DATA NOTE

The genome sequence of a ground beetle, *Dromius quadrimaculatus* (Linnaeus, 1758) (Coleoptera: Carabidae)

[version 1; peer review: 2 approved, 1 approved with reservations]

Liam M. Crowley[1], Roger Booth[2], Michael F. Geiser[2],
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]University of Oxford, Oxford, England, UK
[2]Natural History Museum, London, England, UK

v1 **First published:** 06 Oct 2025, **10**:555
https://doi.org/10.12688/wellcomeopenres.24898.1
Latest published: 06 Oct 2025, **10**:555
https://doi.org/10.12688/wellcomeopenres.24898.1

Abstract

We present a genome assembly from an individual female *Dromius quadrimaculatus* (ground beetle; Arthropoda; Insecta; Coleoptera; Carabidae). The genome sequence has a total length of 778.78 megabases. Most of the assembly (95.73%) is scaffolded into 14 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 16.57 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Dromius quadrimaculatus; ground 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 06 Oct 2025	✓ view	✓ view	? view

1. **Min Li**, Taiyuan Normal University, Jinzhong, China
2. **Liang Lu**, Hebei Normal University, Shijiazhuang, China
3. **Nadège Guiglielmoni**, University of Cologne Center for Molecular Medicine Cologne, Cologne, Germany
Universite Libre de Bruxelles Departement de Biologie des Organismes (Ringgold ID: 168978), Brussels, Belgium

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: Crowley LM: Investigation, Resources; **Booth R**: Investigation, Resources; **Geiser MF**: 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.

How to cite this article: Crowley LM, Booth R, Geiser MF *et al.* **The genome sequence of a ground beetle, *Dromius quadrimaculatus* (Linnaeus, 1758) (Coleoptera: Carabidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10**:555 https://doi.org/10.12688/wellcomeopenres.24898.1

First published: 06 Oct 2025, **10**:555 https://doi.org/10.12688/wellcomeopenres.24898.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Adephaga; Caraboidea; Carabidae; Lebiinae; Lebiini; *Dromius*; *Dromius quadrimaculatus* (Linnaeus, 1758) (NCBI:txid346772)

Background

Dromius quadrimaculatus (Linnaeus, 1758) is a small arboreal carabid (5–6.4 mm) with four pale elytral spots. It is typically >5 mm (larger than similar spotted arboreal carabids), with the apical spots reaching the hind elytral margin and lacking the basal pore-puncture present in *Calodromius* and *Philorhizus* (Telfer & Walters, 2014). It is found on trunks and under bark, mostly on deciduous trees (sycamore, alder, willow, lime, apple) but also recorded from spruce (Habitas, 2025).

It can be found throughout the year under tree bark and among lichen and moss on tree trunks (Telfer & Walters, 2014). It is native to the Palaearctic and the Near East and is widespread across much of Europe, including Great Britain and Ireland; within Britain it is frequent in England and Wales and more local in Scotland and Ireland (NBN Atlas Partnership, 2025; Peters, 2025). Adults and larvae are predatory on small arthropods and can be encountered year-round; the species is also known as the Great Four-spot Treerunner (Peters, 2025).

In a canopy-crane study of a temperate floodplain forest near Leipzig, *D. quadrimaculatus* dominated the canopy catch (~70% of carabids in traps) and was absent from ground pitfall traps, indicating an arboricolous life history (Arndt & Hielscher, 2007). Activity peaked in late July to early August, with slightly higher captures on ash (*Fraxinus excelsior*). The same study explicitly classed *D. quadrimaculatus* among the arboricolous carabids at the site (Arndt & Hielscher, 2007).

We present the first high-quality genome for the genus *Dromius*, one of 37 genomes available for the family Carabidae as of August 2025 (data obtained via NCBI datasets, O'Leary *et al.*, 2024). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Blaxter *et al.*, 2022).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was a female adult *Dromius quadrimaculatus* (specimen ID Ox000312, ToLID icDroQuad1; Figure 1), collected from Wytham Great Wood, United Kingdom (latitude 51.771, longitude -1.338) on 2019-10-30. The specimen was collected and identified by Liam Crowley (University of Oxford). A second specimen was used for Hi-C sequencing (specimen ID NHMUK015058720, ToLID icDroQuad2). It was collected from Buxton Heath, England, United Kingdom (latitude 52.75, longitude 1.22) on 2022-07-06. The specimen was collected and identified by Roger Booth (Natural History Museum). A third specimen was used for RNA sequencing (specimen ID NHMUK015073950, ToLID icDroQuad4). It was collected from Pullingshill Wood, Marlow, England, United Kingdom (latitude 51.57, longitude –0.82) on 2022-07-23. The specimen was collected by David Lees and identified by Michael Geiser (Natural History Museum). For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

Figure 1. Photograph of the *Dromius quadrimaculatus* (icDroQuad1) specimen used for genome sequencing.

