DATA NOTE

The genome sequence of the Deathwatch beetle, *Xestobium rufovillosum* (De Geer, 1774)

[version 1; peer review: 2 approved]

Eliot E. Jefferys[1], Peter W. H. Holland[2], Paula Thomas[1], Martin Hugman[1],
University of Oxford and Wytham Woods 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]National Museum of the Royal Navy, Portsmouth Historic Dockyard, HM Naval Base PP66, Portsmouth, England, PO1 3NH, UK
[2]Department of Biology, University of Oxford, Oxford, England, OX1 3SZ, UK

v1 **First published:** 21 Oct 2024, **9**:619
https://doi.org/10.12688/wellcomeopenres.23210.1
Latest published: 21 Oct 2024, **9**:619
https://doi.org/10.12688/wellcomeopenres.23210.1

Abstract

We present a genome assembly from an individual female *Xestobium rufovillosum* (the Deathwatch beetle; Arthropoda; Insecta; Coleoptera; Ptinidae). The genome sequence has a total length of 475.50 megabases. Most of the assembly is scaffolded into 13 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 19.19 kilobases in length.

Keywords

Xestobium rufovillosum, Deathwatch beetle, genome sequence, chromosomal, Coleoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 21 Oct 2024	✓ view	✓ view

1. **Arun Arumugaperumal**, Rajalakshmi Engineering College, Chennai, India

2. **Chenyang Cai**, Chinese Academy of Sciences, Beijing, China

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: Jefferys EE: Conceptualization, Resources, Writing – Review & Editing; **Holland PWH**: Conceptualization, Resources, Writing – Original Draft Preparation; **Thomas P**: Resources; **Hugman M**: Resources;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] 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: Jefferys EE, Holland PWH, Thomas P *et al.* **The genome sequence of the Deathwatch beetle, *Xestobium rufovillosum* (De Geer, 1774) [version 1; peer review: 2 approved]** Wellcome Open Research 2024, **9**:619 https://doi.org/10.12688/wellcomeopenres.23210.1

First published: 21 Oct 2024, **9**:619 https://doi.org/10.12688/wellcomeopenres.23210.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Bostrichiformia; Bostrichoidea; Ptinidae; Ernobiinae; *Xestobium*; *Xestobium rufovillosum* (De Geer, 1774) (NCBI:txid879067).

Background

The Deathwatch beetle *Xestobium rufovillosum* (Coleoptera: Anobiidae) is widespread in historic wooden buildings in Britain and across Europe. The beetle gained its common name from a superstition founded on the rhythmic tapping sound made by the beetles. Echoing around the walls of a silent house the mysterious sounds were once thought to be a harbinger of death, perhaps the sound of the grim reaper tapping with impatience (Birch & Menendez, 1991). In reality, the noise is made by the adult beetles drumming their head against wood as a method of intraspecific communication between sexes (Birch & Keenlyside, 1991). The larval stages, which may last for several years, form extensive feeding tunnels and cause structural damage to oak timbers, especially those already subjected to infection by wood-decaying fungi (Fisher, 1937). Experiments have shown that female *X. rufovillosum* are drawn to odours released by wood-decaying fungi, favouring oviposition in timbers or decaying trees most affected by these fungi (Belmain *et al.*, 2002). The genetic basis of this behaviour and the nature of the odorant receptors involved are unknown. Much also remains to be learned concerning the relationship between beetle physiology and the fungi and bacteria present in decaying wood.

Here we report the complete genome sequence of *X. rufovillosum* collected as an adult beetle from the timbers of the historic warship HMS Victory. The genome sequences of several cobiont or contaminating species of fungi and bacteria were also obtained. HMS Victory was launched in 1765 and survived the Battle of Trafalgar in 1805 where she served as Nelson's flagship. The ship has been in dry dock since 1922 which may have exposed timbers to an increased risk of damage by fungi and wood-boring beetles, through cross-contamination from other ships or adjoining historic buildings, introduction of timbers for repair, and high humidity in enclosed spaces (Davis, 2024). Records of *X. rufovillosum* in the oak timbers of HMS Victory date back to at least 1932, since when the beetles have survived sustained eradication and control measures (Davis, 2024). A complete genome sequence of *X. rufovillosum* will facilitate research into biochemical adaptations for wood-feeding, the relationship with wood-decaying fungi, and the physiology and population dynamics of an important pest species.

