DATA NOTE

The genome sequence of a rove beetle, *Tachyporus hypnorum* (Fabricius, 1775) (Coleoptera: Staphylinidae)

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

Roger Booth[1], Maxwell V. L. Barclay[1],
Natural History Museum Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Natural History Museum, London, England, UK

v1 **First published:** 19 Mar 2026, **11**:182
https://doi.org/10.12688/wellcomeopenres.26176.1
Latest published: 19 Mar 2026, **11**:182
https://doi.org/10.12688/wellcomeopenres.26176.1

Abstract

We present a genome assembly from an individual female *Tachyporus hypnorum* (rove beetle; Arthropoda; Insecta; Coleoptera; Staphylinidae). The genome sequence has a total length of 531.37 megabases. Most of the assembly (63.74%) is scaffolded into 13 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 19.75 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

Tachyporus hypnorum, rove 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 19 Mar 2026	? view	? view	✓ view

1. **Henrik Lantz**, SciLifeLab, Uppsala University, Uppsala, Sweden
2. **Wai Lok So**, The Chinese University of Hong Kong, Hong Kong, Hong Kong
3. **Nikoletta Andrea Nagy**, University of Debrecen, Debrecen, Hungary

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: Booth R: Investigation, Resources, Writing – Original Draft Preparation; **Barclay MVL**: 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: Booth R, Barclay MVL, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of a rove beetle, *Tachyporus hypnorum* (Fabricius, 1775) (Coleoptera: Staphylinidae) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2026, **11**:182 https://doi.org/10.12688/wellcomeopenres.26176.1

First published: 19 Mar 2026, **11**:182 https://doi.org/10.12688/wellcomeopenres.26176.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Staphyliniformia; Staphylinoidea; Staphylinidae; Tachyporinae group; Tachyporinae; *Tachyporus*; *Tachyporus hypnorum* (Fabricius, 1775) (NCBI:txid346861).

Background

Tachyporus hypnorum (Fabricius, 1775) is a rove beetle in the family Staphylinidae. It is narrow and elongate in shape, broadest across the base of the pronotum and short elytra and with the exposed abdomen tapering apically. It is 3–4 mm in length. It has a black head, a pronotum which is almost black on the disc and with the sides yellowish, brownish-orange elytra with blackish marks basally and laterally, and a black abdomen with narrow pale-orange apical margins to the segments, and pale yellowish appendages. Its distinctively coloured pronotum, general shape and size mean that, with experience, it can be recognised in the field (Duff, 2024).

It is a general predator of smaller insects, including aphids, Collembola and Diptera larvae (Good & Giller, 1991) and can be found on the ground, in moss, decaying vegetation and in tussocks, etc. Because it also climbs up low vegetation, it can be collected with a sweep net, unlike some other members of the genus *Tachyporus* which do not climb vegetation.

Tachyporus hypnorum occurs widely throughout the British Isles, up to and including the Shetland Islands and can be numerous at times. It is widespread throughout Europe except for the far north.

We present a chromosome-level genome sequence for *Tachyporus hypnorum*, generated using the Tree of Life pipeline from a specimen collected from Winterton Dunes, England, UK (Figure 1). This assembly is the first high-quality genome for the genus *Tachyporus* and one of 56 genomes available for the family Staphylinidae as of March 2026 (data obtained via NCBI datasets, O'Leary *et al.*, 2024).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Tachyporus hypnorum* (specimen ID NHMUK015059229, ToLID icTacHypn4; Figure 1), collected from Winterton Dunes, England, UK (latitude 52.73, longitude 1.69) on 2022-07-05. The specimen was collected and identified by Roger Booth. A second specimen was used

Figure 1. Photograph of the *Tachyporus hypnorum* (icTacHypn4) specimen used for genome sequencing.

for Hi-C sequencing (specimen ID NHMUK014400239, ToLID icTacHypn3). It was collected from Parsons Green, England, UK (latitude 51.4751, longitude −0.1879) on 2021-05-12. The specimen was collected and identified by Maxwell Barclay.