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) (see the protocol). 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 are available on protocols.io.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The icDroQuad1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). 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 icDroQuad4 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. 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.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was 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, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the icDroQuad2 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

RNA library preparation and sequencing

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

Genome assembly

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

The HiFi reads were assembled using Hifiasm (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 in YaHS (Zhou *et al.*, 2023) with 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. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 18 breaks, 38 joins, and removal of three haplotypic duplications. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation.

PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all 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), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Dromius quadrimaculatus* specimen generated 28.08 Gb (gigabases) from 1.89 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 790.46 Mb, with a heterozygosity of 2.11% and repeat content of 44.56% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 33× coverage. Hi-C sequencing produced 94.21 Gb from 623.89 million reads, which were 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.

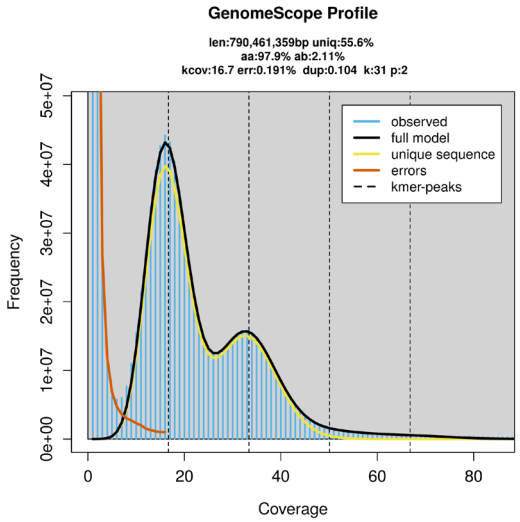

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB65203.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	icDroQuad1	icDroQuad2	icDroQuad4
Specimen ID	Ox000312	NHMUK015058720	NHMUK015073950
BioSample (source individual)	SAMEA7520206	SAMEA112964344	SAMEA115574951
BioSample (tissue)	SAMEA7520298	SAMEA112975509	SAMEA115575072
Tissue	whole organism	whole organism	whole organism
Instrument	Sequel II	Illumina NovaSeq 6000	Illumina NovaSeq X
Run accessions	ERR12040336	ERR11872568	ERR13999075
Read count total	1.89 million	623.89 million	83.79 million
Base count total	28.08 Gb	94.21 Gb	12.65 Gb

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 778.78 Mb in 425 scaffolds, with 192 gaps, and a scaffold N50 of 60.83 Mb (Table 2).

Most of the assembly sequence (95.73%) was assigned to 14 chromosomal-level scaffolds, representing 12 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome X was identified by alignment to GCA_943142095.1.

Table 2. Genome assembly statistics.

Assembly name	icDroQuad1.1
Assembly accession	GCA_963989225.1
Alternate haplotype accession	GCA_963989235.1
Assembly level	chromosome
Span (Mb)	778.78
Number of chromosomes	14
Number of contigs	617
Contig N50	25.78 Mb
Number of scaffolds	425
Scaffold N50	60.83 Mb
Sex chromosomes	X
Organelles	Mitochondrion: 16.57 kb

