DATA NOTE

The genome sequence of the cryptorhynchine weevil, *Kyklioacalles roboris* (Stüben, 2003) (Coleoptera: Curculionidae)

[version 1; peer review: 2 approved]

Liam M. Crowley[1],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]University of Oxford, Oxford, England, UK

v1 **First published:** 06 Jan 2026, **11**:11
https://doi.org/10.12688/wellcomeopenres.25416.1
Latest published: 06 Jan 2026, **11**:11
https://doi.org/10.12688/wellcomeopenres.25416.1

Abstract

We present a genome assembly from an individual male *Kyklioacalles roboris* (cryptorhynchine weevil; Arthropoda; Insecta; Coleoptera; Curculionidae). The assembly contains two haplotypes with total lengths of 698.93 megabases and 609.02 megabases. Most of haplotype 1 (87.44%) is scaffolded into 20 chromosomal pseudomolecules, including the X and Y sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 20.21 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

Kyklioacalles roboris; cryptorhynchine weevil; genome sequence; chromosomal; Coleoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 06 Jan 2026	✓ view	✓ view

1. **Nikoletta Andrea Nagy**, University of Debrecen, Debrecen, Hungary

2. **João Pedro Marques**, BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology (Ringgold ID: 637326), Vairão, Portugal
University of Porto Research Centre in Biodiversity and Genetic Resources (Ringgold ID: 123201), Vairão, Portugal
Juliana Sofia Alves, BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology, Vairão, Portugal
University of Porto Research Centre in Biodiversity and Genetic Resources (Ringgold

ID: 123201), Vairão, Portugal
University of Porto Faculty of Sciences
(Ringgold ID: 131674), Porto, Portugal

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;

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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of the cryptorhynchine weevil, *Kyklioacalles roboris* (Stüben, 2003) (Coleoptera: Curculionidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:11
https://doi.org/10.12688/wellcomeopenres.25416.1

First published: 06 Jan 2026, **11**:11 https://doi.org/10.12688/wellcomeopenres.25416.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Curculionoidea; Curculionidae; Cryptorhynchinae; *Kyklioacalles*; *Kyklioacalles roboris* (Stüben, 2003) (NCBI:txid501664)

Background

Kyklioacalles roboris (Curtis, 1834) is a small cryptorhynchine weevil found in broadleaf woodland litter, especially among dead leaves and twig accumulations around oak growths such as witches' brooms and epicormic shoots (Gurney, 2018). Adults have dense tufted scales on the pronotum and elytra, and the pronotum shows pronounced lateral bumps (Gurney, 2018). In Britain it is the least often encountered of the three similar species in this group (Gurney, 2018).

The genus *Kyklioacalles* was separated from *Acalles* by Stüben in 1999 based on male endophallus characters, with later work refining generic limits and intrageneric groupings (Stüben & Astrin, 2010). The species belongs to the hidden snout weevils (Curculionidae: Cryptorhynchinae), a hyperdiverse lineage with many thousands of species worldwide. Adults commonly feign death by folding the legs and tucking the rostrum into a ventral groove. Phylogenetic and biogeographical analyses indicate a Late Cretaceous origin for Cryptorhynchinae with later dispersal into the Western Palaearctic (Letsch *et al.*, 2020).

We present a chromosome-level genome sequence for *Kyklioacalles roboris*, based on a specimen collected from Wytham Woods, Oxfordshire, UK (Figure 1), as part of the Darwin Tree of Life Project. This genome sequence will enable molecular investigations of cryptorhynchine evolution across this hyperdiverse clade.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Kyklioacalles roboris* (specimen ID Ox003277, ToLID icKykRobo1; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude –1.338) on 2022-10-31. The specimen was collected and identified by Liam Crowley (University of Oxford). Sample metadata were collected in line with the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

Figure 1. Photograph of the *Kyklioacalles roboris* (icKykRobo1) 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 icKykRobo1 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 in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. Sheared DNA was purified by manual 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 6.51 ng/μL and a yield of 299.46 ng, with a fragment size of 13.2 kb.

