to 8 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer completeness

Table 1. Specimen and sequencing data for *Asellus aquaticus*.

Project information			
Study title	Asellus aquaticus (water hoglouse)		
Umbrella BioProject	PRJEB74712		
Species	Asellus aquaticus		
BioSpecimen	SAMEA110451522		
NCBI taxonomy ID	92525		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	qmAseAqua29	SAMEA110451707	whole organism
Hi-C sequencing	qmAseAqua15	SAMEA7521464	whole organism
RNA sequencing	qmAseAqua9	SAMEA7521400	whole organism
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR12893035	1.63e+09	245.6
PacBio Sequel IIe	ERR12875181	5.74e+05	5.28
PacBio Revio	ERR12875178	8.21e+06	86.15
PacBio Sequel II	ERR12875179	1.55e+06	13.83
PacBio Sequel IIe	ERR12875180	2.30e+06	26.07
RNA Illumina HiSeq 4000	ERR12893036	2.98e+07	4.5

indicates the proportion of expected *k*-mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

The combined primary and alternate assemblies achieve an estimated QV of 53.3. The *k*-mer recovery for the primary haplotype is 67.72%, and for the alternate haplotype 66.38%; the combined primary and alternate assemblies have a *k*-mer recovery of 97.95%. BUSCO v.5.5.0 analysis using the arthropoda_odb10 reference set (*n* = 1,013) identified 88.4% of the expected gene set (single = 86.3%, duplicated = 2.1%).

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The assembly achieves the EBP reference standard of 6.C.Q53.

Methods

Sample acquisition and DNA barcoding

An adult *Asellus aquaticus* (specimen ID Ox001984, ToLID qmAseAqua29) was collected from Wytham Woods, Oxfordshire, United Kingdom (latitude 51.764, longitude -1.337) on 2021-10-10 by potting. The specimen was collected and identified by Jessica Thomas (Wellcome Sanger Institute) and preserved on dry ice.

Two other adult specimens were used for Hi-C and RNA sequencing. The specimen used for Hi-C sequencing (specimen ID NHMUK014360867, ToLID qmAseAqua15) was collected from Shropshire Union Canal, United Kingdom (latitude 52.975, longitude -2.510) on 2019-03-18, using a kick-net. The specimen used for RNA sequencing (specimen ID NHMUK014361077, ToLID qmAseAqua9) was collected from Tewin Bury Farm, Tewin, England, UK (latitude 51.810, longitude -0.164). The specimens were collected by the Environment Agency on behalf of the Natural History Museum.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (Pereira *et al.*, 2022). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io (Beasley *et al.*, 2023).

Table 2. Genome assembly data for *Asellus aquaticus*.

Genome assembly		
Assembly name	qmAseAqua29.1	
Assembly accession	GCA_964212115.1	
Alternate haplotype accession	GCA_964212135.1	
Assembly level for primary assembly	chromosome	
Span (Mb)	1,977.80	
Number of contigs	3,069	
Number of scaffolds	1,243	
Longest scaffold (Mb)	297.65	
Assembly metric	Measure	Benchmark
Contig N50 length	1.77 Mb	≥ 1 Mb
Scaffold N50 length	253.56 Mb	= chromosome N50
Consensus quality (QV)	Primary: 53.4; alternate: 53.3; combined: 53.3	≥ 40
k-mer completeness	Primary: 67.72%; alternate: 66.38%; combined: 97.95%	$\geq 95\%$
BUSCO*	C:88.4%[S:86.3%,D:2.1%], F:3.5%,M:8.2%,n:1,013	$S > 90\%$; $D < 5\%$
Percentage of assembly mapped to chromosomes	93.57%	$\geq 90\%$
Sex chromosomes	None	localised homologous pairs
Organelles	Mitochondrial genome: 16.53 kb	complete single alleles

* BUSCO scores based on the arthropoda_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

Metadata collection for samples adhered to the Darwin Tree of Life project standards described by [Lawniczak et al. \(2022\)](#).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io ([Denton et al., 2023b](#)). The qmAseAqua29 sample was prepared for DNA extraction by weighing and dissecting it on dry ice ([Jay et al., 2023](#)). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor ([Denton et al., 2023a](#)). HMW DNA was extracted using the Manual MagAttract v1 protocol ([Strickland et al., 2023b](#)). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system ([Todorovic et al., 2023](#)). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate

shorter fragments and concentrate the DNA ([Strickland et al., 2023a](#)). 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. For this sample, the extracted DNA had a Qubit concentration of 12.0 ng/μL and a yield of 2,400.00 ng. Spectrophotometric measurements indicated 260/280 and 260/230 ratios of 1.92 and 6.05, respectively.

