



DATA NOTE

The genome sequence of a small dung beetle, *Volinus sticticus* (Panzer, 1798), formerly known as *Aphodius sticticus*

[version 1; peer review: 3 approved]

František Sládeček¹, Liam M. Crowley²,
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

¹Department of Zoology, University of South Bohemia, České Budějovice, Czech Republic

²Department of Biology, University of Oxford, Oxford, England, UK

v1 First published: 18 Apr 2025, 10:195
<https://doi.org/10.12688/wellcomeopenres.23692.1>
Latest published: 18 Apr 2025, 10:195
<https://doi.org/10.12688/wellcomeopenres.23692.1>

Abstract

We present a genome assembly from an individual male specimen of the small dung beetle, *Aphodius sticticus* (Arthropoda; Insecta; Coleoptera; Scarabaeidae). The genome sequence has a total length of 628.40 megabases. Most of the assembly (98.96%) is scaffolded into 11 chromosomal pseudomolecules, including the X and Y sex chromosomes. The mitochondrial genome has also been assembled and is 20.8 kilobases in length.

Keywords

Aphodius sticticus, small dung 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			
18 Apr 2025	view	view	view

1. **Hume Douglas** , Agriculture and Agri-Food Canada, Ottawa, Canada
2. **Lindsey C Perkin**, United States Department of Agriculture ARS, College Station, USA
3. **Doga Cedden** , University of Göttingen, Göttingen, France

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: Sládeček F: 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 [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.

Copyright: © 2025 Sládeček F *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: Sládeček F, Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of a small dung beetle, *Volinus sticticus* (Panzer, 1798), formerly known as *Aphodius sticticus* [version 1; peer review: 3 approved]** Wellcome Open Research 2025, 10:195 <https://doi.org/10.12688/wellcomeopenres.23692.1>

First published: 18 Apr 2025, 10:195 <https://doi.org/10.12688/wellcomeopenres.23692.1>

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*; *Volinus*; *Aphodius sticticus* (Panzer, 1798) (NCBI:txid207213)

Background

The subfamily Aphodiinae is a diverse group of beetles within the scarab beetle family, Scarabaeidae, and has a global distribution. While many aphodiines are dung beetles, some species are associated with decaying vegetable matter or burrows of other beetles, such as Geotrupidae ([British Scarabs, 2025](#)). In Britain and Ireland, 46 aphodiine species have been recorded ([Duff, 2018](#)). Different species of Aphodiinae commonly coexist within dung piles by occupying distinct ecological niches, including preferences for dung type, age, soil conditions, or exposure to sunlight, which helps to minimise competition among species ([British Scarabs, 2025](#)). Recent taxonomic revisions have reclassified many former members of the genus *Aphodius* into separate genera, reflecting advances in our understanding of their evolutionary relationships ([British Scarabs, 2025](#)). The species formerly known as *Aphodius sticticus* is now placed within the genus *Volinus*.

Volinus sticticus is a small dung beetle, measuring approximately 4.5–5.0 mm in length, with distinctive black streaks on yellow-brown elytra. It is considered a “dweller” spending its entire life cycle within dung ([British Scarabs, 2025](#)). It shows a preference for horse and sheep dung in shaded habitats.

Volinus sticticus is distributed across parts of the Palaearctic region, with records primarily from Europe. Most occurrences are concentrated in western and central Europe, with additional records in countries such as Turkey and Albania ([GBIF Secretariat, 2023](#)). It is fairly common in England and Wales and can occur in numbers where it is found ([NBN Atlas Partnership, 2024](#)).

We present a chromosome-level genome sequence for *Aphodius sticticus*, based on a male specimen from Blackdown Farm, Gloucestershire, United Kingdom ([Figure 1](#)).

Genome sequence report

The genome of *Aphodius sticticus* ([Figure 1](#)) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 22.88 Gb (gigabases) from 2.16 million reads, providing an estimated 36-fold coverage. Chromosome conformation Hi-C sequencing produced 120.19 Gb from 795.96 million reads. Specimen and sequencing details are summarised in [Table 1](#).

Assembly errors were corrected by manual curation, including 54 missing joins or mis-joins and 5 haplotypic duplications. This reduced the scaffold number by 9.3%. The final assembly has a total length of 628.40 Mb in 77 sequence scaffolds, with 283 gaps, and a scaffold N50 of 79.0 Mb ([Table 2](#)).