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 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 25M 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 whole organism of the icKykRobo1 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 in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The organelle genomes were 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 41 breaks and 48 joins. This reduced the scaffold count by 0.7% and increased the scaffold N50 by 2.1%. 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 both 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 *Kyklioacalles roboris* specimen generated 40.45 Gb (gigabases) from 3.57 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 616.98 Mb, with a heterozygosity of 0.61% and repeat content of 46.12% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 63× coverage. Hi-C sequencing produced 293.49 Gb from 1 943.62 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

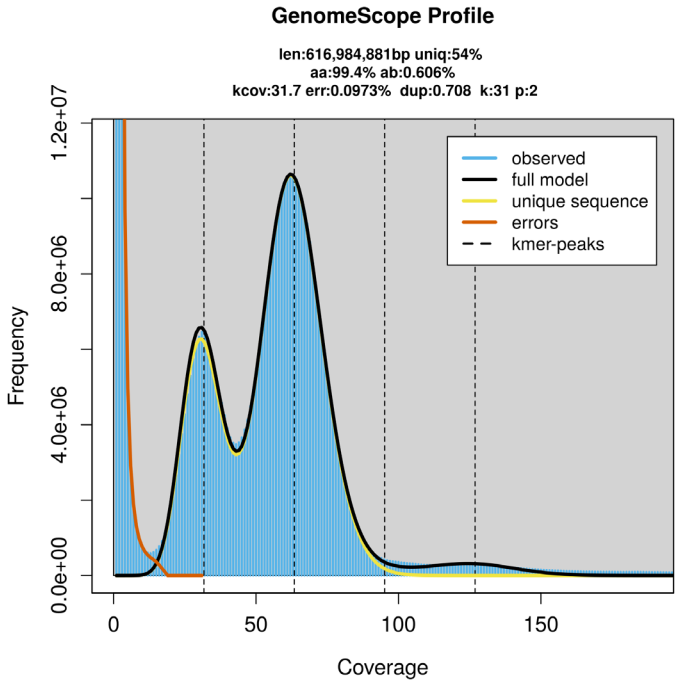

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB76839.

Platform	PacBio HiFi	Hi-C
ToLID	icKykRobo1	icKykRobo1
Specimen ID	Ox003277	Ox003277
BioSample (source individual)	SAMEA113425821	SAMEA113425821
BioSample (tissue)	SAMEA113426007	SAMEA113426007
Tissue	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq 6000
Run accessions	ERR13304160	ERR13317840
Read count total	3.57 million	1 943.62 million
Base count total	40.45 Gb	293.49 Gb

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 698.93 Mb in 964 scaffolds, with 467 gaps, and a scaffold N50 of 30.38 Mb (Table 2).

Most of the haplotype 1 assembly sequence (87.44%) was assigned to 20 chromosomal-level scaffolds, representing 18 autosomes and the X and Y sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosomes X and Y were assigned based on read coverage statistics and Hi-C signal.

The mitochondrial genome was also assembled (length 20.21 kb, OZ125640.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Table 2. Genome assembly statistics.

Assembly name	icKykRobo1.hap1.1	icKykRobo1.hap2.1
Assembly accession	GCA_964211925.1	GCA_964211975.1
Assembly level	chromosome	scaffold
Span (Mb)	698.93	609.02
Number of chromosomes	20	scaffold-level
Number of contigs	1 431	689
Contig N50	2.97 Mb	4.11 Mb
Number of scaffolds	964	424
Scaffold N50	30.38 Mb	30.51 Mb
Longest scaffold length (Mb)	56.91-	
Sex chromosomes	X and Y	-
Organelles	Mitochondrion: 20.21 kb	-