Genome sequence report

The genome of an adult female *Xestobium rufovillosum* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 26.43 Gb (gigabases) from 3.34 million reads, providing approximately 54-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 82.65 Gb from 547.32 million reads, yielding an approximate coverage of 174-fold. Specimen and sequencing information is summarised in Table 1.

Figure 1. Photograph of the *Xestobium rufovillosum* (icXesRufo2) specimen used for genome sequencing.

Manual assembly curation corrected 63 missing joins or misjoins and 15 haplotypic duplications, reducing the assembly length by 1.12% and the scaffold number by 58.23%, and increasing the scaffold N50 by 13.64%. The final assembly has a total length of 475.50 Mb in 32 sequence scaffolds, with 111 gaps, and a scaffold N50 of 36.9 Mb (Table 2). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.42%) of the assembly sequence was assigned to 13 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 3). Genome is from the homogametic sex, but no X chromosome could be identified. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 65.0 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.4.3 completeness of 99.1% (single = 98.6%, duplicated = 0.5%), using the endopterygota_odb10 reference set (n = 2,124).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/879067.

Methods

Sample acquisition and DNA barcoding

Specimens of *Xestobium rufovillosum* were collected by hand from HMS Victory, Portsmouth Historic Dockyard, Hampshire, United Kingdom (latitude 50.8, longitude –1.11)

Table 1. Specimen and sequencing data for *Xestobium rufovillosum*.

Project information			
Study title	Xestobium rufovillosum (deathwatch beetle)		
Umbrella BioProject	PRJEB70741		
Species	*Xestobium rufovillosum*		
BioSample	SAMEA113426066		
NCBI taxonomy ID	879067		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	icXesRufo2	SAMEA113427020	Whole organism
Hi-C sequencing	icXesRufo1	SAMEA113427014	Whole organism
RNA sequencing	icXesRufo1	SAMEA113427014	Whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Illumina NovaSeq 6000 (Hi-C)	ERR12342490	5.47e+08	82.65
PacBio Revio	ERR12340376	3.34e+06	26.43
Illumina NovaSeq 6000 (RNA)	ERR12342491	5.36e+07	8.1

on 2023-05-04 by Eliot Jefferys and Peter Holland (University of Oxford). Eliot Jefferys formally identified the specimens, which were preserved directly at –80 °C. The genome sequence was based on specimen ID Ox003377 (ToLID icXesRufo2) for PacBio HiFi sequencing. A second specimen (ID Ox003366, ToLID icXesRufo1) was used for Hi-C sequencing and for RNA sequencing (ID Ox003366, ToLID icXesRufo1).

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 specimens and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI). 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).

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 core procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The icXesRufo2 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 Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from whole organism tissue of icXesRufo1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit.

Table 2. Genome assembly data for *Xestobium rufovillosum*, icXesRufo2.2.

Genome assembly		
Assembly name	icXesRufo2.2	
Assembly accession	GCA_963966045.2	
Accession of alternate haplotype	*GCA_963966065.2*	
Span (Mb)	475.50	
Number of contigs	144	
Contig N50 length (Mb)	7.3	
Number of scaffolds	32	
Scaffold N50 length (Mb)	36.9	
Longest scaffold (Mb)	53.6	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	65.0	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:99.1%[S:98.6%,D:0.5%], F:0.5%,M:0.4%,n:2,124	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.42%	*≥ 95%*
Sex chromosomes	Not identified	*localised homologous pairs*
Organelles	Mitochondrial genome: 19.19 kb	*complete single alleles*

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from Rhie *et al.* (2021).

** BUSCO scores based on the endopterygota_odb10 BUSCO set using version 5.4.3. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Xestobium_rufovillosum/dataset/GCA_963966045.2/busco.

Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Hi-C preparation

Tissue from the icXesRufo1 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation.

Library preparation and sequencing

Pacific Biosciences SMRTbell libraries were constructed using the Revio HiFi prep kit, according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences Revio instrument.

Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq 6000 instrument.

For Hi-C library preparation, DNA was fragmented to a size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit following manufacturers' instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications

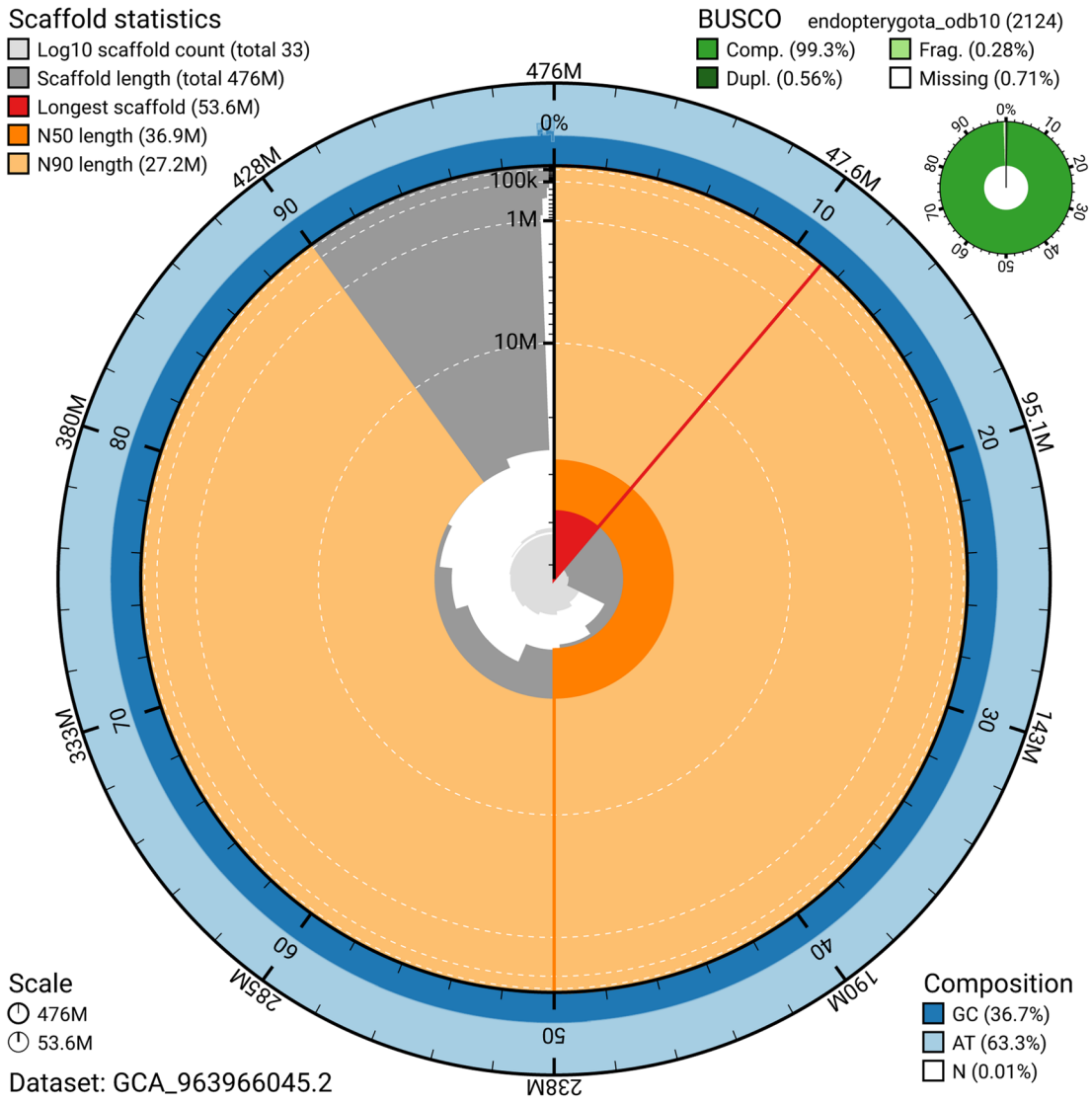

Figure 2. Genome assembly of *Xestobium rufovillosum*, icXesRufo2.2: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 475,562,900 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (53,601,044 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (36,916,890 and 27,162,678 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the endopterygota_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963966045.2/dataset/GCA_963966045.2/snail.