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 icTacHypn4 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by power-mashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. We used centrifuge-mediated fragmentation to produce DNA fragments in the 8–10 kb range, following the Covaris g-TUBE protocol for ultra-low input (ULI). 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. For this sample, the final post-shearing DNA had a Qubit concentration of 2.1 ng/μL and a yield of 273.00 ng.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Prior to library preparation, the DNA was fragmented to ~10 kb. Ultra-low-input (ULI) libraries were prepared using the PacBio SMRTbell® Express Template Prep Kit 2.0 and gDNA Sample Amplification Kit. Samples were normalised to 20 ng DNA. Single-strand overhang removal, DNA damage repair, and end-repair/A-tailing were performed according to the manufacturer's instructions, followed by adapter ligation. A 0.85× pre-PCR clean-up was carried out with Promega ProNex beads.

The DNA was evenly divided into two aliquots for dual PCR (reactions A and B), both following the manufacturer's protocol. A 0.85× post-PCR clean-up was performed with ProNex beads. DNA concentration was measured using a Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with the Qubit HS Assay Kit, and fragment size was assessed on an Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit. PCR reactions A and B were then pooled, ensuring a total mass of ≥500 ng in 47.4 μl.

The pooled sample underwent another round of DNA damage repair, end-repair/A-tailing, and hairpin adapter ligation. A 1× clean-up was performed with ProNex beads, followed by DNA quantification using the Qubit and fragment size analysis using the Agilent Femto Pulse. Size selection was performed on the Sage Sciences PippinHT system, with target fragment size determined by Femto Pulse analysis (typically 4–9 kb). Size-selected libraries were cleaned with 1.0× ProNex beads and normalised to 2 nM before sequencing.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the icTacHypn3 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 to create 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.

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. 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).

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 99 breaks and 387 joins. This reduced the scaffold count by 10.9%, increased the scaffold N50 by 345.4%, and reduced the total assembly length by 4.3%. The curation process is described 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 *Tachyporus hypnorum* specimen generated 43.17 Gb (gigabases) from 4.46 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 488.51 Mb, with a heterozygosity of 1.88% and repeat content of 57.12% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 80× coverage. Hi-C sequencing produced 108.02 Gb from 715.36 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