RNA was extracted from whole organism tissue of qmAseAqua9 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.

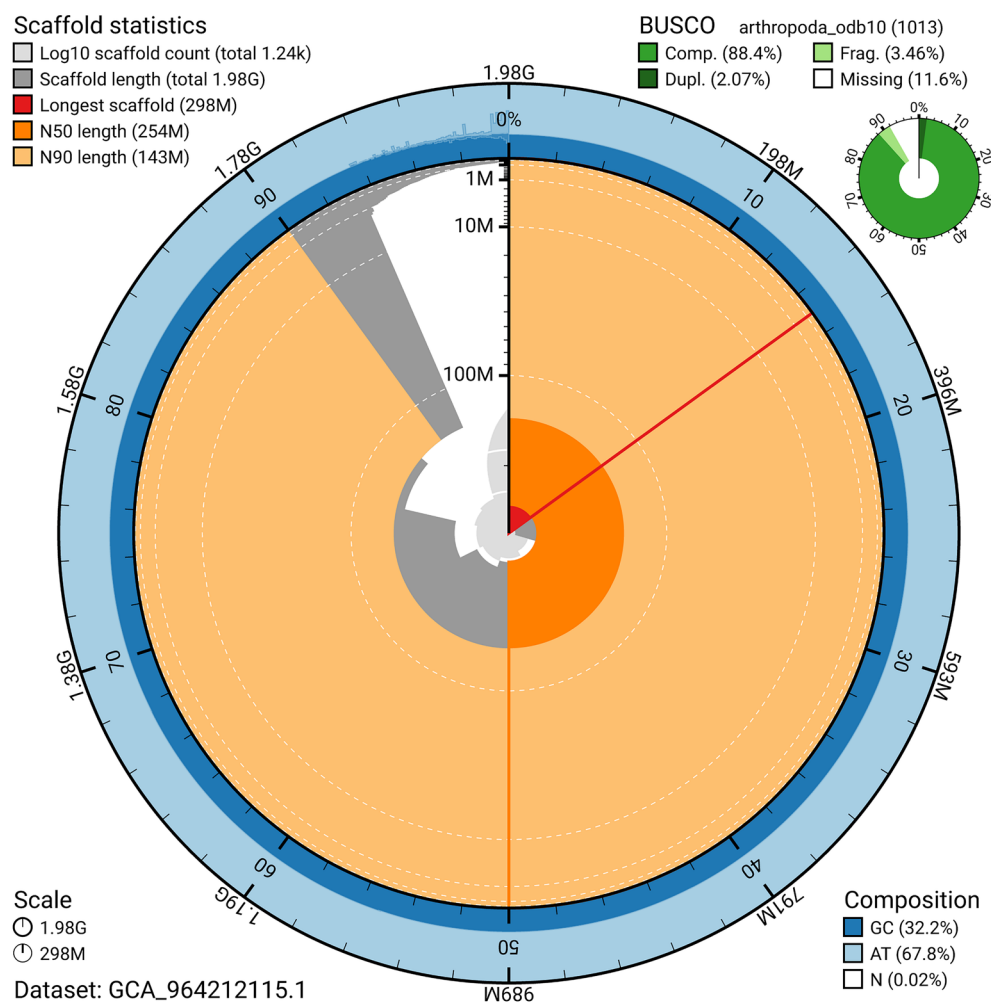


Figure 2. Genome assembly of *Asellus aquaticus*, qmAseAqua29.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964212115.1/dataset/GCA_964212115.1/snail.