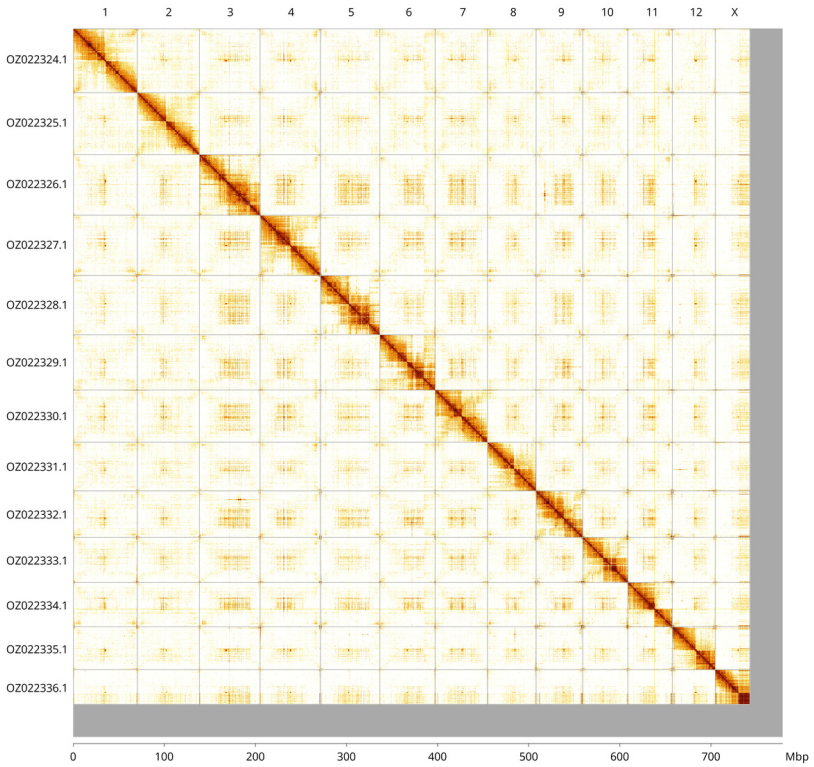

Figure 3. Hi-C contact map of the *Dromius quadrimaculatus* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Dromius quadrimaculatus* icDroQuad1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ022324.1	1	70.83	32.50
OZ022325.1	2	68.38	32.50
OZ022326.1	3	66.70	33.50
OZ022327.1	4	66.65	33
OZ022328.1	5	65.50	33.50
OZ022329.1	6	60.83	33.50
OZ022330.1	7	57.66	33
OZ022331.1	8	53.56	33
OZ022332.1	9	51.35	33
OZ022333.1	10	49.90	33
OZ022334.1	11	48.81	33.50
OZ022335.1	12	47.34	33.50
OZ022337.1	13	0.32	32
OZ022336.1	X	37.72	33

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.

The combined primary and alternate assemblies achieve an estimated QV of 58.3. The *k*-mer completeness is 68.16% for the primary assembly, 65.44% for the alternate haplotype, and 97.98% for the combined assemblies (Figure 4).

BUSCO v.5.5.0 analysis using the endopterygota_odb10 reference set (n = 2 124) identified 99.2% of the expected gene set (single = 98.2%, duplicated = 1.1%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **7.C.Q59**, meeting the recommended reference standard.