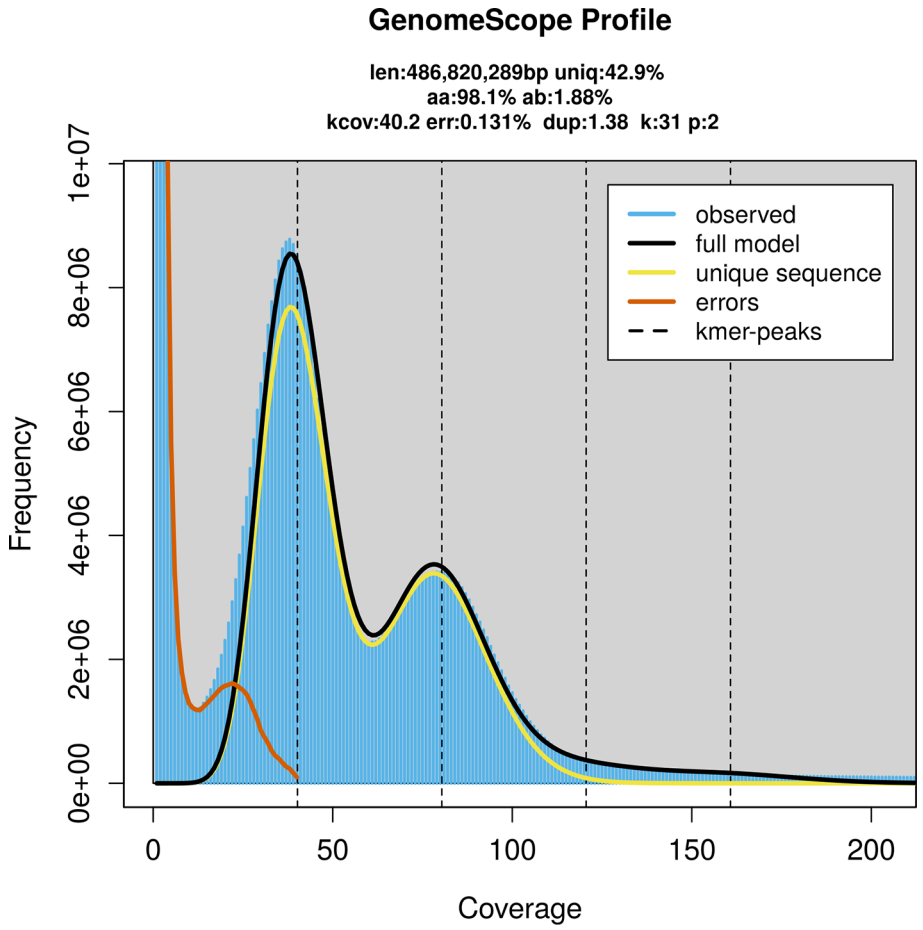

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB76757.

Platform	PacBio HiFi	Hi-C
ToLID	icTacHypn4	icTacHypn3
Specimen ID	NHMUK015059229	NHMUK014400239
BioSample (source individual)	SAMEA112963073	SAMEA9359432
BioSample (tissue)	SAMEA112963166	SAMEA9359499
Tissue	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq 6000
Run accessions	ERR13304145	ERR13317823
Read count total	4.46 million	715.36 million
Base count total	43.17 Gb	108.02 Gb

Table 2. **Genome assembly statistics.**

Assembly name	icTacHypn4.1
Assembly accession	GCA_965783885.1
Alternate haplotype accession	GCA_965783505.1
Assembly level	chromosome
Span (Mb)	531.37
Number of chromosomes	13
Number of contigs	1 393
Contig N50	1.06 Mb
Number of scaffolds	702
Scaffold N50	19.01 Mb
Sex chromosomes	X
Organelles	Mitochondrion: 19.75 kb