Hi-C sample preparation and crosslinking

Hi-C data were generated from the whole organism of the qmAseAqua15 sample using the Arima-HiC v2 kit (Arima Genomics) with 20–50 mg of frozen tissue (stored at -80°C). As per manufacturer's instructions, tissue was fixed, and the DNA crosslinked using a TC buffer with a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagenode Power Masher-II. The crosslinked DNA was digested using a restriction enzyme master mix, then biotinylated and ligated. A clean up was performed with SPRIselect beads prior to library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and

Qubit HS Assay Kit, and sample biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Library preparation and sequencing

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

PacBio HiFi

At a 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), depending on genome size and sequencing depth

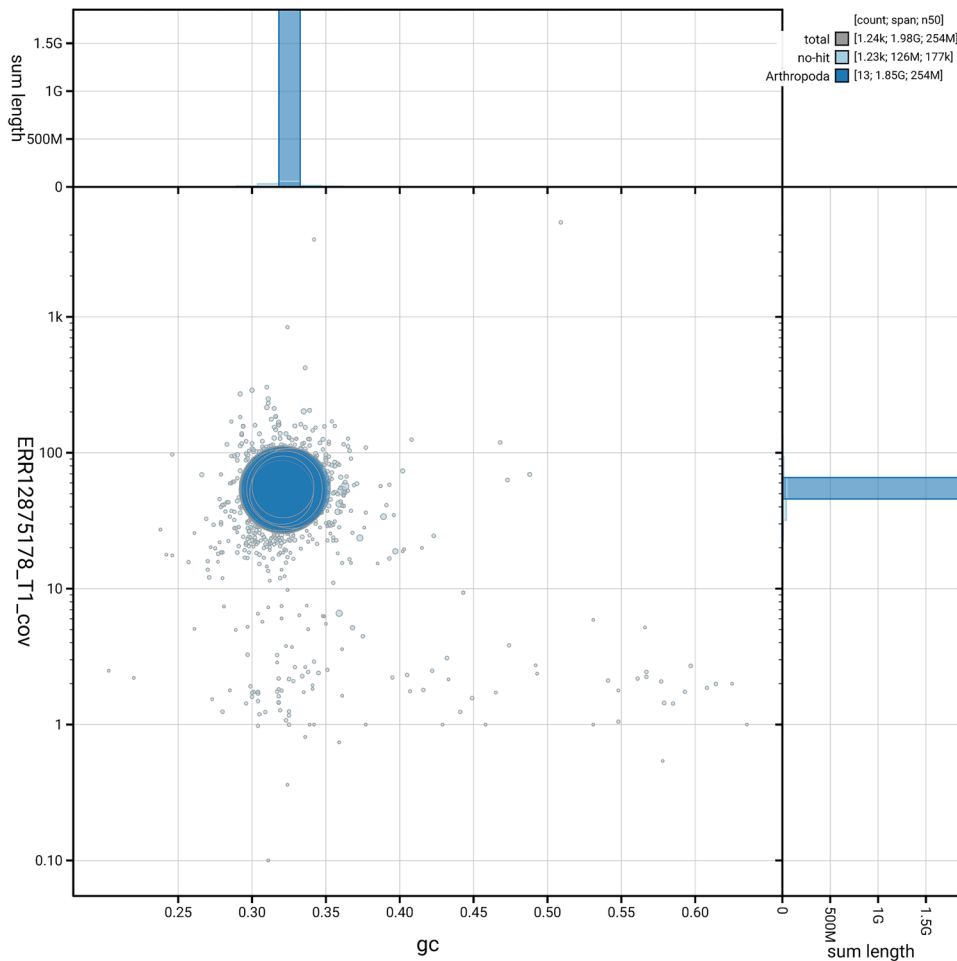


Figure 3. Genome assembly of *Asellus aquaticus*, qmAseAqua29.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964212115.1/dataset/GCA_964212115.1/blob.

required. Libraries were prepared using the SMRTbell Prep Kit 3.0 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. Size-selection and clean-up were carried out using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using the Qubit Fluorometer v4.0 (ThermoFisher 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 the gDNA 55kb BAC analysis kit.