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

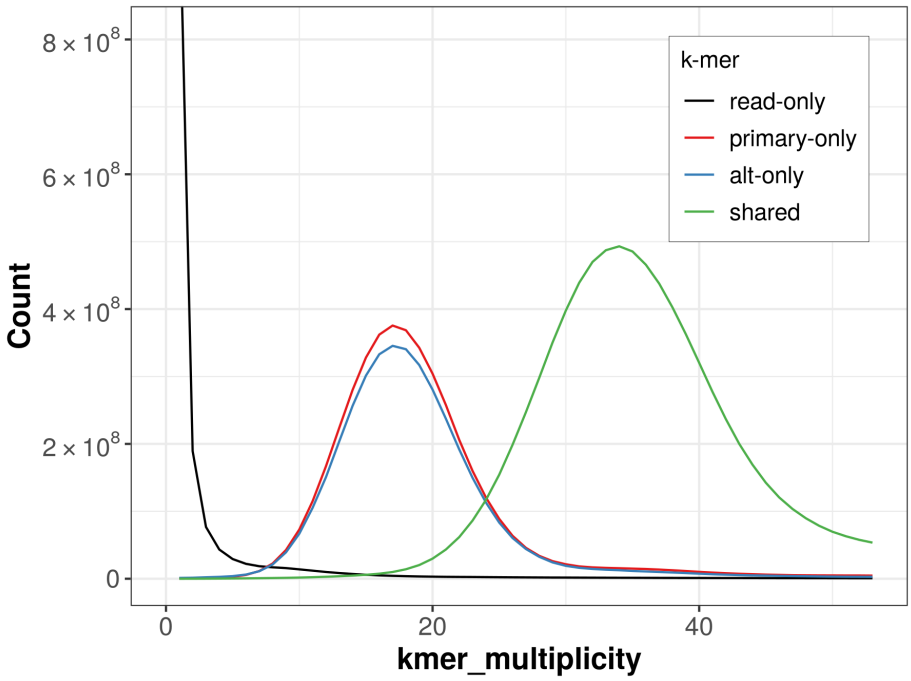

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

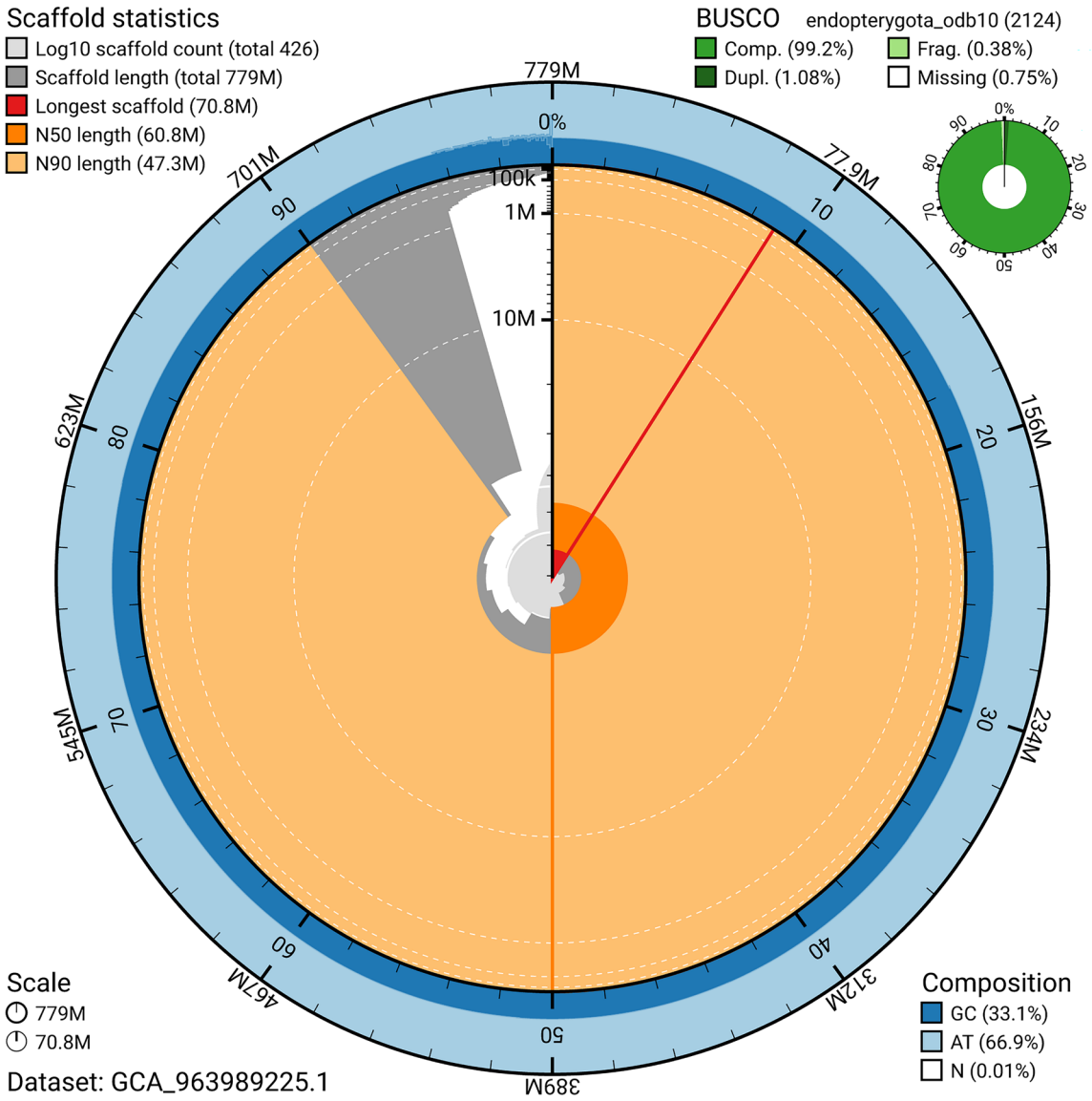

Figure 5. Assembly metrics for icDroQuad1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 can be accessed on the BlobToolKit viewer.

to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. 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