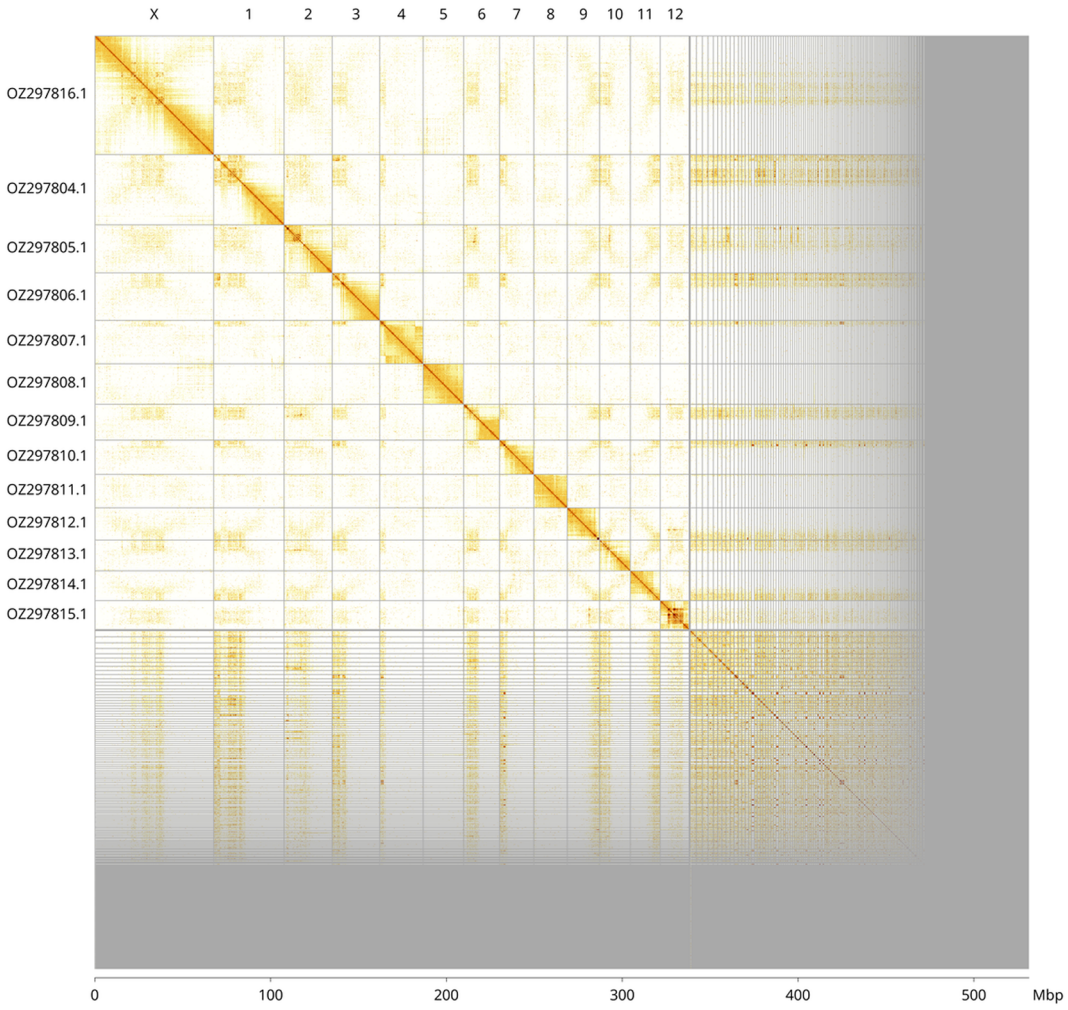

Figure 3. **Hi-C contact map of the *Tachyporus hypnorum* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

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 531.37 Mb in 702 scaffolds, with 691 gaps, and a scaffold N50 of 19.01 Mb (Table 2).

Most of the assembly sequence (63.74%) was assigned to 13 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 and Table 3). Chromosome X was assigned based on BUSCO ancestral hits.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Tachyporus hypnorum* icTacHypn4.

INSDC accession	Molecule	Length (Mb)	GC%
OZ297804.1	1	40.09	34
OZ297805.1	2	27.35	35
OZ297806.1	3	27.06	34.50
OZ297807.1	4	24.62	34.50
OZ297808.1	5	23.06	33.50
OZ297809.1	6	20.48	34.50
OZ297810.1	7	19.49	35
OZ297811.1	8	19.01	33
OZ297812.1	9	18.30	34
OZ297813.1	10	17.53	34
OZ297814.1	11	17.04	33.50
OZ297815.1	12	17.01	34
OZ297816.1	X	67.67	33.50

