DATA NOTE

The genome sequence of the click beetle, *Ampedus sanguinolentus sanguinolentus* (Schrank, 1776)

[version 1; peer review: 3 approved]

Duncan Sivell[1], Dmitry Telnov[1-3], Michael F. Geiser[1], Maxwell V. L. Barclay[1], 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

[1]Natural History Museum, London, England, UK
[2]Daugavpils University, Daugavpils, Latvia
[3]Institute of Biology, University of Latvia, Rīga, Latvia

v1 **First published:** 24 Feb 2025, **10**:96
https://doi.org/10.12688/wellcomeopenres.23712.1
Latest published: 24 Feb 2025, **10**:96
https://doi.org/10.12688/wellcomeopenres.23712.1

Abstract

We present a genome assembly from a female specimen of *Ampedus sanguinolentus sanguinolentus* (click beetle; Arthropoda; Insecta; Coleoptera; Elateridae). The assembly contains two haplotypes with total lengths of 1,574.76 megabases and 1,572.87 megabases. Most of haplotype 1 (97.13%) is scaffolded into 10 chromosomal pseudomolecules, while haplotype 2 is a scaffold-level assembly. The mitochondrial genome has also been assembled and is 15.99 kilobases in length.

Keywords

Ampedus sanguinolentus sanguinolentus, click beetle, genome sequence, chromosomal, Coleoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 24 Feb 2025	✓ view	✓ view	✓ view

1. **Hume Douglas**, Agriculture and Agri-Food Canada, Ottawa, Canada
2. **Ruiqi Li**, University of Colorado Boulder, Boulder, USA
3. **Erich D Jarvis**, Rockefeller University, Millbrook, USA

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: Sivell D: Investigation, Resources; **Telnov D**: Writing – Original Draft Preparation, Writing – Review & Editing; **Geiser MF**: Investigation, Resources; **Barclay MVL**: Investigation, Resources, 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.

How to cite this article: Sivell D, Telnov D, Geiser MF *et al.* **The genome sequence of the click beetle, *Ampedus sanguinolentus sanguinolentus* (Schrank, 1776) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:96 https://doi.org/10.12688/wellcomeopenres.23712.1

First published: 24 Feb 2025, **10**:96 https://doi.org/10.12688/wellcomeopenres.23712.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Elateriformia; Elateroidea; Elateridae; Elaterinae; Ampediini; *Ampedus*; *Ampedus sanguinolentus sanguinolentus* (Schrank, 1776) (NCBI:txid941989).

Background

The Elateridae Leach, 1815, commonly known as click beetles, is a large family of the order Coleoptera with about 10,000 species worldwide (Costa *et al.*, 2010). Adult and immature (also called 'wireworms') elaterids, rather homogenous morphologically, are one of the ecologically and economically most important groups of beetles. The family is of cosmopolitan distribution, most species-rich in the tropical regions. Click beetle larvae are usually saproxylic, herbivorous, (feeding on plant roots) or opportunistic predators of small invertebrates. Adults are mostly phytophagous, feed on green plants and their juices; some species are anthophilous. One of the English names of the family refers to the peculiar click mechanism adult beetles possess, allowing them to spring into the air to avoid predators – sometimes with a distinct click. Seventeen extant subfamilies are recognised within Elateridae (Bouchard *et al.*, 2011). In the British fauna, there are 69 established species of click beetles known at present (Duff, 2020; Mendel, 2024). The genus *Ampedus* Dejean, 1833 is very speciose in the Palaearctic Region with over 250 known species; in Europe about 80 species (Cate, 2007), 12 of which are present in the British fauna (Mendel, 2024).

Ampedus sanguinolentus sanguinolentus (Schrank, 1776) *is placed* in the subfamily Elaterinae tribe Ampediini (Cate, 2007). The species is attributed to the nominate subgenus. Adults of the species are recognised by the spindle-shaped black median spot on the elytra (which can be reduced or absent, or expanded to cover most of elytra in some aberrant specimens) on the generally bright red elytra, and the comparatively sparsely punctured lateral sides of the pronotal disc (Duff, 2020; Mendel, 2024). It is hoped that genomic information may contribute towards a better understanding of subspecies concept of *Ampedus sanguinolentus*.