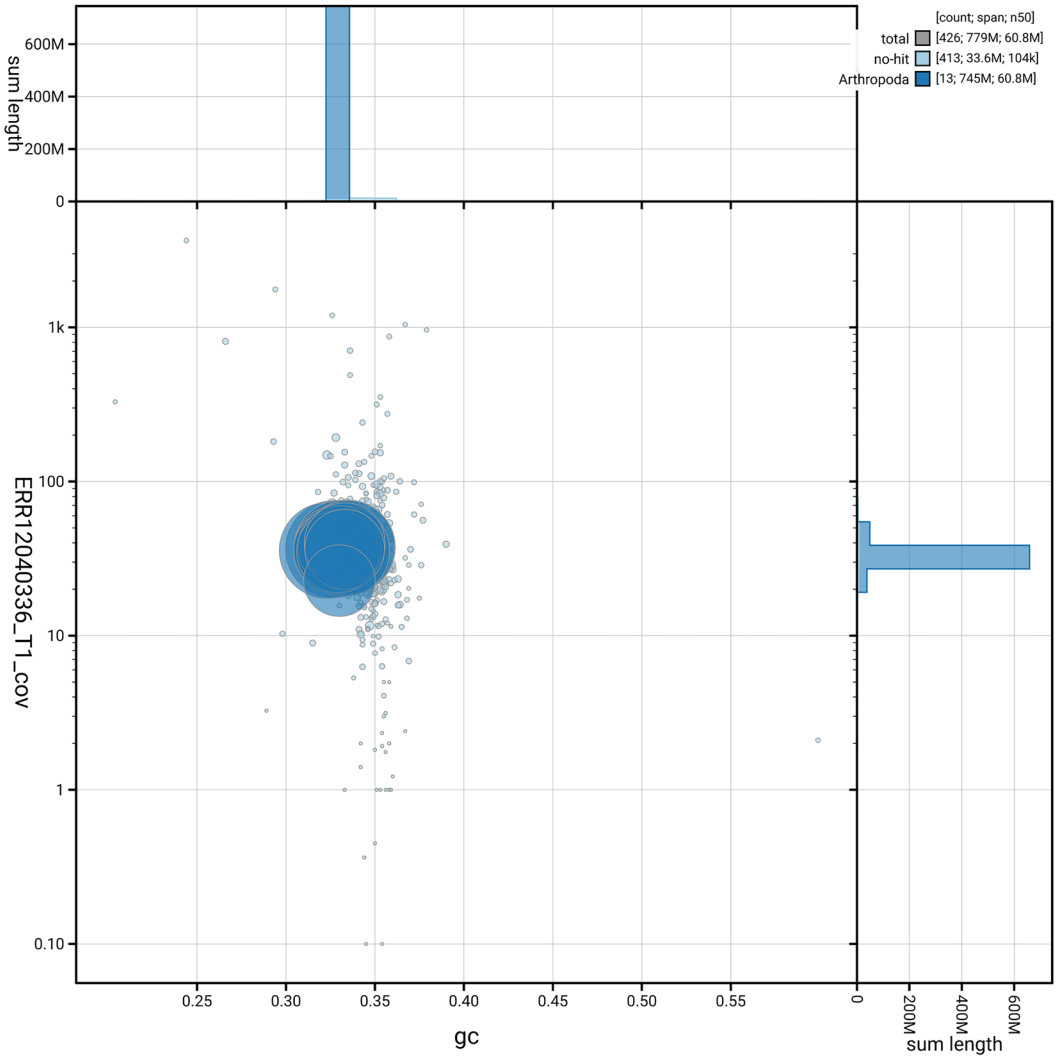

Figure 6. BlobToolKit GC-coverage plot for icDroQuad1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Dromius quadrimaculatus* assembly.

Measure	Value	Benchmark
EBP summary (primary)	7.C.Q59	6.C.Q40
Contig N50 length	25.78 Mb	≥ 1 Mb
Scaffold N50 length	60.83 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.2; alternate: 57.6; combined: 58.3	≥ 40
k-mer completeness	Primary: 68.16%; alternate: 65.44%; combined: 97.98%	≥ 95%
BUSCO	C:99.2% [S:98.2%; D:1.1%]; F:0.4%; M:0.4%; n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	95.73%	≥ 90%

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: Dromius quadrimaculatus. Accession number PRJEB65203. The genome sequence is released openly for reuse. The *Dromius quadrimaculatus* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and 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.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

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

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC

Software	Version	Source
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/ blobtoolkit	0.4.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/ curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