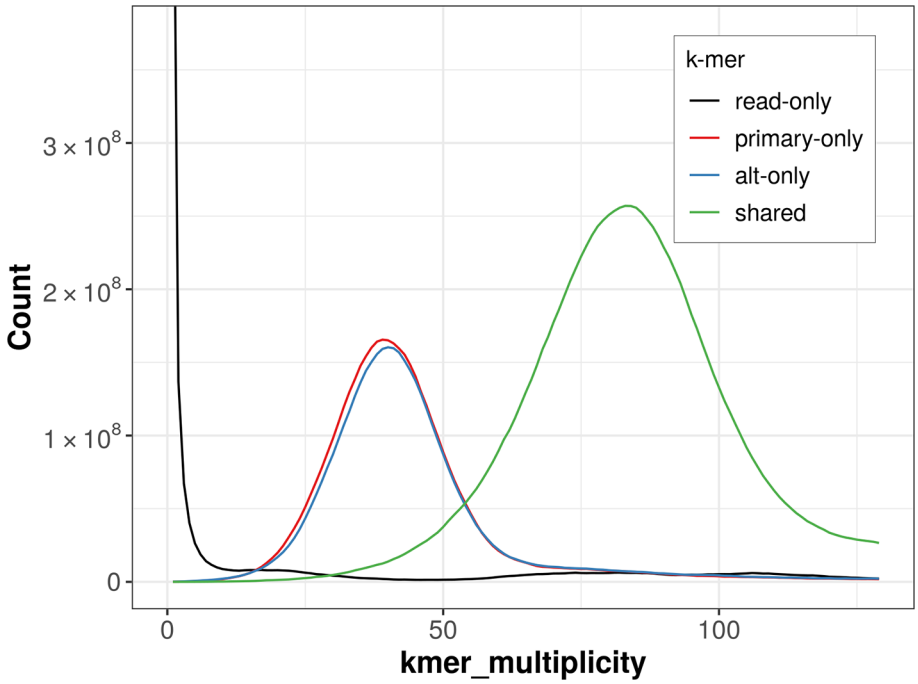

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

The mitochondrial genome was also assembled (length 19.75 kb, OZ297817.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 58.9. The *k*-mer completeness is 69.25% for the primary assembly, 67.61% for the alternate haplotype, and 96.23% for the combined assemblies (Figure 4).

BUSCO v.6.0.0 analysis using the endopterygota_odb10 reference set (n = 2 124) identified 98.9% of the expected gene set (single = 97.1%, duplicated = 1.7%). 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) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **6.7.Q59**.