Ampedus sanguinolentus sanguinolentus is a trans-Palaearctic subspecies widely distributed from Portugal and the British Isles towards eastern Siberia and north-eastern China (Cate, 2007). It is recorded from 39 European countries, including the UK (Cate, 2007; Mannerkoski *et al.*, 2010). It is generally much less abundant in western, southern and southeastern Europe than in the zone of temperate and boreal forests. Outside Europe, the species is known from Morocco and widely distributed from western Turkey and Armenia to the Russian Far East and northern China (Cate, 2007). The subspecies *nippon* Kishii, 1982 is restricted to the Japanese Archipelago. The European extent of occurrence and area of occupancy of this species are both strongly exceed the thresholds for a threatened species (Mannerkoski *et al.*, 2010).

This is a forest species. Larvae of *Ampedus sanguinolentus* are omnivorous and facultatively predatory, developing in white- and brown-rot decaying wood of various deciduous, sometimes also coniferous trees and are polyphagous, developing in *Alnus* spp., *Betula* spp., *Fagus sylvatica*, *Quercus* spp., *Tilia* spp. preferring wet wood (Koch, 1992; Mendel, 2024; Nikitsky *et al.*, 1996). They usually develop in stumps or fallen trunks, sometimes larvae are found in soil, at the roots of heather, or even in pony dung (Hyman, 1992; Mendel, 2024 and references therein). Larval development takes two years (four in the northern and eastern parts of the range), adults are found throughout the year in rotten wood or peat earth (Mendel, 2024; Nikitsky *et al.*, 1996). The species is described as eurytopic, sylvicol, xylodetriticol, corticol and arboricol (Koch, 1992). Adult beetles occur in forests, on forest edges, hedgerows and clearings, and in Britain are most frequently found on lowland heaths associated with *Betula* or *Pinus* (M.V.L. Barclay personal observation). Adults feed on pollen of various flowering plants, often umbellifers and *Crataegus* spp., but also *Pinus* and many others. In Britain, adults are reported as active from March to August (Hyman, 1992). The adult specimen used in the current study was sampled in May. In the northern and eastern parts of the distribution area the adults appear in May to August.

Ampedus sanguinolentus sanguinolentus in the United Kingdom is primarily distributed in southern and southeastern England (Mendel, 2024). The northernmost records known from Bewdley Forest, Oxford, and Worcester, date to the first half of the 19th century, and the beetle was likely more widely distributed through England in the past (Hyman, 1992; Mendel, 2024). The species was not listed in the national Red Data Book (Shirt, 1987), the present status of the species in the United Kingdom is Notable (A), meaning that it is estimated to occur in 16 to 30 of the 10-km squares of the UK National Grid (Hyman, 1992). Although not widespread, it can be common where it occurs and is one of the more frequently recorded species of the genus *Ampedus* in the UK.

The adult used for sequencing was obtained in Thursley Common, Surrey, southern England, sampled by D. Sivell on 30.v.2022 and identified by Michael F. Geiser.

Genome sequence report

Sequencing data

The genome of a specimen of *Ampedus sanguinolentus* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 70.95 Gb from 7.03 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 1,544.42 Mb, with a heterozygosity of 1.88% and repeat content of 45.73%. These values provide an initial assessment of genome complexity and the challenges anticipated during assembly. Based on this estimated genome size, the sequencing data provided approximately 45.0x coverage of the genome. Chromosome conformation Hi-C data produced 112.89 Gb from 747.60 million reads. Table 1 summarises the specimen and sequencing information, including the BioProject, study name, BioSample numbers, and sequencing data for each technology.

Figure 1. Photographs of the *Ampedus sanguinolentus* (icAmpSanu1) specimen used for genome sequencing.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The assembly was improved by manual curation, which corrected 43 misjoins or missing joins. These interventions decreased the scaffold count by 1.98% and increased the scaffold N50 by 1.27%. The final haplotype 1 assembly has a total length of 1,574.76 Mb in 346 scaffolds, with 395 gaps, and a scaffold N50 of 146.6 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics for haplotype 1, 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 (97.13%) was assigned to 10 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, it was noted that the assembly is of the homogametic sex, but no X chromosome could be confidently assigned.