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

Arndt E, Hielscher S: **Ground beetles (Coleoptera: Carabidae) in the forest canopy: Species composition, seasonality, and year-to-year fluctuation.** In: M. Unterseher, W. Morawetz, S. Klotz, *et al.* (eds), *The Canopy of a Temperate Floodplain Forest: Results from Five Years of Research at the Leipzig Canopy Crane*. Leipzig, Germany: University of Leipzig, 2007; 106–110.
Reference Source

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

Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **Sequence locally, think globally: the Darwin Tree of Life Project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, 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

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

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98.
PubMed Abstract | Publisher Full Text | Free Full Text

Habitas: **Species account: *Dromius quadrimaculatus***. 2025.
Reference Source

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

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

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

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

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

Li H: **Minimap2: pairwise alignment for nucleotide sequences.**

Bioinformatics. 2018; **34**(18): 3094–3100.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Manni M, Berkeley MR, 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.
Reference Source

NBN Atlas Partnership: ***Dromius quadrimaculatus*: Species profile**. 2025.
Reference Source

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**

Peters D: ***Dromius quadrimaculatus***. 2025.
Reference Source

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

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

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

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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Telfer M, Walters J: **Guides to British Beetles 15: *Demetrias*, *Paradromius*, *Dromius*, *Calodromius* and *Philorhizus* (Coleoptera: Carabidae).** 2014.
Reference Source

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

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

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

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 1

Reviewer Report 05 January 2026

https://doi.org/10.21956/wellcomeopenres.27425.r142545

?

Nadège Guiglielmoni
[1] University of Cologne Center for Molecular Medicine Cologne, Cologne, Germany
[2] Universite Libre de Bruxelles Departement de Biologie des Organismes (Ringgold ID: 168978), Brussels, Brussels, Belgium

This study presents a chromosome-level genome assembly of a female individual of the arboreal carabid species *Dromius quadrimaculatus.* PacBio HiFi and Hi-C reads, from two different individuals, were combined to produce chromosome candidates using hifiasm and YaHS, and MitoHiFi for the mitochondrial genome. Genome evaluation supports correctness and completeness of the assembly.

I have a few small modifications to suggest:

1) I would spell out that the species has XY chromosomes.

2) I would detail for all the sequencing datasets the sex of the specimen, especially for those who may want to reuse them.

3) I am confused why this plot from Merqury was incorporated with pri/alt/shared. If you used the primary assembly, I would expect to see the spectra-cn plot of the primary. And is the QV indeed for the primary of pri+alt?

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: Genome assembly, animal genomics, animal parthenogenesis

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, however I have significant reservations, as outlined above.

Reviewer Report 02 January 2026

https://doi.org/10.21956/wellcomeopenres.27425.r142548

Liang Lu
Hebei Normal University, Shijiazhuang, Hebei, China

I have read the main text of the manuscript, which reports a very commonly found (in the UK) ground beetle, Dromius quadrimaculatus (Linnaeus, 1758). It is a formulaic paper that reports an NGS sequence with detailed description of the background of the sample information in taxonomy, biology, collection, etc. The authors kindly presented in this manuscript clear images that make their identification trustable. The workflows of sample preparation, DNA extraction, sequencing, and assembly are fine and aligned with previous work/publications by Wellcome Sanger Institute. But why this time they do not include annotation? So I suggest the author should provide details of prediction and annotation, methods, results (with the assembly), and a report on the results.

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: Zoology (Coleoptera), Phylogenomics

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

Reviewer Report 12 November 2025

https://doi.org/10.21956/wellcomeopenres.27425.r137098

Min Li
Taiyuan Normal University, Jinzhong, China

This is a high-quality genomic data report with standardized experimental design, mature technical implementation, and reliable data quality, fully meeting the publication standards for open science and genomics research.
I suggest only a few minor comments:

1. Were the three sequenced specimens identified by both morphological method and DNA barcoding process? Before DNA extraction, were the specimens undergone aseptic treatment?
2. Clarify the sex of the specimen (specimen ID NHMUK015058720, ToLID icDroQuad2) used for Hi-C sequencing to ensure consistency between assembly results and sex-related chromosome annotations.
3. Supplement key steps of the ASCC decontamination pipeline (e.g., contamination sequence identification criteria, filtering thresholds) to illustrate how the purity of the genome assembly is ensured.
4. If the annotations of the genome are also included, it would be more comprehensive.

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: molecular evolutionary biology; molecular ecology; DNA metabarcoding

technology

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.