Samples were sequenced using both the Sequel IIe system and Revio (Pacific Biosciences). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. For Revio sequencing, prepared libraries were normalised to 2 nM, and 15 μ L was used for making complexes. Primers were annealed

and polymerases were bound to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration (in the range 200–300 pM), and spiked with a Revio sequencing internal control. Samples were sequenced 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 runs, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, the biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size-selected to 400–600 bp using SPRIselect beads. DNA was then enriched using Arima-HiC v2 Enrichment beads. The NEBNext Ultra

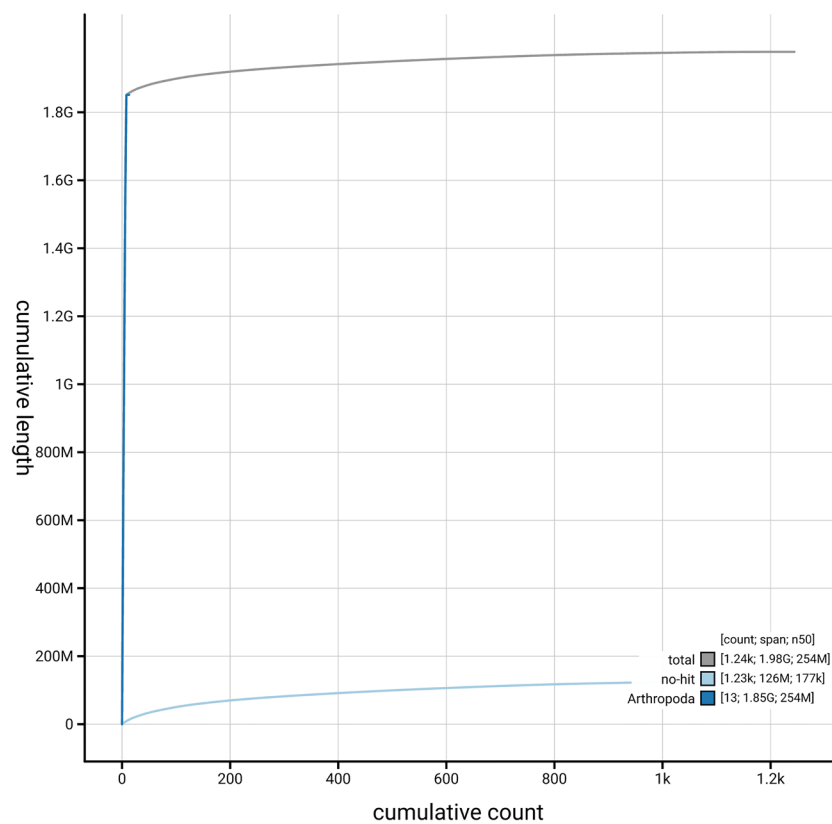


Figure 4. Genome assembly of *Asellus aquaticus*, qmAseAqua29.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964212115.1/dataset/GCA_964212115.1/cumulative.

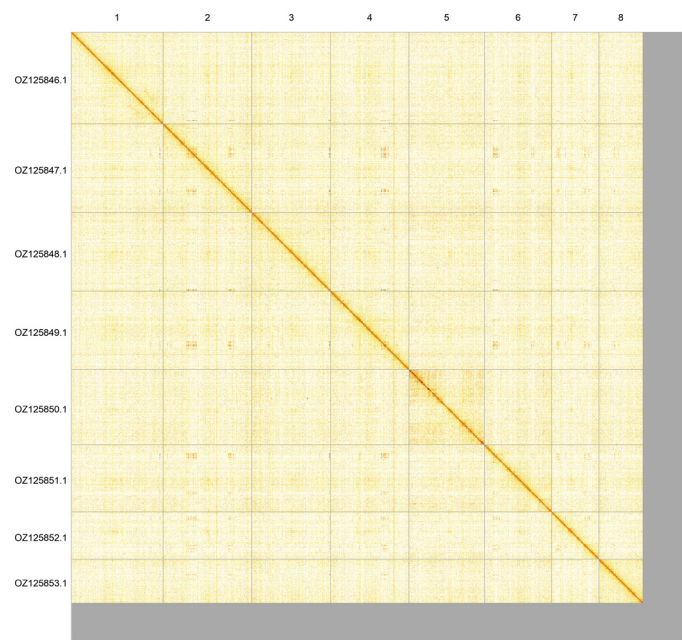


Figure 5. Genome assembly of *Asellus aquaticus*: Hi-C contact map of the qmAseAqua29.1 assembly produced in PretextView. Chromosomes are shown in order of size and labelled with chromosome numbers (top) and chromosome accession numbers (left).