Figure 1. Photograph of the *Aphodius sticticus* (icAphStic4) specimen used for genome sequencing.

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 (98.96%) was assigned to 11 chromosomal-level scaffolds, representing 9 autosomes and the X and Y sex chromosomes. These chromosome-level scaffolds, confirmed by the Hi-C data, are named in order of size ([Figure 5](#); [Table 3](#)). During manual curation, chromosomes X and Y were assigned based on read coverage statistics.

While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to an alternate 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, and as a separate fasta file.

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 primary haplotype has a QV of 60.6, and the combined primary and alternate assemblies achieve an estimated QV of 60.1. The *k*-mer completeness for the primary haplotype is 80.97%, and for the alternate haplotype it is 77.35%, while the combined primary and alternate assemblies achieve a *k*-mer completeness of 99.07%. BUSCO analysis using the endopterygota_odb10 reference set ($n = 2,124$) indicated a completeness score of 99.4% (single = 97.8%, duplicated = 1.6%).

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

Project information			
Study title	Aphodius sticticus		
Umbrella BioProject	PRJEB70743		
Species	<i>Aphodius sticticus</i>		
BioSample	SAMEA112774809		
NCBI taxonomy ID	207213		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	icAphStic4	SAMEA112774903	Whole organism
Hi-C sequencing	icAphStic2	SAMEA112774901	Whole organism
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR12342492	7.96e+08	120.19
PacBio Revio	ERR12340114	2.16e+06	22.88

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.Q60.

Methods

Sample acquisition and DNA barcoding

Specimens of *Aphodius sticticus* were collected from Black-down Farm, Gloucestershire, United Kingdom (latitude 52.04, longitude -1.70) on 2022-08-25 by potting. The specimens were collected by František Sládeček (University of South Bohemia) and Liam Crowley (University of Oxford), identified by František Sládeček, and then preserved on dry ice. The specimen used for PacBio HiFi sequencing was specimen ID Ox002960 (ToLID icAphStic4), while another specimen (specimen ID Ox002958, ToLID icAphStic2) was used for Hi-C scaffolding of the assembly.

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 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 icAphStic4 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.

Table 2. Genome assembly data for *Aphodius sticticus*, icAphStic4.1.

Genome assembly		
Assembly name	icAphStic4.1	
Assembly accession	GCA_963966075.1	
Accession of alternate haplotype	GCA_963966085.1	
Span (Mb)	628.40	
Number of contigs	361	
Number of scaffolds	77	
Longest scaffold (Mb)	112.87	
Assembly metrics*		Benchmark
Contig N50 length (Mb)	4.8	$\geq 1 \text{ Mb}$
Scaffold N50 length (Mb)	79.0	= chromosome N50
Consensus quality (QV)	Primary: 60.6; alternate: 59.6; combined 60.1	≥ 40
k-mer completeness	Primary: 80.97%; alternate: 77.35%; combined: 99.07%	$\geq 95\%$
BUSCO**	C:99.3%[S:97.7%,D:1.6%], F:0.2%,M:0.5%,n:2,124	$S > 90\%$, $D < 5\%$
Percentage of assembly mapped to chromosomes	98.96%	$\geq 90\%$
Sex chromosomes	XY	localised homologous pairs
Organelles	Mitochondrial genome: 20.8 kb	complete single alleles

* Assembly metric benchmarks are adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024.

** 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/Aphodius_sticticus/dataset/GCA_963966075.1/busco.

Hi-C sample preparation

Tissue from the whole organism of the icAphStic2 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. Tissue (stored at -80°C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde, final concentration 2%. 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.