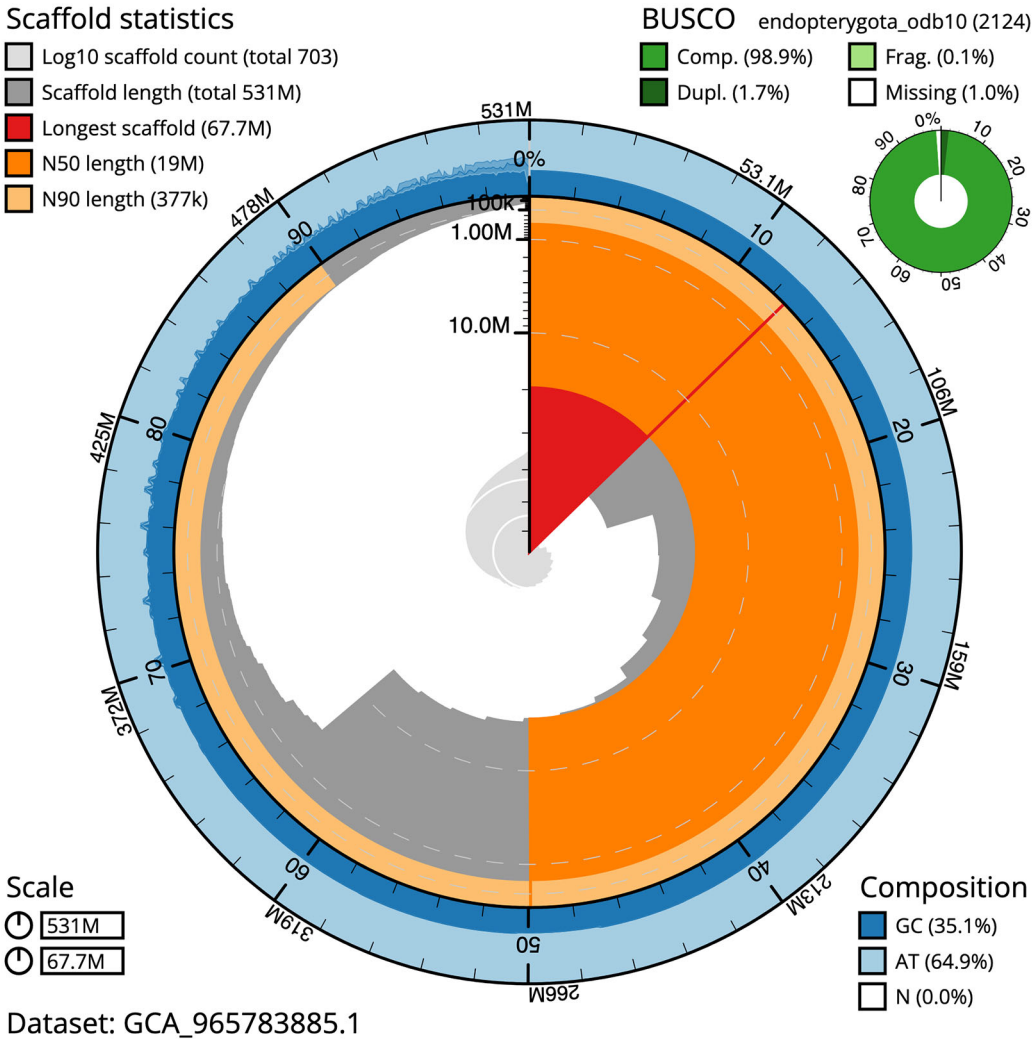

Figure 5. Assembly metrics for icTacHypn4.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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.