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 in GenBank.

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were calculated

Table 1. Specimen and sequencing data for *Ampedus sanguinolentus*.

Project information			
Study title	Ampedus sanguinolentus		
Umbrella BioProject	PRJEB78328		
Species	*Ampedus sanguinolentus*		
BioSpecimen	SAMEA114805610		
NCBI taxonomy ID	941989		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	icAmpSanu1	SAMEA114805729	whole organism
Hi-C sequencing	icAmpSanu1	SAMEA114805729	whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq X	ERR13389759	7.48e+08	112.89
PacBio Revio	ERR13420775	7.03e+06	70.95

Table 2. Genome assembly data for *Ampedus sanguinolentus*.

Genome assembly	**Haplotype 1**	**Haplotype 2**
Assembly name	icAmpSanu1.hap1.1	icAmpSanu1.hap2.1
Assembly accession	GCA_964234815.1	GCA_964234845.1
Assembly level	chromosome	scaffold
Span (Mb)	1,574.76	1,572.87
Number of contigs	741	689
Number of scaffolds	346	253
Longest scaffold (Mb)	257.77	None
Assembly metrics* (benchmark)	**Haplotype 1**	**Haplotype 2**
Contig N50 length (≥ 1 Mb)	6.82 Mb	5.4 Mb
Scaffold N50 length (= chromosome N50)	146.6 Mb	143.64 Mb
Consensus quality (QV) (≥ 40)	66.6	66.3
k-mer completeness	70.01%	69.87%
Combined *k*-mer completeness (≥ 95%)	99.52%	
BUSCO** (S > 90%; D < 5%)	C:99.3%[S:97.2%,D:2.2%], F:0.2%,M:0.5%,n:2,124	C:99.6%[S:97.4%,D:2.3%], F:0.2%,M:0.2%,n:2,124
Percentage of assembly mapped to chromosomes (≥ 90%)	97.13%	-
Sex chromosomes (localised homologous pairs)	Not identified	-
Organelles (one complete allele)	Mitochondrial genome: 15.99 kb	-

* BUSCO scores based on the endopterygota_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.

for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer completeness 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.

For haplotype 1, the estimated QV is 66.6, and for haplotype 2, the QV is 66.3. When the two haplotypes are combined, the assembly achieves an estimated QV of 66.4. The *k*-mer completeness for haplotype 1 is 70.01%, and for haplotype 2 it is 69.87%, while the combined assemblies achieve a *k*-mer completeness of 99.52%. BUSCO 5.5.0 analysis of the haplotype 1 assembly using the endopterygota_odb10 reference set (n = 2,124) achieved a completeness score of 99.3% (single = 97.2%, duplicated = 2.2%).

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.Q66**.

Methods

Sample acquisition and DNA barcoding

An adult *Ampedus sanguinolentus* (specimen ID NHMUK014440683, ToLID icAmpSanu1) was collected from Thursley Common, Waverley, England, United Kingdom (latitude 51.16, longitude –0.7) on 2022-05-30. The specimen was collected by Duncan Sivell (Natural History Museum), identified by Michael Geiser (Natural History Museum) and preserved by dry freezing (–80 °C).