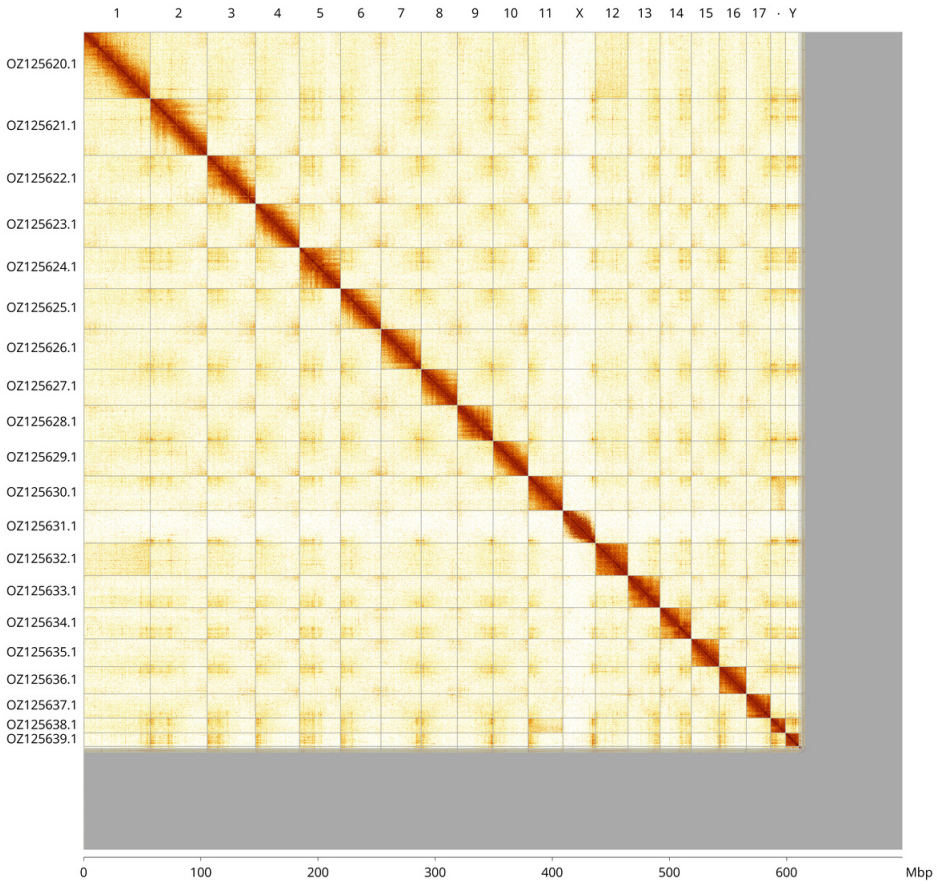

Figure 3. Hi-C contact map of the *Kyklioacalles roboris* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Kyklioacalles roboris* icKykRobo1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ125620.1	1	56.91	34.50
OZ125621.1	2	48.63	34.50
OZ125622.1	3	41.22	35
OZ125623.1	4	37.52	34.50
OZ125624.1	5	35.12	35
OZ125625.1	6	34.58	35
OZ125626.1	7	34.19	34.50
OZ125627.1	8	30.92	34
OZ125628.1	9	30.38	34.50
OZ125629.1	10	30.07	34.50
OZ125630.1	11	30.03	35
OZ125632.1	12	27.70	35.50
OZ125633.1	13	27.49	35
OZ125634.1	14	26.69	35
OZ125635.1	15	23.83	35.50
OZ125636.1	16	23.10	35
OZ125637.1	17	20.75	35
OZ125638.1	18	12.72	36
OZ125631.1	X	27.86	34.50
OZ125639.1	Y	11.46	35.50

For haplotype 1, the estimated QV is 61.6, and for haplotype 2, 63.0. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.2. The *k*-mer completeness is 89.18% for haplotype 1, 83.76% for haplotype 2, and 98.94% for the combined haplotypes (Figure 4).

BUSCO analysis using the endopterygota_odb10 reference set (n = 2 124) identified 98.7% of the expected gene set (single = 97.8%, duplicated = 0.9%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

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 haplotype 1, is **6.7.Q61**.