Table 3. Chromosomal pseudomolecules in the genome assembly of *Asellus aquaticus*, qmAseAqua29.

INSDC accession	Name	Length (Mb)	GC%
OZ125846.1	1	297.65	32
OZ125847.1	2	287.49	32
OZ125848.1	3	254.79	32
OZ125849.1	4	253.56	32
OZ125850.1	5	243.89	32.5
OZ125851.1	6	217.78	32
OZ125852.1	7	152.5	32
OZ125853.1	8	143.02	32
OZ125854.1	MT	0.02	34

II DNA Library Prep Kit (New England Biolabs) was used for end repair, A-tailing, and adapter ligation, following a modified protocol in which library preparation is carried out while the DNA remains bound to the enrichment beads. PCR amplification was performed using KAPA HiFi HotStart mix and custom dual-indexed adapters (Integrated DNA Technologies) in a 96-well plate format. Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, samples were amplified for 10–16 PCR cycles. Post-PCR clean-up was carried out using SPRISelect beads. The libraries were quantified using the Accuclear Ultra High Sensitivity dsDNA Standards Assay kit (Biotium) and normalised to 10 ng/μL before sequencing. Hi-C sequencing was performed on the Illumina NovaSeq 6000 instrument using 150 bp paired-end reads.

RNA

Poly(A) RNA-Seq libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina HiSeq 4000 instrument.

Genome assembly, curation and evaluation

Assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts (*k* = 31) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

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

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023) and OATK (Zhou *et al.*, 2024).

Assembly curation

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

Assembly quality assessment

The Merquery.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer databases (*k* = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The blobtoolkit pipeline is a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoAT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these outputs into a blobdir for visualisation.

The blobtoolkit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), relying on the Conda package manager, the Bioconda initiative

(Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

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

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the ‘**Darwin Tree of Life Project Sampling Code of Practice**’, which can be found in full on the Darwin Tree of Life website

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	666652151335353eef2fcd58880bcef5bc2928e1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Goat CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.5-r587	https://github.com/chhy123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
OATK	1.0	https://github.com/c-zhou/oatk
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

here. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

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

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

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

Data availability

European Nucleotide Archive: *Asellus aquaticus* (water hoglouse). Accession number PRJEB74712; <https://identifiers.org/ena.embl/PRJEB74712>. The genome sequence is released openly for reuse. The *Asellus aquaticus* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project and the

European Reference Genome Atlas Pilot Project (ERGA-PI). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12157525>.

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12159242>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>.

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

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

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

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

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.4783558>.

References

- Altschul SF, Gish W, Miller W, et al.: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Balázs G, Biró A, Fišer Z, et al.: **Parallel morphological evolution and habitat-dependent sexual dimorphism in cave- vs. surface populations of the *Asellus aquaticus* (Crustacea: Isopoda: Asellidae) species complex.** *Ecol Evol.* 2021; **11**(21): 15389–15403.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bateman A, Martin MJ, Orchard S, et al.: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Beasley J, Uhl R, Forrest LL, et al.: **DNA barcoding SOPs for the Darwin Tree of Life Project.** *protocols.io.* 2023; [Accessed 25 June 2024].
[Publisher Full Text](#)
- Blaxter M, Mieszkowska N, Di Palma F, et al.: **Sequence locally, think globally: the Darwin Tree of Life project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Crowley L, Allen H, Barnes I, et al.: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- da Veiga Leprevost F, Grüning BA, Alves Aflitos S, et al.: **BioContainers:**

an open-source and community-driven framework for software standardization. *Bioinformatics*. 2017; **33**(16): 2580–2582.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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](#)

