




DATA NOTE

**REVISED** **The genome sequence of a dung beetle, *Aphodius (Calamosternus) granarius* Linnaeus, 1767**

[version 2; peer review: 2 approved]

Darren J. Mann<sup>1</sup>, Liam M. Crowley <sup>2</sup>,  
 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

<sup>1</sup>Oxford University Museum of Natural History, Oxford, England, UK<sup>2</sup>University of Oxford, Oxford, England, UK

**V2** First published: 28 Jul 2025, **10**:369  
<https://doi.org/10.12688/wellcomeopenres.24306.1>  
 Latest published: 04 Sep 2025, **10**:369  
<https://doi.org/10.12688/wellcomeopenres.24306.2>

**Abstract**

We present a genome assembly from a female specimen of *Aphodius granarius* (dung beetle; Arthropoda; Insecta; Coleoptera; Scarabaeidae). The genome sequence has a total length of 397.20 megabases. The assembly is scaffolded into 10 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 22.13 kilobases.



**Keywords**



*Aphodius granarius*, dung beetle, genome sequence, chromosomal, Coleoptera



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

**Open Peer Review****Approval Status**  

	1	2
<b>version 2</b> (revision) 04 Sep 2025		 <a href="#">view</a>
<b>version 1</b> 28 Jul 2025	 <a href="#">view</a>	

1. **Manuel Barrios-Izas** , University of San Carlos of Guatemala, Zacapa, Guatemala
2. **Benjamin James Hunt** , University of Exeter, Exeter, UK

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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** Mann DJ: Investigation, Resources; Crowley LM: Investigation, Resources;

**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 Mann DJ *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:** Mann DJ, Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of a dung beetle, *Aphodius (Calamosternus) granarius* Linnaeus, 1767 [version 2; peer review: 2 approved]** Wellcome Open Research 2025, 10:369 <https://doi.org/10.12688/wellcomeopenres.24306.2>

**First published:** 28 Jul 2025, 10:369 <https://doi.org/10.12688/wellcomeopenres.24306.1>

**REVISED Amendments from Version 1**

In the new version of this data note we have provided a labelled diagram of the Hi-C contact map to make it easier to interpret. We have also added information about previous sequencing of the family Scarabaeidae.

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

**Species taxonomy**

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Scarabaeiformia; Scarabaeoidea; Scarabaeidae; Aphodiinae; *Aphodius*; *Calamosternus*; *Aphodius granarius* Linnaeus, 1767 (NCBI: txid207157)

**Background**

The genome of a dung beetle, *Aphodius granarius*, was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *Aphodius granarius*, based on a specimen from Wytham Woods, Oxfordshire, United Kingdom (Figure 1). Although numerous genomes exist for the family Scarabaeidae, this assembly provides the first chromosomally complete sequence for *Aphodius granarius*, enabling comparative analyses (data obtained via NCBI datasets, O'Leary *et al.*, 2024).

**Genome sequence report****Sequencing data**

The genome of a specimen of *Aphodius granarius* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 23.28 Gb from 2.77 million reads, which were used to assemble the genome. GenomeScope analysis estimated the haploid genome size at 342.83 Mb, with a heterozygosity



**Figure 1.** Photograph of the *Aphodius granarius* (icAphGran1) specimen used for genome sequencing.

of 0.89% and repeat content of 52.37%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 64x coverage. Hi-C sequencing produced 94.41 Gb from 625.24 million reads, used to scaffold the assembly. RNA sequencing data were also generated and are available in 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 34 misjoins or missing joins and removed three haplotypic duplications. These interventions decreased the scaffold count by 2.31% and increased the scaffold N50 by 67.92%. The final assembly has a total length of 397.20 Mb in 590 scaffolds, with 69 gaps, and a scaffold N50 of 24.5 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 (69.36%) was assigned to 10 chromosomal-level scaffolds, representing 9 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, Chromosome X identified via synteny to the genome of *Cetonia aurata* (GCA\_949128085.1).

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 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 60.2. The *k*-mer recovery for the primary haplotype is 80.96%, and for the alternate haplotype 75.94%; the combined primary and alternate assemblies have a *k*-mer recovery of 96.92%. BUSCO v.5.5.0 analysis using the endopterygota\_odb10 reference set ( $n = 2,124$ ) identified 99.6% of the expected gene set (single = 98.3%, duplicated = 1.3%).