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

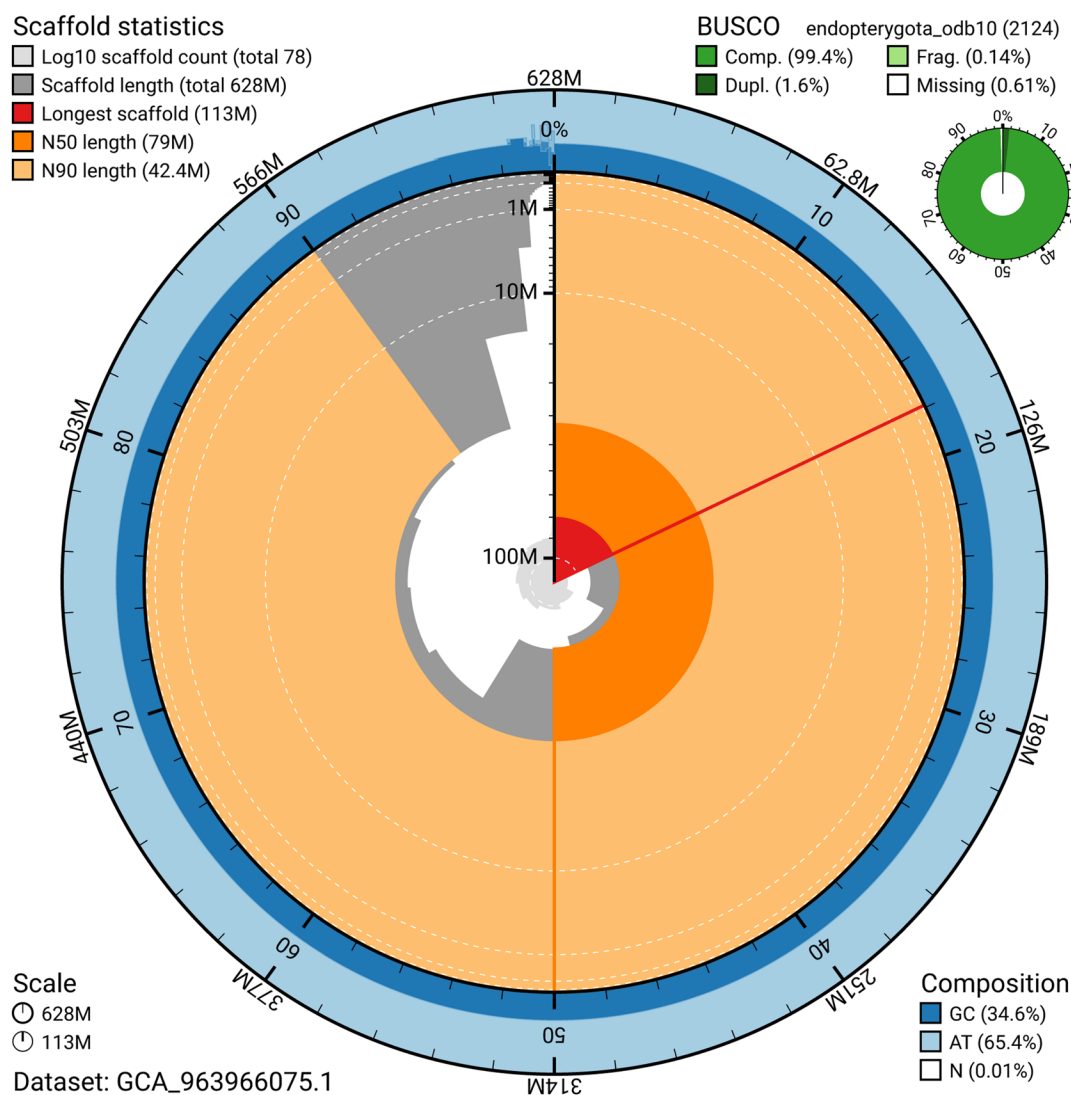


Figure 2. Genome assembly of *Aphodius sticticus*, icAphStic4.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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_963966075.1/dataset/GCA_963966075.1/snail.