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol*. 2017; **35**(4): 316–319.
[PubMed Abstract](#) | [Publisher Full Text](#)

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

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

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

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

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](#)

Gregory S: **Woodlice and waterlice (Isopoda: Oniscidea & Asellota) in Britain and Ireland.** Field Studies Council/Centre for Ecology & Hydrology, 2009.
[Reference Source](#)

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods*. 2018; **15**(7): 475–476.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

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): g1aa153.
[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](#)

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

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

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res*. 2022; **7**: 187.
[Publisher Full Text](#)

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

Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol*

Evol. 2021; **38**(10): 4647–4654.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

Pereira L, Sivell O, Sivess L, *et al.*: **DTOL: taxon-specific standard operating procedure for the terrestrial and freshwater arthropods working group.** 2022.
[Publisher Full Text](#)

Pérez-Moreno JL, Balázs G, Wilkins B, *et al.*: **The role of isolation on contrasting phylogeographic patterns in two cave crustaceans.** *BMC Evol Biol*. 2017; **17**(1): 247.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

Protas ME, Trontelj P, Patel NH: **Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*.** *Proc Natl Acad Sci U S A*. 2011; **108**(14): 5702–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun*. 2020; **11**(1): 1432.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

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

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

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

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

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

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res*. 2024; **9**: 339.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

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

Verovnik R, Konec M: ***Asellus aquaticus*: a model system for historical biogeography.** In: White, W. B., Culver, D. C., and Pipan, T. (eds.) *Encyclopedia of caves (Third Edition)*. Academic Press, 2019; 76–84.
[Publisher Full Text](#)

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: a *de novo* assembly tool for complex plant organelle genomes.** *bioRxiv*. 2024.
[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 13 January 2026

<https://doi.org/10.21956/wellcomeopenres.26797.r143369>

© 2026 Schwentner 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.



Martin Schwentner

Natural History Museum Vienna, Vienna, Austria

The author, Jessica Thomas, provides the first chromosome-scale genome assembly of the freshwater isopod species *A. aquaticus*, which is of evolutionary as well as ecological importance. The results are highly relevant, well presented and the underlying methods and protocols clearly explained and presented. All relevant data is shown and a part from two minor questions I am very satisfied with the manuscript and have nothing to add or criticize

1. is there a citation for the ASCC pipeline?

2. The analyses relies on large number of scripts (I guess including many own pipelines etc). It would be great if collection of scripts and commands used herein were provided as well (maybe they are and I just missed 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: phylogenomics

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

Reviewer Report 03 November 2025

<https://doi.org/10.21956/wellcomeopenres.26797.r125959>

© 2025 Cronin 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.

**Tara Cronin** 

Clemson university, Clemson, USA

Article – The genome sequence of the pond louse, *Asellus aquaticus* (Linnaeus, 1758)

Summary – In this study, the author presents the first documentation of the genome assembly of the pond louse, *Asellus aquaticus*. It is part of the Darwin Tree of Life project which aims to sequence all eukaryotic species in the Atlantic Archipelago of Britain and Ireland. The authors produced a genome sequence of 1,977.80 Mb, with the majority of the assembly being scaffolded into 8 chromosomal pseudomolecules. Additionally, the authors assembled the mitochondrial genome of this species. The authors measured the QV and k-mer completeness, along with BUSCO scores for the assembly. These measures resulting in a QV of 53.3, k-mer recovery of 97.95%, and a BUSCO score of 88.4%

Comments related to questions-

1. Is the rationale for creating the dataset clearly described?

The dataset created is part of a much larger project, The Darwin Tree of Life project, which aims to sequence all name eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Additionally, the emergence of several isolated subterranean populations that have evolved cave-associated traits has made this organism a model taxon to study parallel adaptations. This genomic resource will allow for characterization of genomic changes driving evolution of these subterranean populations. The authors describe both the project and the need/importance of this genome clearly.