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on

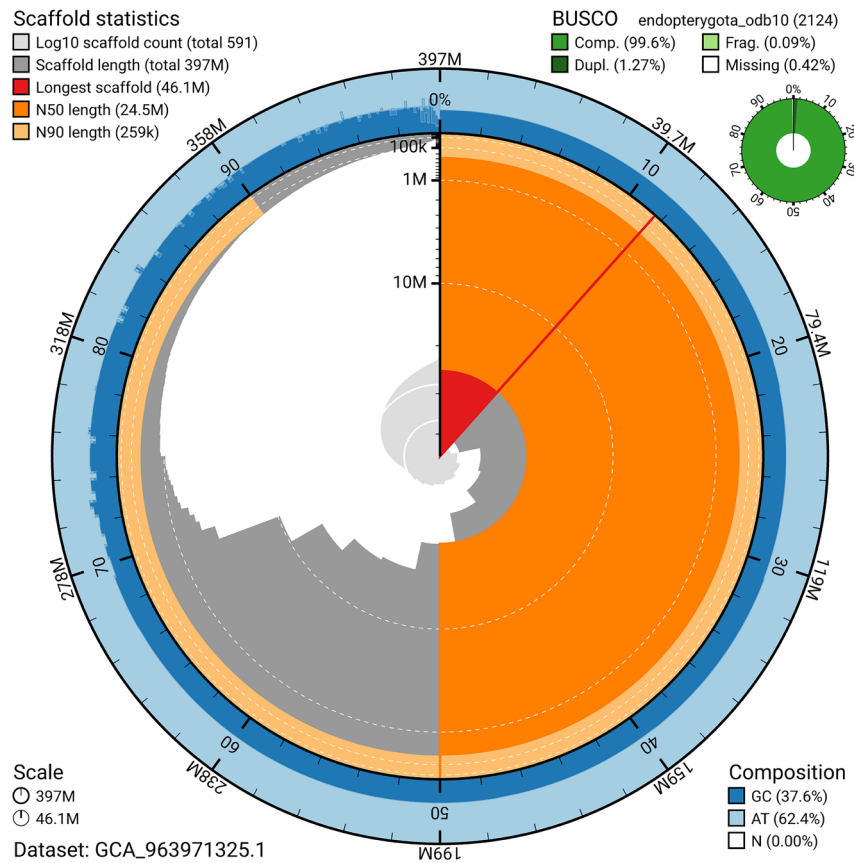
**Table 1. Specimen and sequencing data for *Aphodius granarius*.**

Project information			
Study title	Aphodius granarius		
Umbrella BioProject	PRJEB61913		
Species	<i>Aphodius granarius</i>		
BioSpecimen	SAMEA112232679		
NCBI taxonomy ID	207157		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	icAphGran1	SAMEA112233156	whole organism
Hi-C sequencing	icAphGran1	SAMEA112233156	whole organism
RNA sequencing	icAphGran3	SAMEA112233158	whole organism
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR11439643	6.25e+08	94.41
PacBio Sequel IIe	ERR11435988	2.77e+06	23.28
RNA Illumina NovaSeq X	ERR13093631	1.93e+08	29.21

**Table 2. Genome assembly data for *Aphodius granarius*.**

Genome assembly		
Assembly name	icAphGran1.1	
Assembly accession	GCA_963971325.1	
Alternate haplotype accession	GCA_963971265.1	
Assembly level for primary assembly	chromosome	
Span (Mb)	397.20	
Number of contigs	659	
Number of scaffolds	590	
Longest scaffold (Mb)	46.1	
Assembly metric	Measure	Benchmark
Contig N50 length	4.08 Mb	$\geq 1$ Mb
Scaffold N50 length	24.5 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.7; alternate: 60.6; combined: 60.2	$\geq 40$
k-mer completeness	Primary: 80.96%; alternate: 75.94%; combined: 96.92%	$\geq 95\%$
BUSCO*	C:99.6%[S:98.3%,D:1.3%], F:0.1%,M:0.3%,n:2,124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	69.36%	$\geq 90\%$
Sex chromosomes	X	localised homologous pairs
Organelles	Mitochondrial genome: 22.13 kb	complete single alleles

\* 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.



**Figure 2. Genome assembly of *Aphodius granarius*, icAphGran1.1: 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\\_963971325.1/dataset/GCA\\_963971325.1/snail](https://blobtoolkit.genomehubs.org/view/GCA_963971325.1/dataset/GCA_963971325.1/snail).

Assembly Standards [September 2024](#). The assembly achieves the EBP reference standard of **6.7.Q59**.

## Methods

### Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Aphodius granarius* (specimen ID Ox002476, ToLID icAphGran1), collected from Wytham Farm, Berkshire, United Kingdom (latitude 51.783, longitude -1.318) on 2022-06-13 by potting. Another specimen collected on the same occasion was used for RNA sequencing (specimen ID Ox002478, ToLID icAphGran3). The specimens were collected by Darren Mann and Liam Crowley, identified by Darren Mann and preserved on dry ice.