were identified and removed using purge_dups (Guan *et al.*, 2020). 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. 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

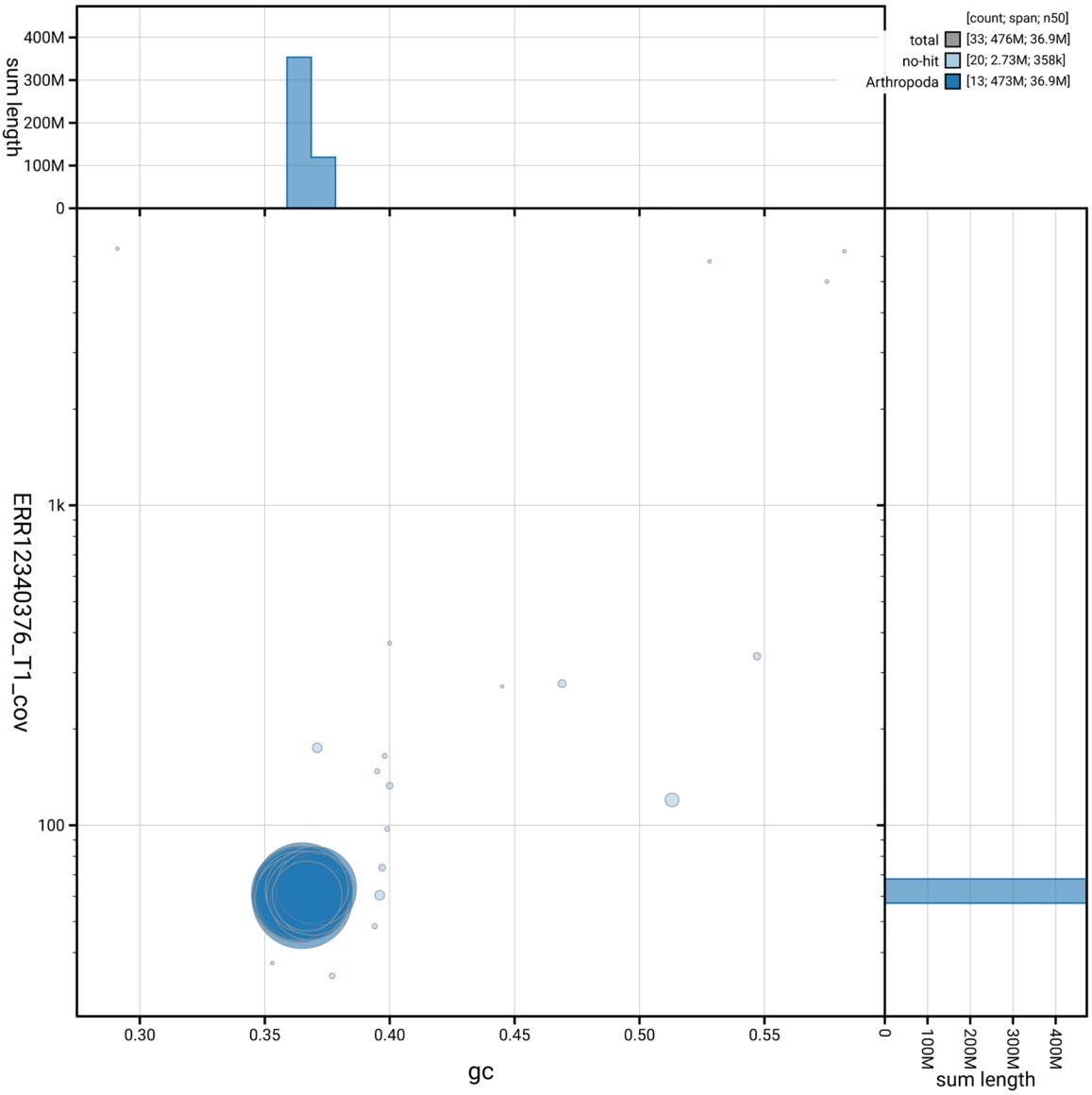

Figure 3. Genome assembly of *Xestobium rufovillosum*: Blot plot of base coverage against GC proportion for sequences in the icXesRufo2.2 assembly. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963966045.2/dataset/GCA_963966045.2/blob.