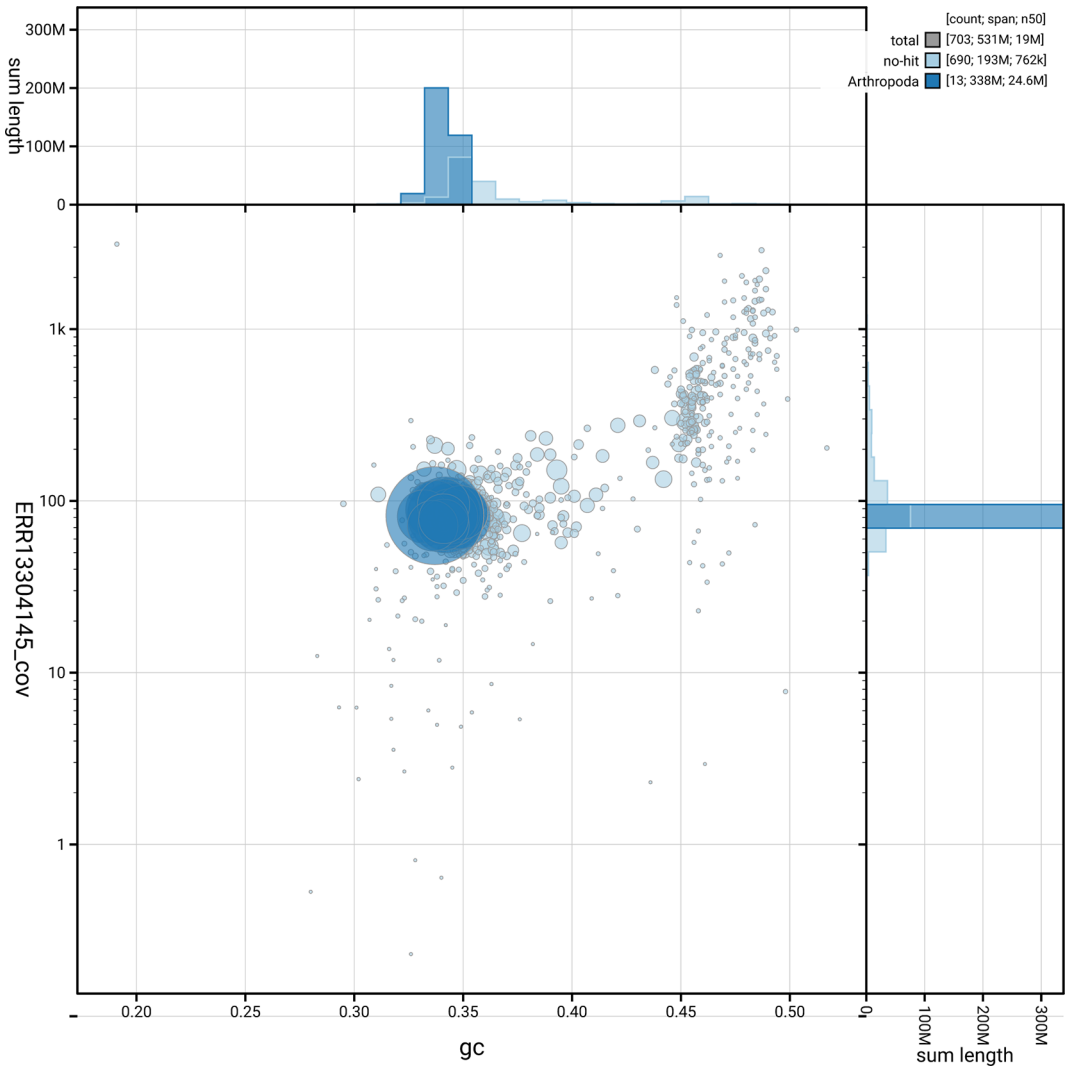

Figure 6. BlobToolKit blob plot for icTacHypn4.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total 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 *Tachyporus hypnorum* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.7.Q59	6.C.Q40
Contig N50 length	1.06 Mb	≥ 1 Mb
Scaffold N50 length	19.01 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.9; alternate: 58.5; combined: 58.9	≥ 40
k-mer completeness	Primary: 69.25%; alternate: 67.61%; combined: 96.23%	≥ 95%
BUSCO	C:98.9% [S:97.1%, D:1.7%], F:0.1%, M:1.0%, n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	63.74%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues.

Author information

Contributors are listed at the following links:

- 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

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. 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: Tachyporus hypnorum. Accession number PRJEB76757. The genome sequence is released openly for reuse. The *Tachyporus hypnorum* 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 Tables 1 and 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.

Table 5. Software versions and sources.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	6.0.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2

Table 5. *Continued*

Software	Version	Source
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
Hifiasm	0.19.8-r603	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	2.28-r1209	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
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.9.0	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

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 M-J, 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

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 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): 170–75.
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.** *Wellcome Open Res.* 2023; **8**: 123.
Publisher Full Text

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

Duff AG: *Beetles of Britain and Ireland. Vol. 2: Staphylinidae.* West Runton: A. G. Duff Publishing; 2024.

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

Good JA, Giller PS: **The diet of predatory Staphylinid beetles – a review of records.** *Entomol. Mon. Mag.* 1991; **127**: 77–89.

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).
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.
Publisher Full Text

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

Li H: **Minimap2: Pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
Publisher 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.
Publisher Full Text

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239).
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.
Publisher 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.
Publisher 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.
Publisher 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.
Publisher 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).
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.* 2020; **2020**: baaa062.
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.** *Wellcome Open Res.* 2024; **9**: 339.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: A Python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
Publisher Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–24.
Publisher Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1).
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ? ✓

Version 1

Reviewer Report 14 April 2026

https://doi.org/10.21956/wellcomeopenres.28829.r151926

✓ **Nikoletta Andrea Nagy**
University of Debrecen, Debrecen, Hungary

The manuscript is well written, and the genome presented can contribute to future research on rove beetles. I have only a few suggestions for the authors.

Background
- when describing the appearance species, I would recommend changing "It is" to "The body of the species"
- suggested change:
  from "Its distinctively coloured pronotum, general shape and size mean that, with experience, it can be recognised in the field"
  to "Based on its distinctively coloured pronotum, general shape and size, the species can be recognised in the field"
Genome assembly
- please include the read filtering method
Sequence data
- the haploid genome size in the main text does not match with the one on Figure 2
Assembly statistics
- please include whether the assembled mitochondrial genome was circularized or linear

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, genome annotation, transcriptome analysis

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 14 April 2026

https://doi.org/10.21956/wellcomeopenres.28829.r151927

?

Wai Lok So
The Chinese University of Hong Kong, Hong Kong, Hong Kong

The manuscript presents a genome assembly of T*achyporus hypnorum* as a part of the Darwin Tree of Life initiative. The assembly methods are technically sound, yielding a genome with 13 pseudo-chromosomes, including the sex chromosome and a complete mitochondrial genome. Below are some comments for the improvement of the current manuscript.