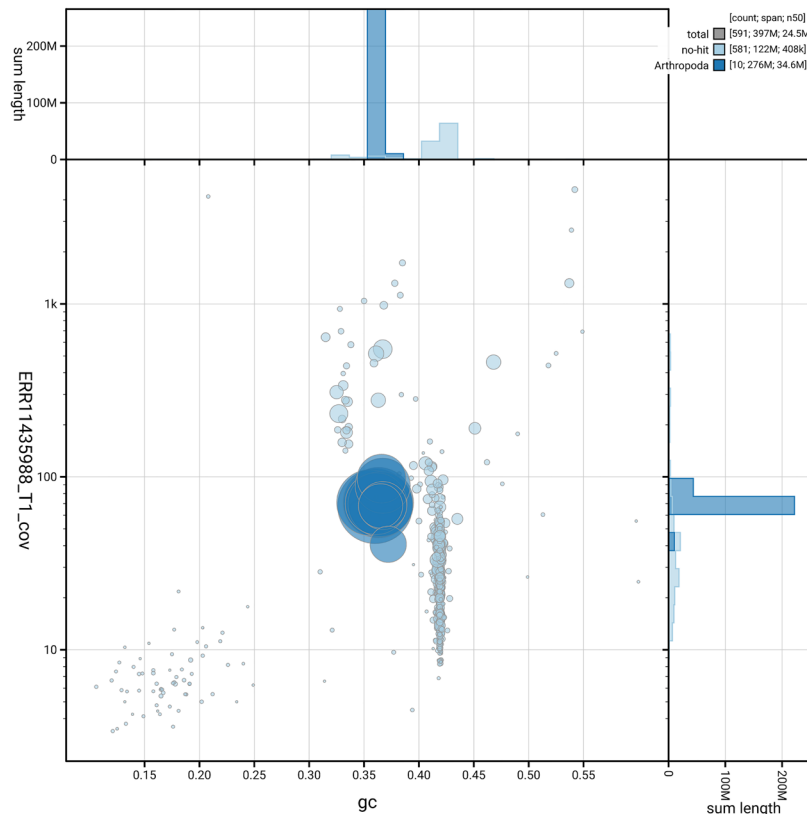
The initial identification was verified by an additional DNA barcoding process according to the framework developed by [Twyford et al. \(2024\)](#). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (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](#)).

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 icAphGran1 sample was prepared for DNA



**Figure 3. Genome assembly of *Aphodius granarius*, icAphGran1.1: 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\\_963971325.1/dataset/GCA\\_963971325.1/blob](https://blobtoolkit.genomehubs.org/view/GCA_963971325.1/dataset/GCA_963971325.1/blob).

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 in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The 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 manual solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). 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 icAphGran3 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. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

#### Hi-C sample preparation and crosslinking

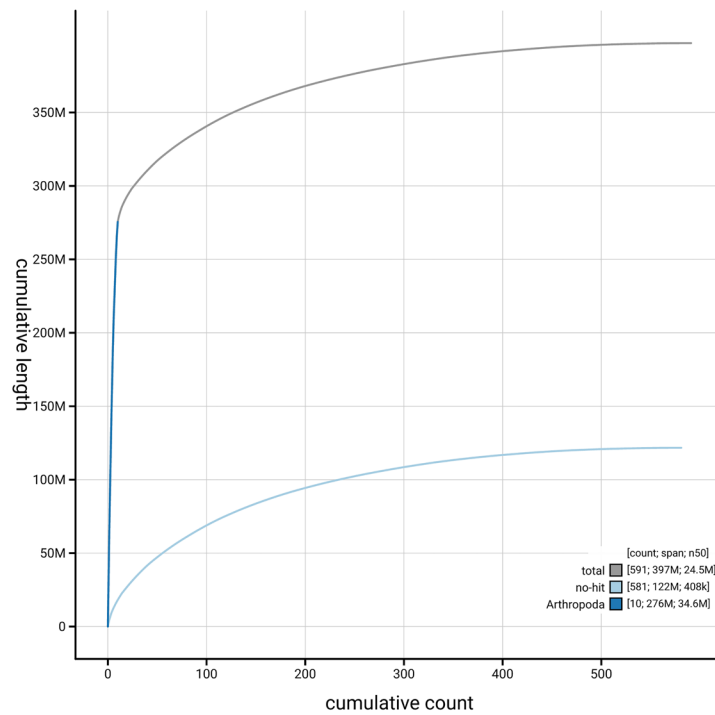
Hi-C data were generated from the whole organism of the icAphGran1 sample using the Arima-HiC v2 kit (Arima Genomics) with 20–50 mg of frozen tissue (stored at  $-80^{\circ}\text{C}$ ). As per manufacturer's instructions, tissue was fixed, and the DNA crosslinked using a TC buffer with a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagnocine 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