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

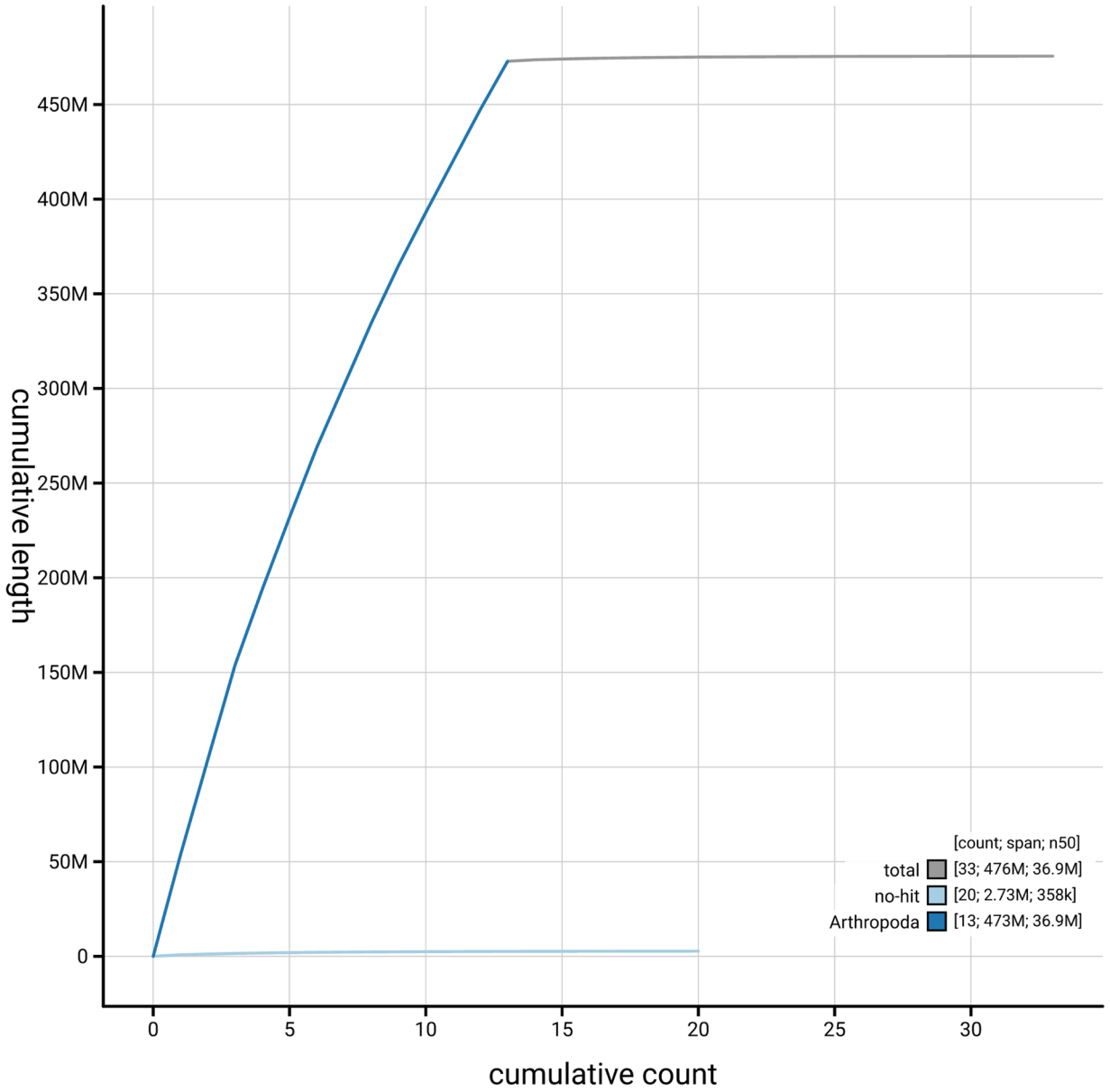

Figure 4. Genome assembly of *Xestobium rufovillosum* icXesRufo2.2: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963966045.2/dataset/GCA_963966045.2/cumulative.

were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

Evaluation of the final assembly

The final assembly was post-processed and evaluated using the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines: sanger-tol/readmapping (Surana *et al.*, 2023a), sanger-tol/genomenote (Surana *et al.*, 2023b), and sanger-tol/blobtoolkit (Muffato *et al.*, 2024). The readmapping pipeline aligns the Hi-C reads using bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The genomenote pipeline converts the Hi-C alignments 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). This pipeline also generates assembly statistics using the NCBI datasets report (Sayers *et al.*, 2024), computes *k*-mer