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

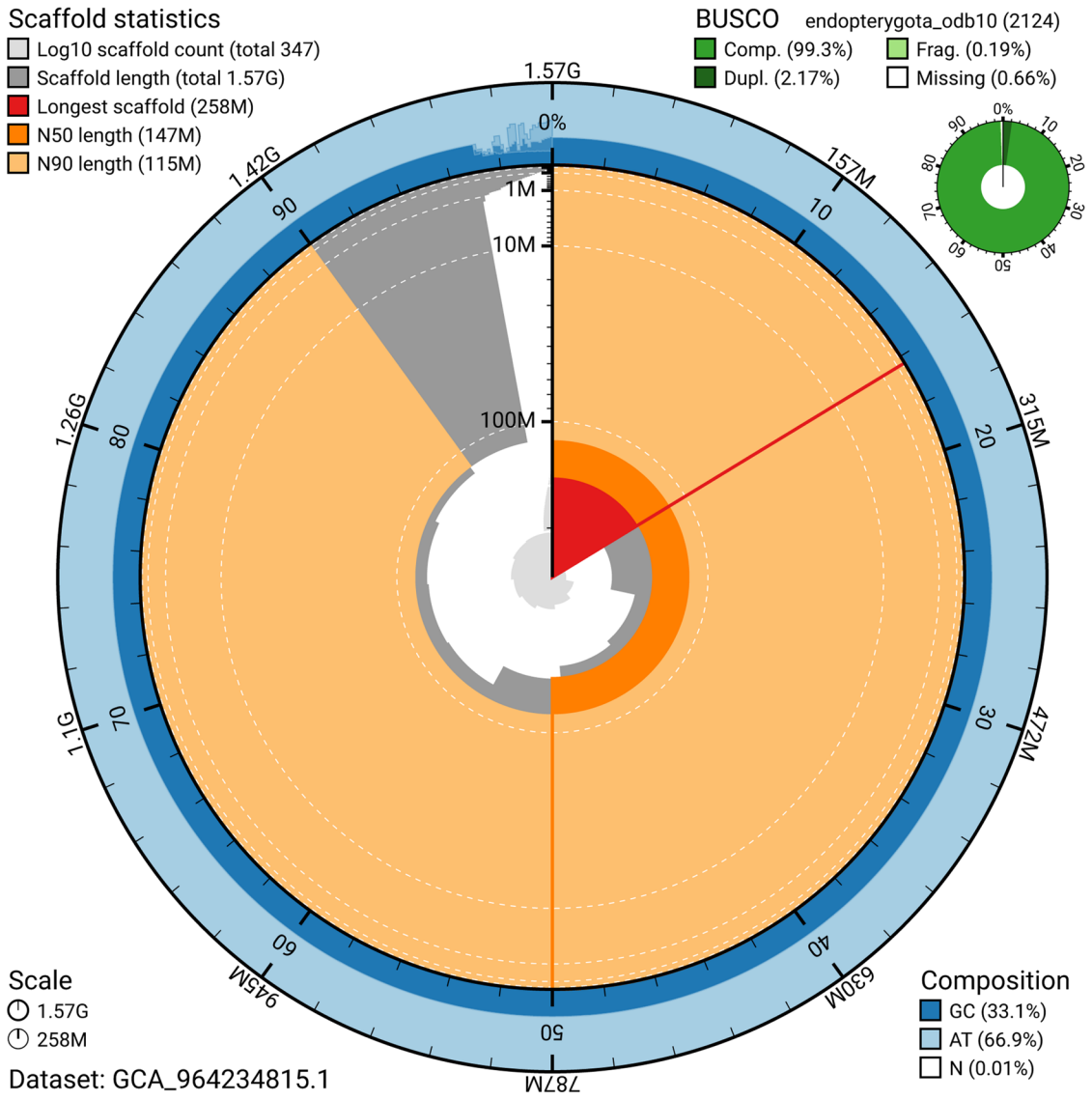

Figure 2. Genome assembly of *Ampedus sanguinolentus*, icAmpSanu1.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around 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 endopterygota_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964234815.1/dataset/GCA_964234815.1/snail.

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).