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

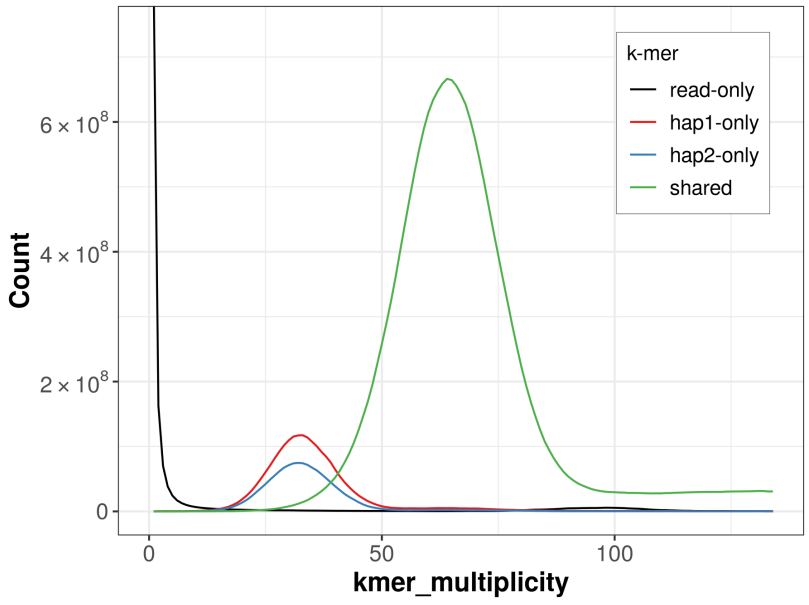

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

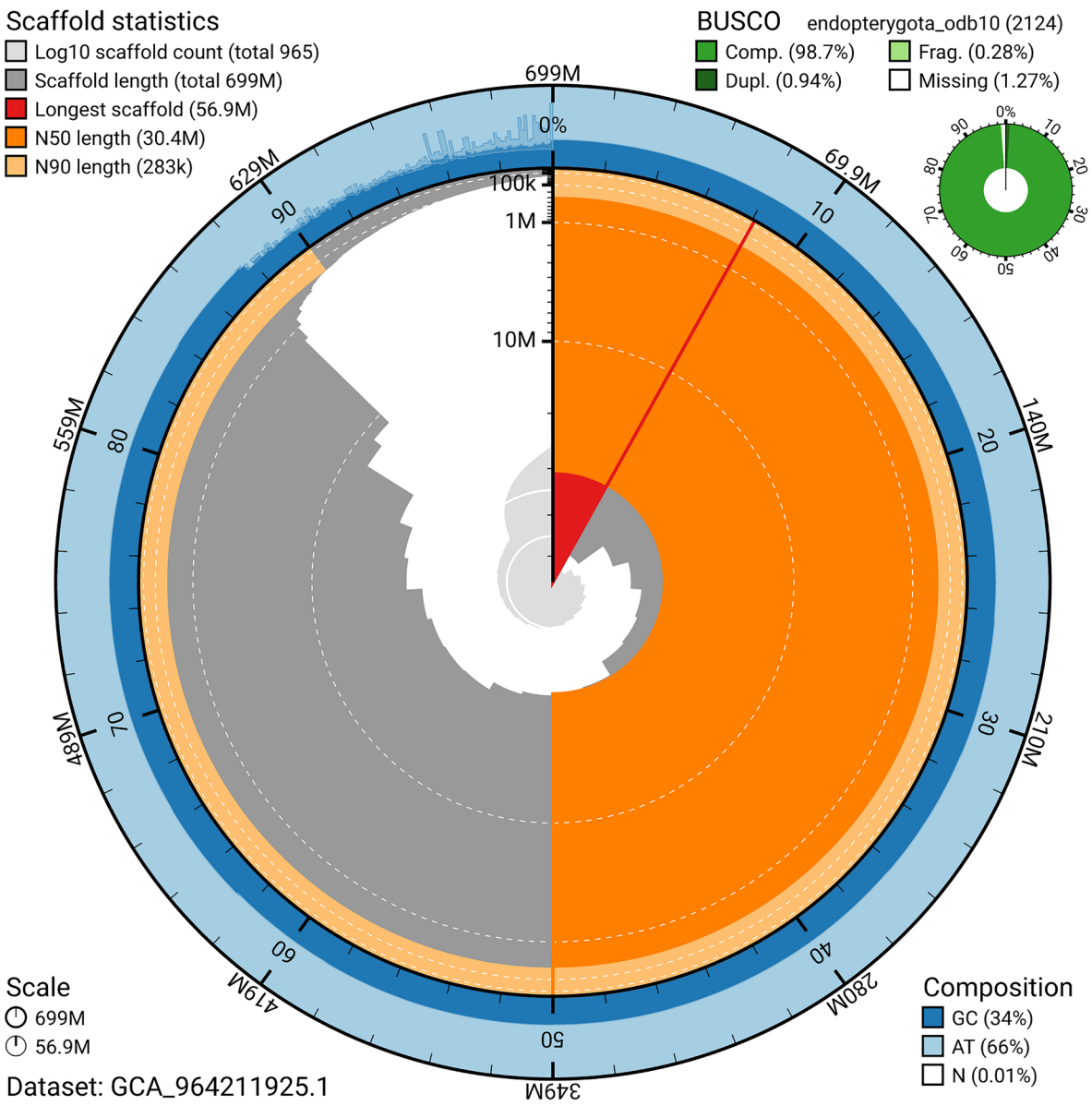

Figure 5. Assembly metrics for icKykRobo1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