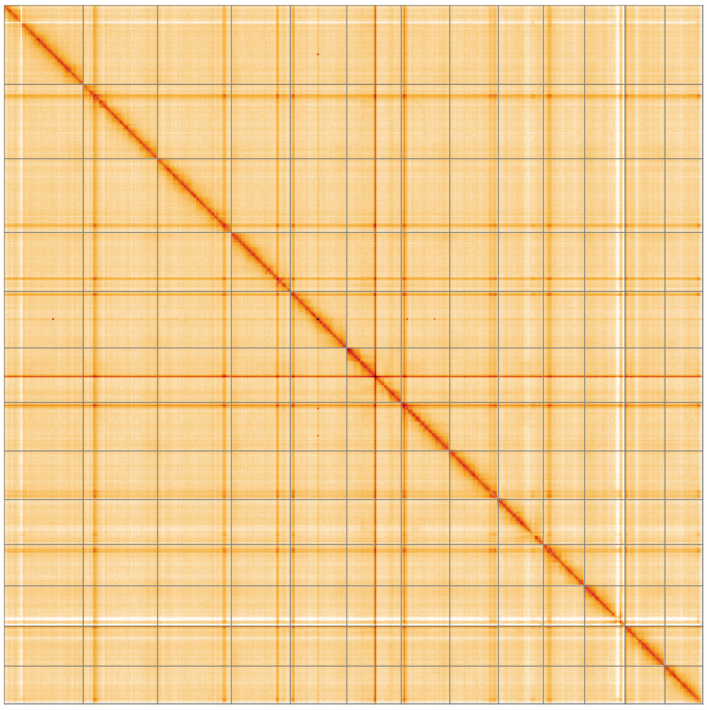

Figure 5. Genome assembly of *Xestobium rufovillosum* icXesRufo2.2: Hi-C contact map of the icXesRufo2.2 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=fArA1obsR7iQaEZssxsq3w.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Xestobium rufovillosum*, icXesRufo2.

INSDC accession	Name	Length (Mb)	GC%
OZ014492.1	1	53.6	36.5
OZ014493.1	2	50.24	36.5
OZ014494.1	3	49.95	36.5
OZ014495.1	4	40.08	36.5
OZ014496.1	5	38.02	36.5
OZ014497.1	6	36.92	37.0
OZ014498.1	7	32.89	36.5
OZ014499.1	8	32.86	37.0
OZ014500.1	9	30.42	36.5
OZ014501.1	10	28.02	37.0
OZ014502.1	11	27.21	37.0
OZ014503.1	12	27.16	37.0
OZ014504.1	13	25.47	36.5
OZ014505.1	MT	0.02	29.5

completeness and QV consensus quality values with FastK and MERQURY.FK, and runs BUSCO (Manni *et al.*, 2021) to assess completeness.