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

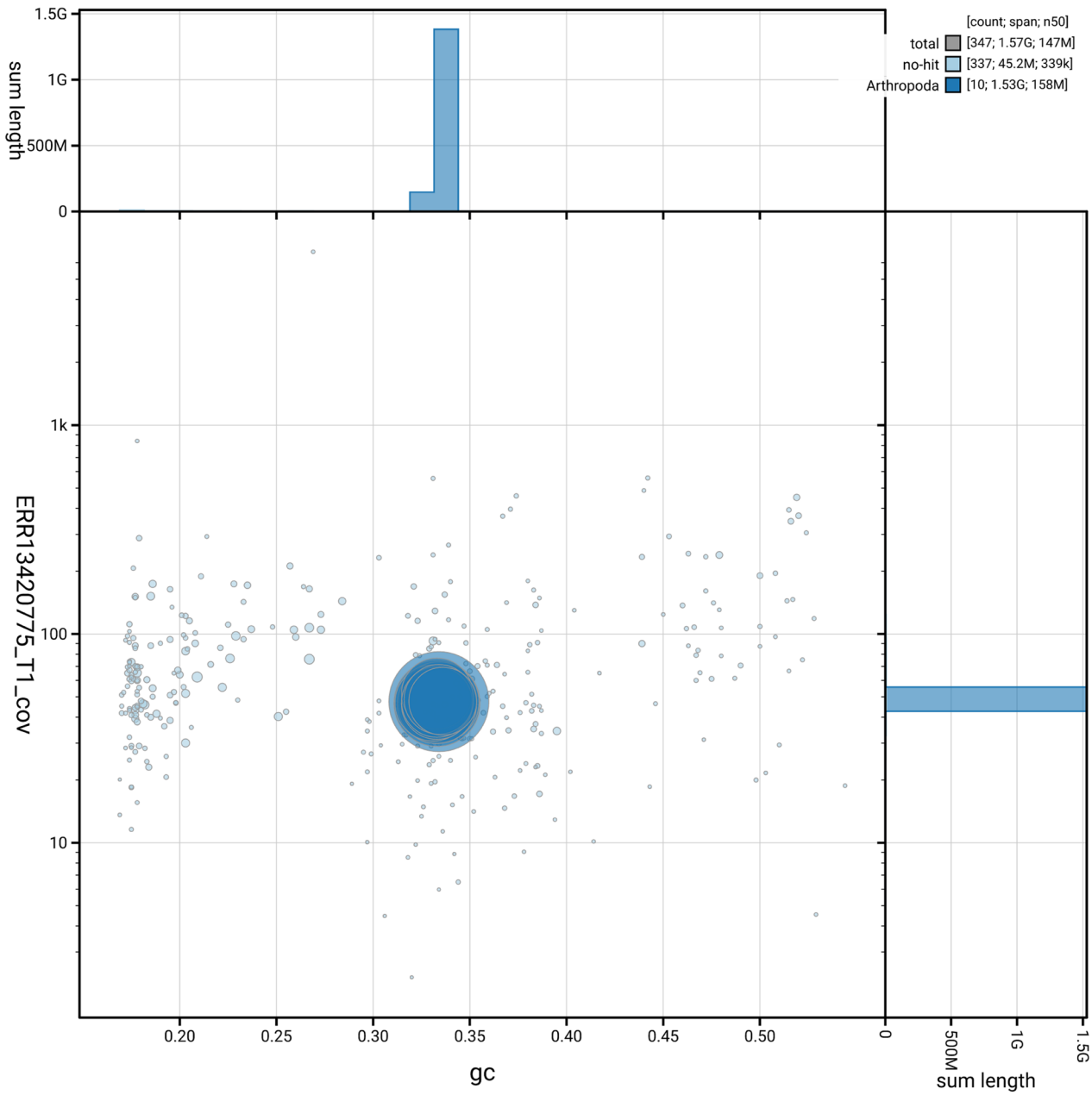

Figure 3. Genome assembly of *Ampedus sanguinolentus*, icAmpSanu1.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum 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_964234815.1/blob.

purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The icAmpSanu1 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 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