2. Are the protocols appropriate and is the work technically sound?

The authors used appropriate protocols and the work appears to be technically sound. The protocols are clearly outlined for the entire Darwin Tree of Life project, allowing for all genomes sequenced for this project to have the same protocol. Numerous other publications have already used this protocol adding to the support of its appropriateness.

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

Yes sufficient details are provided to allow for replication. The authors do a good job describing their methods in the paper, and provide access to all their protocols so that others may replicate this study with ease.

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

The dataset is archived in repositories online allowing for others to access all the raw data and assembly data. Additionally, the figures and tables provide a good presentation of the data for readers to get a good understanding of what is in the repositories.

Minor Revisions –

Line 48, Left column – provide the suborder of the woodlouse in paraentheses.

Line 51, Left column – Write out *Asellus aquaticus* rather than using the abbreviated form, as it is at the beginning of a sentence.

Line 53, Left column – provide a citation for the sentence about the distribution of this species.

Line 57, Left column – Change *Asellus aquaticus* to *A. aquaticus*, as you have already written out the full name and it is not the start of the sentence.

Line 72, Left column - Change *Asellus aquaticus* to *A. aquaticus*, as you have already written out the full name and it is not the start of the sentence.

Line 82, Left column – change Gb to gigabases (Gb) as it is the first time the abbreviation is use and should be defined.

Line 84, Left column – change Mb to megabases (Mb) as it is the first time the abbreviation is use and should be defined.

Line 88, Left column – change 56 to 56x.

Line 89, Left column – remove the comma between reads and and.

Line 56, Right column – provide a link or name of the repositories you are referring to so that readers can quickly find it.

Line 157, Left column – Change *Asellus aquaticus* to *A. aquaticus*, as you have already written out the full name and it is not the start of the sentence.

Line 140, Right column – there is a paraentheses missing after the ToLID number.

Line 154, Right column – It is not clear that this is a standardized protocol the way the sentence currently is written. Mentioning that this protocol is standardized would help the reader know that more detailed methods can be found in that protocol.

Line 158, Right column – state what the relevant DNA barcode region is for this sample to help with clarity for the reader.

Line 207, Left column – to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the PowerMasher II tissue disruptor.

Line 207, Left column – Generally starting sentences with abbreviations should be avoided, change HMW to High Molecular Weight (HMW) as it is the start of the sentence and the first time this term is used.

Line 210, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Megaruptor 3 system.

Line 212, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the AMPure PB beads.

Line 198, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Nanodrop spectrophotometer, Qubit Fluorometer, and the Qubit dsDNA High Sensitivity Assay kit.

Line 201, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the FemtoPulse system.

Line 210, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Qubit RNA Broad-Range Assay kit.

Line 211, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Line 247, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the

Arima-HiC v2 kit.

Line 252, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Diagnocine Power Masher II.

Line 254, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the SPRiselect beads.

Line 256, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the Qubit Fluorometer v4.0.

Line 245, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Qubit HS assay Kit.

Line 246, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Arima-HiC v2 QC beads.

Line 255, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the SMRTbell Prep Kit 3.0.

Line 283, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the AMPure PB beads.

Line 284, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the Qubit Fluorometer v4.0 and the Qubit 1X dsDNA HS assay kit.

Line 287, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the Agilent Femto Pulse and gDNA 55kb BAC analysis kit.

Line 291, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the Sequel IIe system and Revio.

Line 280, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the SMRTbell beads.

Line 292, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Covaris E220 sonicator.

Line 294, Right column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Arima-HiC v2 Enrichment beads.

Line 332, Left column - - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the prep kit.

Line 336, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the KAPA HiFi HotStart mix.

Line 337, Left column - - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the

adapters.

Line 343, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for Biotium.

Line 344, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the manufacturer name and headquarters location for the Illumina NovaSeq.

Line 349, Left column - to keep a consistent style with the rest of the paper, and to increase the ease and ability of the reader to replicate your methods, provide the headquarters location for the RNA library Prep Kit.

Line 340, Right column – Provide a citation for the ASCC pipeline.

Major Revisions – N/A

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: Mitochondrial genomics, behavioral ecology, invertebrate biology, physiology, parasitology

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.