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 (Buchfink *et al.*, 2021) blastp. The genome is also split 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 with DIAMOND blastx. Genome sequences without a hit are chunked with 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 genome assembly and evaluation pipelines were 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 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

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	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 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-r587	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
Merqury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
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	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/genomenote	1.1.1	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.2.1	https://github.com/sanger-tol/readmapping
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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: *Xestobium rufovillosum* (death-watch beetle). Accession number PRJEB70741; https://identifiers.org/ena.embl/PRJEB70741 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Xestobium rufovillosum* 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 University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.12157525.

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.

Acknowledgements

We would like to thank Rachel Trembath for assistance with maintenance of specimens.

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

Belmain SR, Simmonds MSJ, Blaney WM: **Influence of odor from wood-decaying fungi on host selection behavior of deathwatch beetle *Xestobium rufovillosum*.** *J Chem Ecol.* 2002; **28**(4): 741–754.
PubMed Abstract | Publisher Full Text

Birch MC, Keenlyside JJ: **Tapping behavior is a rhythmic communication in the death-watch beetle, *Xestobium rufovillosum* (Coleoptera: Anobiidae).** *J Insect Behav.* 1991; **4**(2): 257–263.
Publisher Full Text

Birch M, Menendez G: **Knocking on wood for a mate.** *New Sci.* 1991; **131**: 42–44.
Reference Source

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

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

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

Davis D: **Wood-borers in historic ships.** *Environ Archaeol.* 2024; **29**(1): 71–79.
Publisher 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

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): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

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

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, 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

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

Muffato M, Butt Z, Challis R, *et al.*: **Sanger-tol/blobtoolkit: v0.3.0 – poliwag**. 2024.
Publisher Full Text

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

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

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

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

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

Sayers EW, Cavanaugh M, Clark K, *et al.*: **GenBank 2024 update.** *Nucleic Acids Res.* 2024; **52**(D1): D134–D137.
PubMed Abstract | Publisher Full Text | Free Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
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.
Publisher 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

Wellcome Sanger Institute: **The genome sequence of the Deathwatch beetle, *Xestobium rufovillosum* (De Geer, 1774).** European Nucleotide Archive. [dataset], accession number PRJEB70741, 2023.

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 07 October 2025

https://doi.org/10.21956/wellcomeopenres.25562.r133737

✔ **Chenyang Cai**
Chinese Academy of Sciences, Beijing, Beijing, China

This manuscript presents a high-quality genome assembly of the Deathwatch beetle *Xestobium rufovillosum*, an ecologically and economically important wood-boring beetle with cultural and historical relevance due to its association with historic buildings and ships, such as HMS Victory. The study is part of the Darwin Tree of Life project and follows their standardized workflow for specimen acquisition, sequencing, assembly, and curation.
This manuscript is well-structured, and provides an important genomic resource for future research on beetle evolution, wood-feeding adaptations, pest control. The genome assembly is of chromosome-level completeness, with excellent metrics (BUSCO completeness 99.3% and the mitochondrial genome is also provided. The methods are described in great detail and include transparent data processing and curation pipelines. The manuscript is appropriate for publication as a genome report. It is better to list the subfamily which the species belongs to.

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: beetles. 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 07 November 2024

https://doi.org/10.21956/wellcomeopenres.25562.r108066

Arun Arumugaperumal
Department of Biotechnology, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India

This is the first published article on the genome sequence of the Deathwatch beetle *Xestobium rufovillosum*. The genome assembly size was reported as 475.5 Mb. The mitogenome was also assembled and the length was 19.9 kb. The methodology used was very apt to the aim of the data article. The genome sequence was evaluated to be a near-complete one. This resource can be used to carry out further research on the organism.

The authors could have given a panel of different views of the specimen instead of a single photograph. The authors could have added a note on the number of protein-coding genes predicted. It was mentioned that there were some symbiotic bacteria and other organism sequences. A link to the data associated with those sequences could be useful for readers in taking forward the research about this beetle.

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: Bioinformatics, 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.