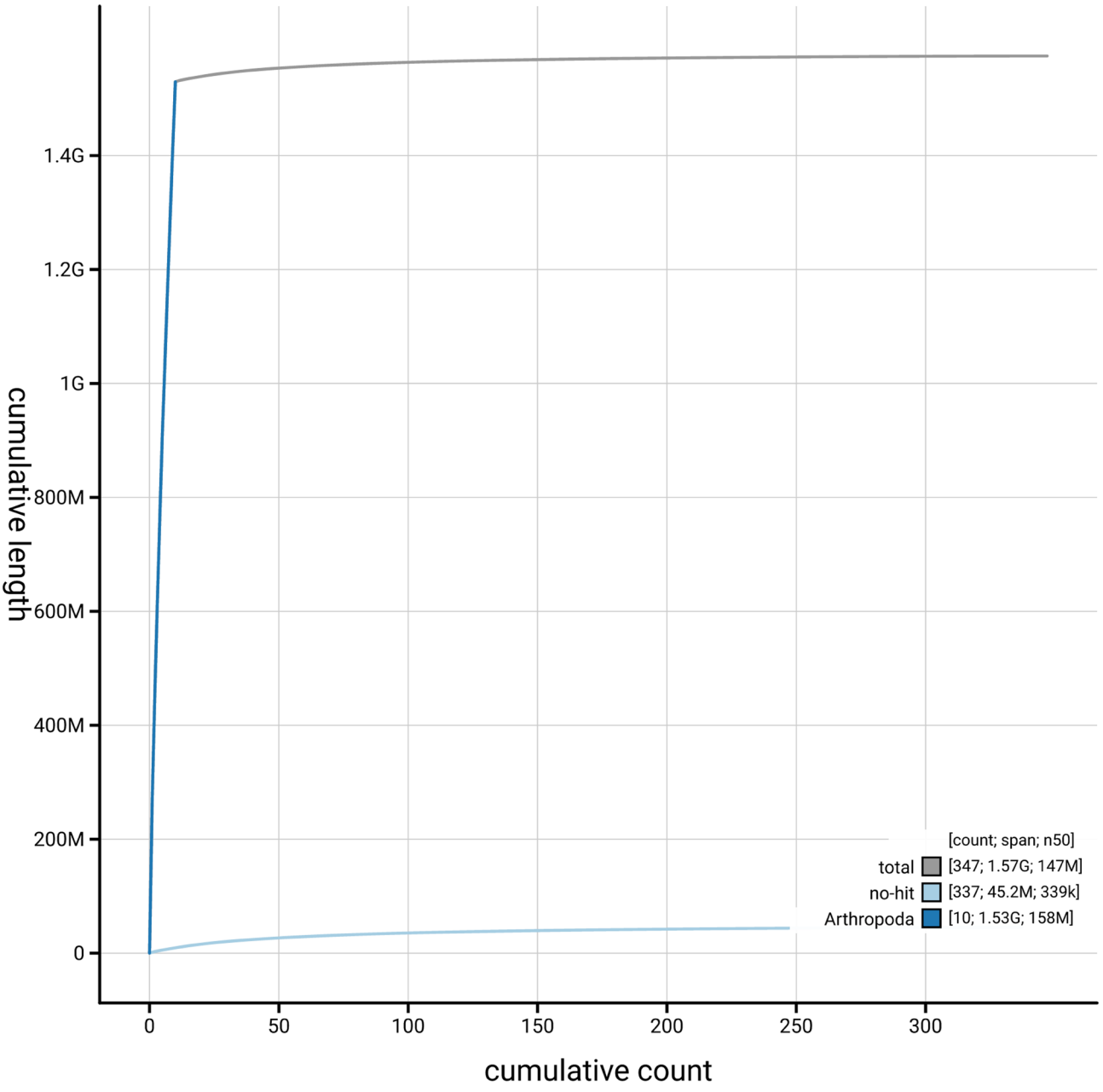

Figure 4. Genome assembly of *Ampedus sanguinolentus* icAmpSanu1.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964234815.1/dataset/GCA_964234815.1/cumulative.

Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Hi-C sample preparation

Tissue from the whole organism of the icAmpSanu1 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, 20–50 mg of frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the Arima-HiC v2 kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest

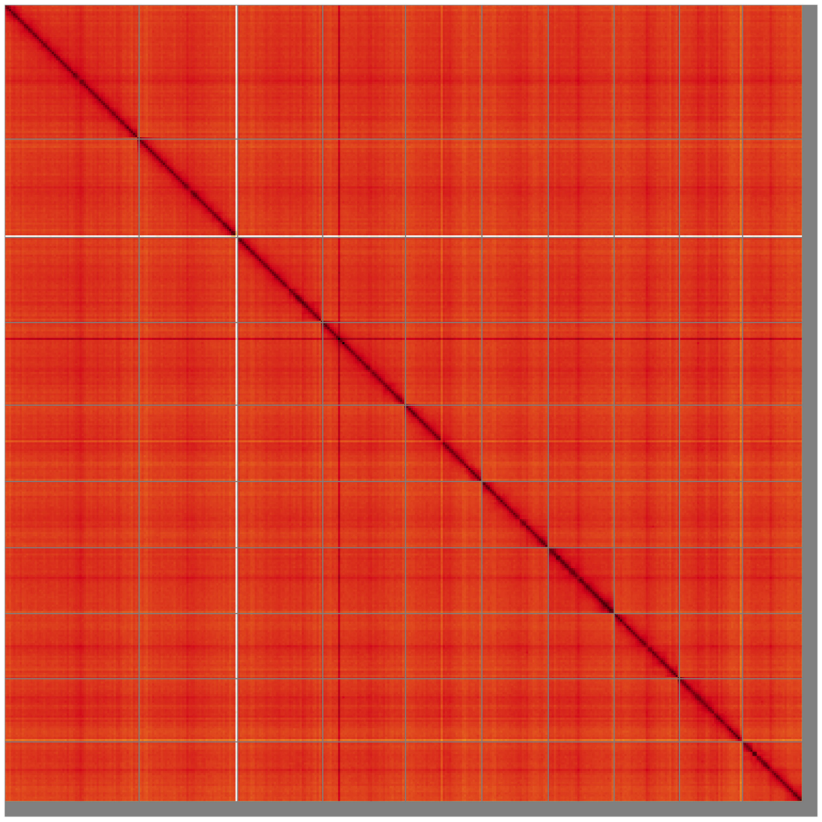

Figure 5. Genome assembly of *Ampedus sanguinolentus* icAmpSanu1.hap1.1: Hi-C contact map of the icAmpSanu1.hap1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=BtJvbz7lTAuxZAOkJOcOsQ.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Ampedus sanguinolentus*, icAmpSanu1.

INSDC accession	Name	Length (Mb)	GC%
OZ174264.1	1	257.77	33.5
OZ174265.1	2	188.73	33.5
OZ174266.1	3	163.24	33.5
OZ174267.1	4	158.49	33.5
OZ174268.1	5	146.6	33
OZ174269.1	6	127.39	33.5
OZ174270.1	7	126.21	33.5
OZ174271.1	8	124.92	33.5
OZ174272.1	9	121.25	33.5
OZ174273.1	10	114.93	33.5
OZ174274.1	MT	0.02	27

remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation. Additionally, the biotinylation percentage was estimated using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and 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, 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.

Samples were sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries were normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases were hybridised to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration (in the range 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 run, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using SPRISelect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, a-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required, determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X 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 assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), resulting in a pair of haplotype-resolved assemblies. The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

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).

Assembly quality assessment

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

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow 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.

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
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
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.5.1	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	1.2a.2	https://github.com/c-zhou/yahs

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 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: Ampedus sanguinolentus. Accession number PRJEB78328; https://identifiers.org/ena.embl/PRJEB78328. The genome sequence is released openly for reuse. The *Ampedus sanguinolentus* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. 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 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

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

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

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

Bouchard P, Bousquet Y, Davies AE, *et al.*: **Family-group names in Coleoptera (Insecta).** *ZooKeys.* 2011; (88): 1–972.
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

Cate PC: **Family Elateridae Leach, 1815.** In: Löbl, I. and Smetana, A. (eds.) *Catalogue of Palaearctic Coleoptera. Volume 4. Elateroidea - Derodontoidea – Bostrichoidea - Lymexyloidea - Cleroidea - Cucujoidea.* Stenstrup: Apollo Books, 2007; **2**: 89–209.