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

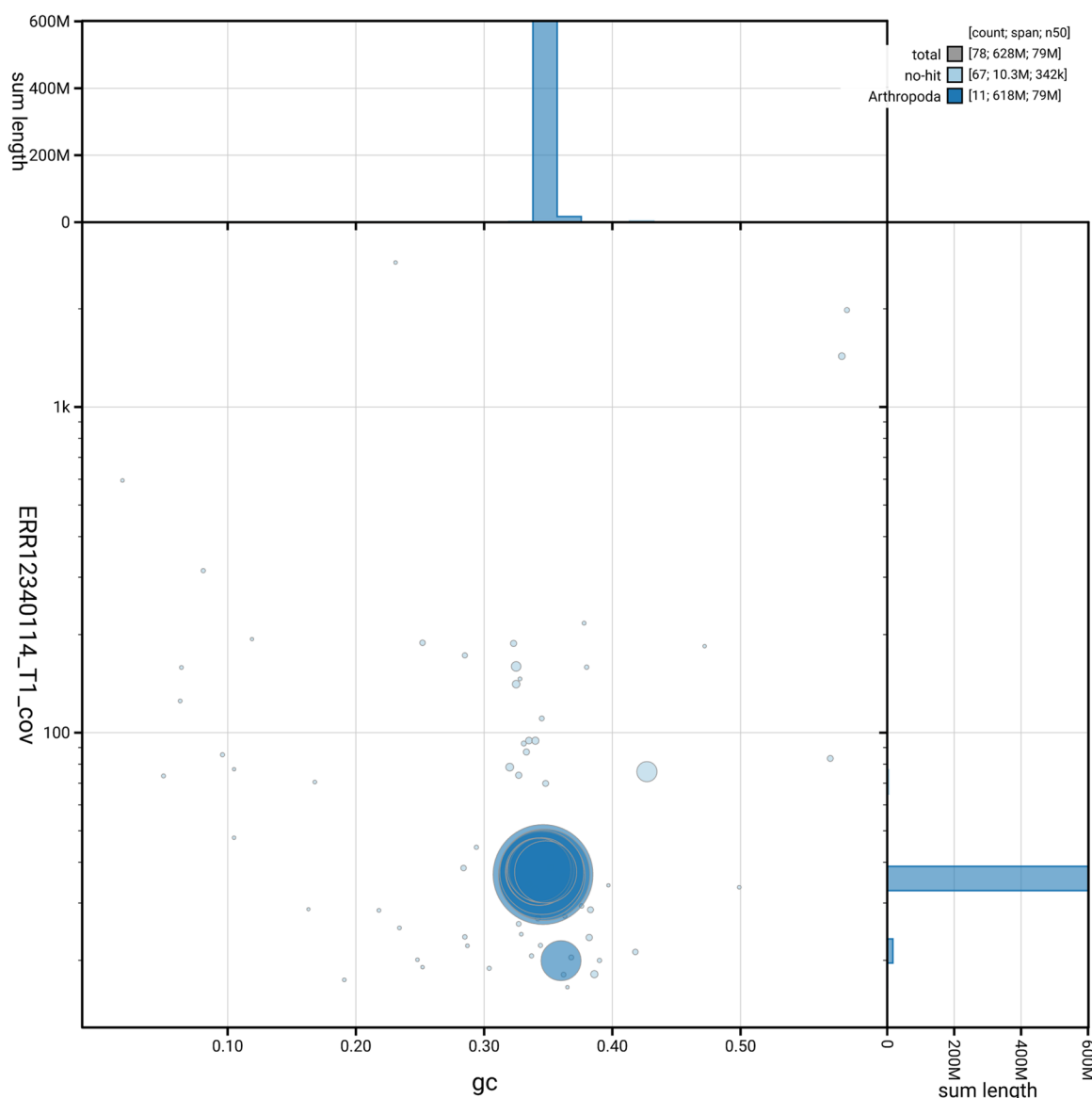


Figure 3. Genome assembly of *Aphodius sticticus*, icAphStic4.1: BlobToolKit GC-coverage plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms 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_963966075.1/dataset/GCA_963966075.1/blob.

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 6000 instrument.

Genome assembly, curation and evaluation

Assembly

The HiFi reads were 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 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