Comments:

1. The authors failed to mention the importance of sequencing this beetle species. It is encouraged the authors to state the rationale for creating such genome resources in the Background.

2. The authors spent a paragraph in Background in illustrating the morphology of the beetle, which is detailed and sufficient. However, the orientation and appearance of the beetle photograph is poor in Figure 1. It gives no clue on the identity of the described beetle. Could the authors also provide a proper and nicer photo for this species in Figure 1?

3. The authors mentioned the sample was dissected and stored in ethanol in Methods. Could the authors provide more information/details on the dissection? What had been dissected and what kind of tissues had been taken out? How much was the tissue? Was the gut remove in the sample?

4. The authors mentioned that they used the centrifuge-mediated fragmentation to produce the DNA fragments in 8-10 kb range. Please kindly provide the profile of the centrifuge in achieving the desired fragment sizes.

5. The authors generated a genome with 13 chromosomes from 63.74% of assembled sequences. Does this chromosome number match the other beetle genomes in the same family or genus level? There are still a lot of unmapped sequences. Could the authors identify those sequences?

Are there any genes found on them?

6. There are some formatting issues that need to be polished, e.g. spacing and numerical value presentation. Please kindly proofread.

Is the rationale for creating the dataset(s) clearly described?
No

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: evolution, genomics, arthropods, myriapods

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 08 April 2026

https://doi.org/10.21956/wellcomeopenres.28829.r151925

?

Henrik Lantz
SciLifeLab, Uppsala University, Uppsala, Sweden

This manuscript presents the genome sequence of a small beetle, Tachyporus hypnorum, assembled as a part of the Darwin Tree of Life project. The species is widespread in Europe and the specimens used here were sampled in England. Despite being only 3-4 mm in length, the species is distinctive enough to be, with experience, identified in the field.

The genome is the first high-quality genome assembled for the genus and thus constitutes an important resource not only for studies of the species but also for comparative studies at higher taxonomical levels. Methods are well-described and follow a high international standard. The assembly is of high quality and is certainly useful for most studies, including looking at genomic rearrangements which requires high contiguity, but fails to meet Earth Biogenome Project

standards as only 63.74% of the assembly could be assigned to chromosomes (should be at least 90%).

I have two main comments and a small one, all intended to improve clarity and transparency.

1. It is not clearly stated what was done to facilitate re-confirmation of the identity of the specimens used in the study. Because of the small size of the species, whole individuals needed to be used and this of course limits what can be preserved, but this should then be explained in the text. If someone wants to confirm the identity of the specimens used, for example if a need arises to describe sub-species, what remains for that researcher to base conclusions on? Was anything at all saved from the specimens, and where is that material stored in that case? Is the photo detailed enough for an expert to base identification on? Where is that photo stored in that case? The use of DNA barcoding to verify the identity of the specimens does strengthen the identification, but is of limited use for within species studies. Remaining voucher material and/or photos should be stored in a Natural History Museum, and enough information should be given in the manuscript to allow for a reader to find that material.

2. 13 large scaffolds were considered chromosome-level, and this number correlates well with the expected number of chromosomes given in Genomes On A Tree, based on close relatives. However, only 63.74% of the assembly is assigned to these chromosomes, and no effort is done in the manuscript to explain what the remaining 36.26% are. As I understand it, heterochromatin is also known from Coleoptera and can constitute a large part of the genome, and perhaps the unassigned part of the assembly is simply that, but then I would like to see this stated in the manuscript.

3. Figure 6, BlobToolKit plot: It seems most likely to me that the plot is done after decontamination, but still, it would be nice to have that clearly stated. I also find it likely that the smaller blobs with no hits are the contigs that could not be assigned to chromosomes, but would appreciate if that also could be detailed in the figure text.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, genome annotation, taxonomy, systematics, phylogeny.

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.