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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

Costa C, Lawrence CF, Policena Rosa S: **4.7. Elateridae Leach, 1815.** In: Leschen, R. A. B., Beutel, R. G., and Lawrence, J. F. (eds.) *Coleoptera, Beetles. Morphology and Systematics (Elateroidea, Bostrichiformia, Cucujiformia partim).* Berlin & New York: De Gruyter, 2010; **2**: 75–103.
Publisher 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

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Duff AG: **Beetles of Britain and Ireland. Geotrupidae to Scraptiidae.** West Runton: A.G. Duff, 2020; **3**.

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

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

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

Hyman PS: **No. 3. A review of the scarce and threatened Coleoptera of Great Britain. Part 1.** Peterborough: The UK Joint Nature Conservation Committee, 1992.
Reference Source

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

Koch K: **Die Käfer Mitteleuropas. Ökologie.** Krefeld: Goecke & Evers, 1992; **3**.
Reference Source

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

Mannerkoski I, Hyvärinen E, Alexander KNA, *et al.*: ***Ampedus sanguinolentus* (Europe assessment)**. 2010.

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

Mendel H: **Click-beetles of Great Britain and Ireland. An Atlas and Natural History.** Wallingford: Biological Records Centre, 2024.
Reference Source

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

Nikitsky NB, Osipov IN, Chemeris MV, *et al.*: **The beetles of the Prioksko-Terrasny Biosphere reserve - xylobiontes, mycetobiontes, and Scarabaeidae (with the review of the Moscow region fauna of the groups).** *Archives of the Zoological Museum Moscow State University.* 1996; **36**: 1–198.

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

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

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

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
PubMed Abstract | Publisher Full Text | Free Full Text

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.*: **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

Shirt DB, (ed.): **British Red Data Books: 2. Insects.** Peterborough: Nature Conservancy Council, 1987.
Reference Source

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

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

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 June 2025

https://doi.org/10.21956/wellcomeopenres.26160.r119541

✔ **Erich D Jarvis**
Rockefeller University, Millbrook, USA

Everything about the article looks great. The only issue I have is the Hi-C contact map of Figure 5. It does not look like a usual map, where you can see boundaries between chromosomes more cleaner than this. What are the two perpandicular white lines? Are you sure the settings are good? It would be useful to label the chromosomes.

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: Genomics and Neuroscience

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

Reviewer Report 24 March 2025

https://doi.org/10.21956/wellcomeopenres.26160.r120126

Ruiqi Li
University of Colorado Boulder, Boulder, Colorado, USA

Sivell presents a high-quality genome of the click beetle, *Ampedus sanguinolentus sanguinolentus.* I only have a few minor comments.

1. The introduction did a great job describing click beetle's ecology, distribution, biology, etc. However, it would be great if the authors could provide another paragraph discussing what biological question we can use this genome to answer. See Li *et al* 2024

2. Figure1. It would be better if the authors could add a scale on the photo.

References

1. Li R, Li J, Lopez JV, Oatley G, et al.: The genome sequence of the smooth giant clam, Tridacna derasa Röding, 1798.*Wellcome Open Res*. 2024; **9**: 375 PubMed Abstract | Publisher Full Text
2. Li R, Li J, Lemer S, Lopez JV, et al.: The genome sequence of a heart cockle, Fragum fragum (Linnaeus, 1758).*Wellcome Open Res*. 2024; **9**: 129 PubMed Abstract | Publisher Full Text

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics

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 06 March 2025

https://doi.org/10.21956/wellcomeopenres.26160.r120118

Hume Douglas
Agriculture and Agri-Food Canada, Ottawa, Canada

-Tribe should be spelled Ampedini
-Not morphologically homogeneous especially after recognition of various groups of soft-bodied elateridae
-I do not know of any evidence that adults mostly eat plants.
-Still 17 extant subfamilies according to Bouchard et al.2024 (type genera), but not the same subfamilies now. This is a more current classification
-What is an aberrant specimen?
-Please mention that this genome could help with understanding elaterid phylogenetics and pest biology
-iNaturalist suggests a strong peak of adult activity in May (https://www.inaturalist.org/taxa/491379-Ampedus-sanguinolentus)
-It would help to have a bit more explanation of Fig. 5.
This will be a useful contribution to beetle biology.
Thanks,
Hume Douglas

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: Elateridae systematics

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.