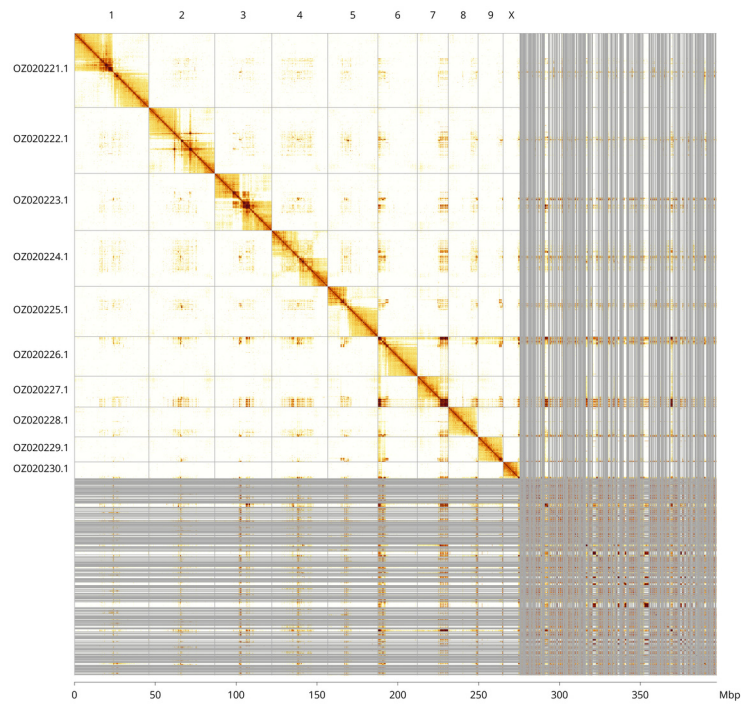
Library preparation and sequencing were performed at the WSI Scientific Operations core.

#### *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



**Figure 4. Genome assembly of *Aphodius granarius*, icAphGran1.1: 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\\_963971325.1/dataset/GCA\\_963971325.1/cumulative](https://blobtoolkit.genomehubs.org/view/GCA_963971325.1/dataset/GCA_963971325.1/cumulative).



**Figure 5. Genome assembly of *Aphodius granarius*: Hi-C contact map of the icAphGran1.1 assembly, visualised using PretextSnpshot.** Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure in HiGlass may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=E19CqaImR6afAxo-Vfj28g>.

**Table 3. Chromosomal pseudomolecules in the genome assembly of *Aphodius granarius*, icAphGran1.**

INSDC accession	Name	Length (Mb)	GC%
OZ020221.1	1	46.1	36
OZ020222.1	2	40.72	36.5
OZ020223.1	3	35.26	35.5
OZ020224.1	4	34.56	36.5
OZ020225.1	5	31.08	36
OZ020226.1	6	24.5	36.5
OZ020227.1	7	19.03	36.5
OZ020228.1	8	18.46	36.5
OZ020229.1	9	15.54	36.5
OZ020230.1	X	10.27	37
OZ020231.1	MT	0.02	21

Biosciences), depending on genome size and sequencing depth 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 the Sequel IIE system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIE was in the range 40–135 pM. 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, 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 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.

### RNA

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

### Genome assembly, 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 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. 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. Sex chromosomes were identified by synteny analysis. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>.