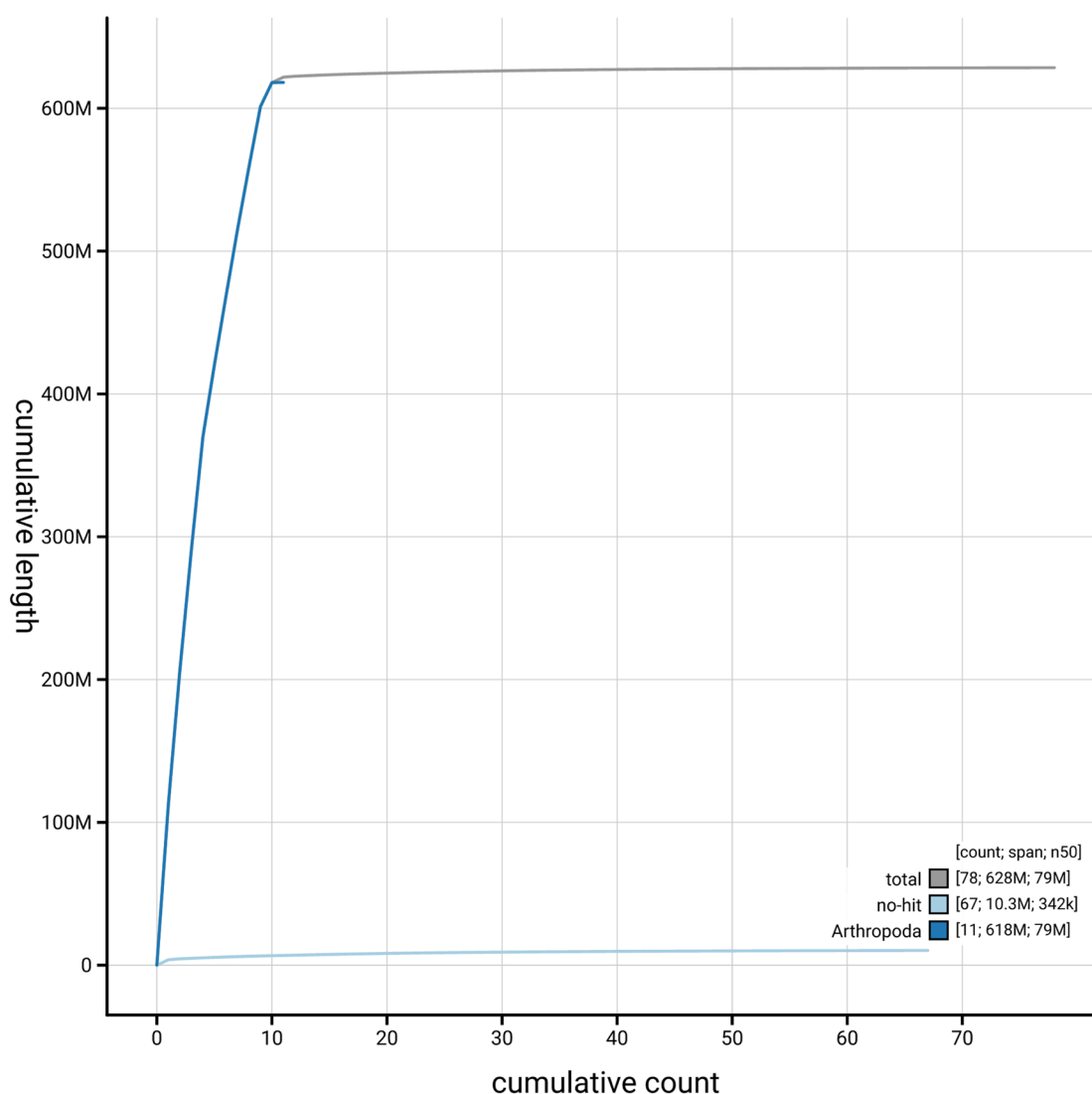


Figure 4. Genome assembly of *Aphodius sticticus* icAphStic4.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963966075.1/dataset/GCA_963966075.1/cumulative.

handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQUERY.FK (Rhie *et al.*, 2020).

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

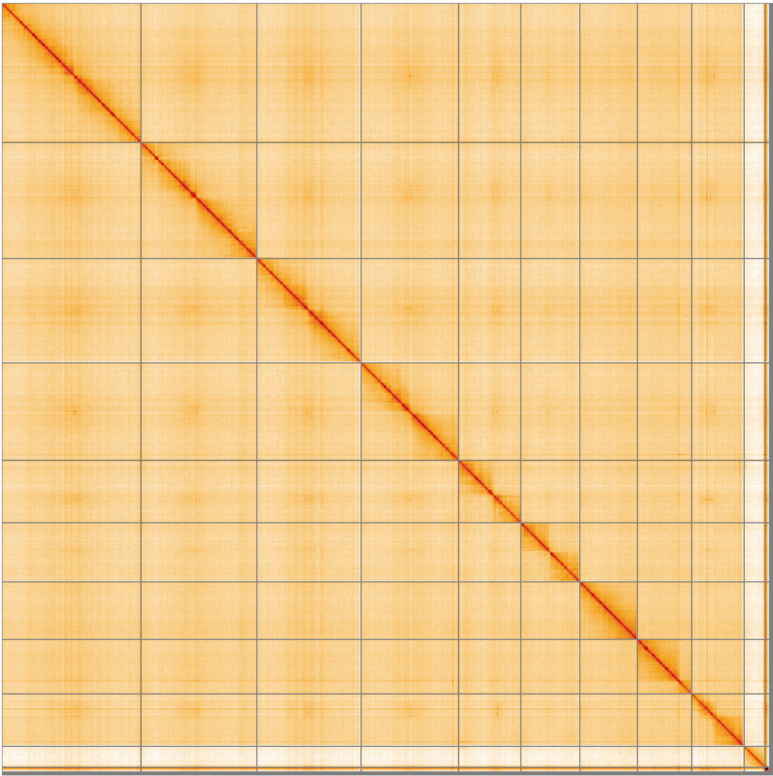


Figure 5. Genome assembly of *Aphodius sticticus* icAphStic4.1: Hi-C contact map of the icAphStic4.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=QScANppxQV2p1Yw8VLrZvA>.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Aphodius sticticus*, icAphStic4.

INSDC accession	Name	Length (Mb)	GC%
OZ014506.1	1	112.75	34.5
OZ014507.1	2	93.81	34.5
OZ014508.1	3	84.26	34.5
OZ014509.1	4	79.04	34.5
OZ014510.1	5	50.49	34.5
OZ014511.1	6	47.68	34.5
OZ014512.1	7	46.68	34.5
OZ014513.1	8	44.06	34.5
OZ014514.1	9	42.37	35.0
OZ014515.1	X	16.86	36.0
OZ014516.1	Y	3.8	42.5
OZ014517.1	MT	0.02	23.0

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. 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 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.

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

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/chhy1p123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
Merquy.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	0.2.5	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/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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 sticticus*. Accession number PRJEB70743; <https://identifiers.org/ena.embl/PRJEB70743>. The genome sequence is released openly for reuse.

The *Aphodius sticticus* 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>.

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 Megaruptor³ 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](#)
- British Scarabs: **Dwellers – Aphodiines**. 2025; [Accessed 28 January 2025]. [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](#)
- da Veiga Leprevost F, Grünig BA, Alves Afrits 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: **Checklist of Beetles of the British Isles.** 3rd Edition. Iver: Pembroley Books, 2018.
[Reference Source](#)
- 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](#)
- GBIF Secretariat: ***Volinus sticticus* (Panzer, 1798) Checklist dataset.** *GBIF Backbone Taxonomy*. 2023; [Accessed 29 January 2025].
[Reference Source](#)
- Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods*. 2018; **15**(7): 475–476.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics*. 2020; **36**(9): 2896–2898.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps.** 2022.
[Reference Source](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. 2021; **10**(1): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io*. 2023.
[Publisher Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One*. 2017; **12**(5): e0177459.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- NBN Atlas Partnership: ***Volinus sticticus* (Panzer, 1798).** *NBN Atlas*. 2024; [Accessed 25 January 2025].
[Reference Source](#)
- 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 SPRL.** *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.*: **Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project [version 1; peer review: 2 approved].** *Wellcome Open Res*. 2024; **9**: 339.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics*. 2023; **24**(1): 288.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.
[Publisher Full Text](#)
- 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 09 June 2025

<https://doi.org/10.21956/wellcomeopenres.26138.r124025>

© 2025 Cedden D. 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.



Doga Cedden 

University of Göttingen, Göttingen, France

The authors report a high-quality chromosome-level genome assembly from the small dung beetle, *Aphodius sticticus*. The background regarding the beetle is sufficient. The authors provide a thorough quality control of the assembly. The genome assembly and generated sequence data are publicly available.

-Please consider providing a scale in Fig. 1.

I have no further comments.

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: I work on coleopteran pests using RNA interference and RNA-seq

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 09 June 2025

<https://doi.org/10.21956/wellcomeopenres.26138.r124019>

© 2025 Perkin L. 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.



Lindsey C Perkin

United States Department of Agriculture ARS, College Station, Texas, USA

The manuscript presents a nice genome for the small dung beetle, *Volinus stricticus*. The authors follow the community standard for collecting, sequencing and assembling reference quality genomes.

The protocols and methods are well defined and easy to follow. The data is supported and summarized by many tables and figures.

The datasets are referenced and in an accessible format.

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: agriculturally significant pests of row crops

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 30 May 2025

<https://doi.org/10.21956/wellcomeopenres.26138.r124018>

© 2025 Douglas H. 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.



Hume Douglas 

Agriculture and Agri-Food Canada, Ottawa, Canada

This article efficiently presents the information that genome users will need. This article is part of a series, and all are well written.

The information about the beetle is sufficient. The species is a good choice for sequencing.

The methods are well presented.

The results are well presented.

The sequence data, mapping and chromosome information will all be appreciated by researchers.

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: Beetle 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.