#### 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 ( $k = 31$ ) 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 was visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The genome was analysed in the blobtoolkit pipeline, a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the Pac-Bio reads in SAMtools (Danecek *et al.*, 2021) 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	<a href="https://github.com/arq5x/bedtools2">https://github.com/arq5x/bedtools2</a>
BLAST	2.14.0	<a href="ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/">ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/</a>
BlobToolKit	4.3.9	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.5.0	<a href="https://gitlab.com/e2lab/busco">https://gitlab.com/e2lab/busco</a>
bwa-mem2	2.2.1	<a href="https://github.com/bwa-mem2/bwa-mem2">https://github.com/bwa-mem2/bwa-mem2</a>
Cooler	0.8.11	<a href="https://github.com/open2c/cooler">https://github.com/open2c/cooler</a>
DIAMOND	2.1.8	<a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>
fasta_windows	0.2.4	<a href="https://github.com/tolkit/fasta_windows">https://github.com/tolkit/fasta_windows</a>
FastK	666652151335353eef2fcd58880bcef5bc2928e1	<a href="https://github.com/thegenemyers/FASTK">https://github.com/thegenemyers/FASTK</a>
Gfastats	1.3.6	<a href="https://github.com/vgl-hub/gfastats">https://github.com/vgl-hub/gfastats</a>
GoAT CLI	0.2.5	<a href="https://github.com/genomehubs/goat-cli">https://github.com/genomehubs/goat-cli</a>
Hifiasm	0.16.1-r375	<a href="https://github.com/chhylyp123/hifiasm">https://github.com/chhylyp123/hifiasm</a>
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
MercuryFK	d00d98157618f4e8d1a9190026b19b471055b22e	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>
Minimap2	2.24-r1122	<a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>
MitoHiFi	3	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
MultiQC	1.14, 1.17, and 1.18	<a href="https://github.com/MultiQC/MultiQC">https://github.com/MultiQC/MultiQC</a>
Nextflow	23.04.1	<a href="https://github.com/nextflow-io/nextflow">https://github.com/nextflow-io/nextflow</a>
PretextView	0.2.5	<a href="https://github.com/sanger-tol/PretextView">https://github.com/sanger-tol/PretextView</a>
purge_dups	1.2.5	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
samtools	1.19.2	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
sanger-tol/ascc	0.1.0	<a href="https://github.com/sanger-tol/ascc">https://github.com/sanger-tol/ascc</a>
sanger-tol/blobtoolkit	0.4.0	<a href="https://github.com/sanger-tol/blobtoolkit">https://github.com/sanger-tol/blobtoolkit</a>
Seqtk	1.3	<a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>
Singularity	3.9.0	<a href="https://github.com/sylabs/singularity">https://github.com/sylabs/singularity</a>
TreeVal	1.2.0	<a href="https://github.com/sanger-tol/treeval">https://github.com/sanger-tol/treeval</a>
YaHS	1.2a.2	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

## 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: *Aphodius granarius*. Accession number PRJEB61913; <https://identifiers.org/ena.embl/PRJEB61913>.

The genome sequence is released openly for reuse. The *Aphodius granarius* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). 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>.

---

## 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: diagenome Megaruptor<sup>®</sup>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](#)
- 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](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCftools.** *GigaScience.* 2021; **10**(2): gjab008.  
[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, Denton A, Yatsenko H, *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): gjaa153.  
[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, Pereira-da-Conceicao LL, Blaxter ML, *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](#)
- Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.  
[Publisher Full Text](#)
- O’Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the Tree of Life with NCBI datasets.** *Sci Data.* 2024; **11**(1): 732.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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.*: **Mercury: 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.* 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, 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 2

Reviewer Report 14 October 2025

<https://doi.org/10.21956/wellcomeopenres.27461.r132136>

© 2025 Hunt B. 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.



**Benjamin James Hunt** 

University of Exeter, Exeter, UK

Mann and co-authors report the genome of the dung beetle *Aphodius granarius*. Assembly was conducted using a high quality and well-established pipeline as part of the Darwin Tree of Life Project, with the full process and associated metadata documented in the manuscript. The assembly achieves a good QV score and is very complete and low in duplication levels on the basis of its BUSCO score. Scaffolding to chromosome level appears to have been less successful than many of the ToL projects, with just under 70% of the genome assigned to chromosome-scale scaffolds. The genome is not yet annotated but this is underway - it would perhaps have been preferable to wait until this was completed before publication.

Two minor formatting notes:

- \* In 'Background', *Aphodius granarius* is not italicised in the final sentence.
- \* In 'Data availability', the link to the ENA accession is broken.

**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:** Entomology, bioinformatics

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

---

**Version 1**

Reviewer Report 02 September 2025

<https://doi.org/10.21956/wellcomeopenres.26806.r130746>

© 2025 Barrios-Izas 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.



**Manuel Barrios-Izas** 

University of San Carlos of Guatemala, Zacapa, Guatemala

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

Yes, the authors describe this is a project that aim to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. The information will be useful for further research and is available on ENA.

Are the protocols appropriate and is the work technically sound?

Yes, the authors use standardized protocols from field through wet lab and bioinformatics lab.

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

- Figure 1, otherwise than db number, is a bad quality uninformative image.
- Please add the version of all the software used, as this will improve reproducibility.
- When you state that 'amplicons were sequenced and compared to the BOLD database, confirming the species identification', do you mean that you performed a BLAST search, or that you constructed a species tree?

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

All datasets are available at ENA PRJEB61913

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

Yes

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

Yes

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

Partly

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

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Entomology, bioinformatics, molecular biology

**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.**

---