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

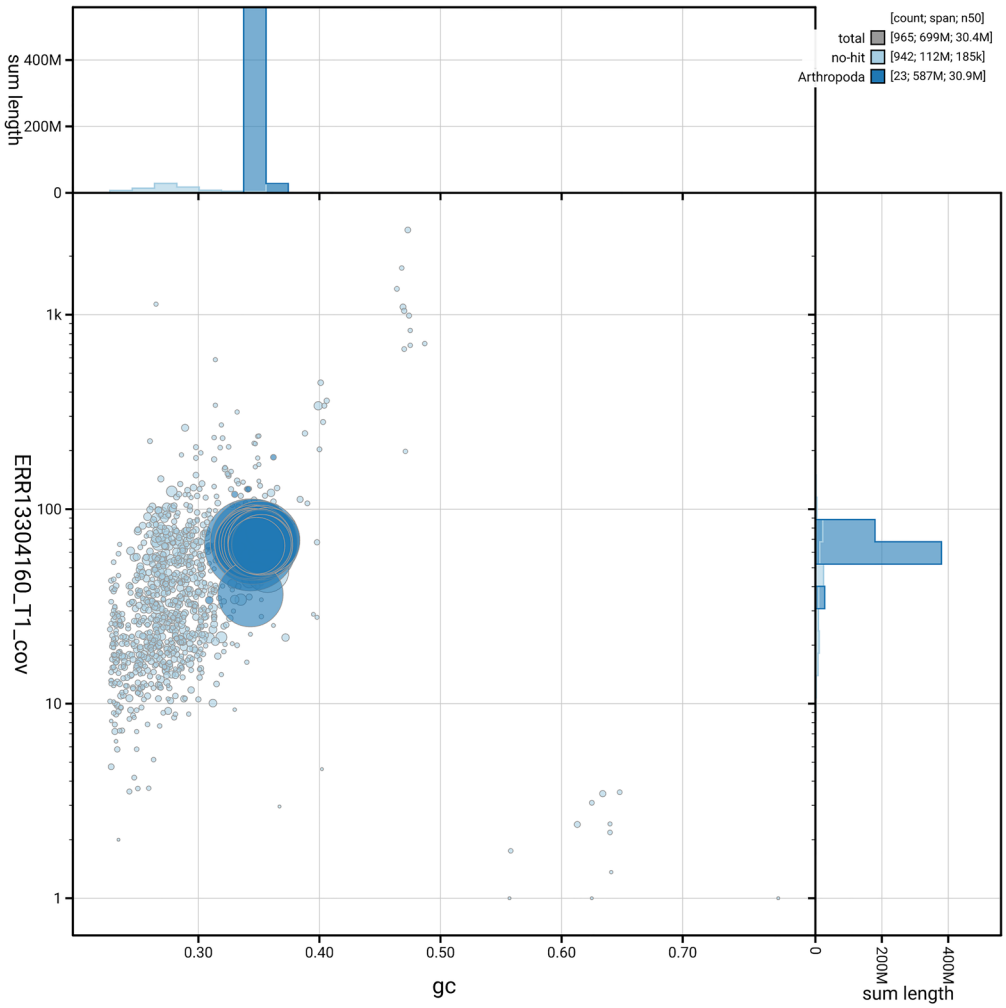

Figure 6. BlobToolKit GC-coverage plot for icKykRobo1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Kyklioacalles roboris* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q61	6.C.Q40
Contig N50 length	2.97 Mb	≥ 1 Mb
Scaffold N50 length	30.38 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 61.6; haplotype 2: 63.0; combined: 62.2	≥ 40
k-mer completeness	Haplotype 1: 89.18%; Haplotype 2: 83.76%; combined: 98.94%	≥ 95%
BUSCO	C:98.7% [S:97.8%; D:0.9%]; F:0.3%; M:1.0%; n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	87.44%	≥ 90%

- 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: Kyklioacalles roboris. Accession number PRJEB76839. The genome sequence is released openly for reuse. The *Kyklioacalles roboris* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

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 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	0.19.8-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	2.24-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	23.10.0	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	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	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 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

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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
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

Gurney M: **An identification guide to typical weevils (Curculionoidea): species guide 6, version 1.04**. 2018.
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

Letsch H, Balke M, Toussaint EFA, *et al.*: **Historical biogeography of the hyperdiverse hidden snout weevils (Coleoptera, Curculionidae, Cryptorhynchinae).** *Syst Entomol.* 2020; **45**(2): 312–326.
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

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**

Stüben PE, Astrin JJ: **Molecular phylogeny of the weevil genus *Kyklioacalles* Stüben, with descriptions of a new subgenus *Glaberacalles* and two new species (Curculionidae: Cryptorhynchinae).** *Zootaxa.* 2010; **2662**: 28–52.
Publisher Full Text

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

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

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

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

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 11 February 2026

https://doi.org/10.21956/wellcomeopenres.28002.r145613

✔ **João Pedro Marques**
[1] BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology (Ringgold ID: 637326), Vairão, Porto, Portugal
[2] University of Porto Research Centre in Biodiversity and Genetic Resources (Ringgold ID: 123201), Vairão, Porto, Portugal

Juliana Sofia Alves
[1] BIOPOLIS Association - Research Network on Biodiversity and Evolution Biology, Vairão, Porto, Portugal
[2] University of Porto Research Centre in Biodiversity and Genetic Resources (Ringgold ID: 123201), Vairão, Porto, Portugal
[3] University of Porto Faculty of Sciences (Ringgold ID: 131674), Porto, Porto District, Portugal

This manuscript presents a chromosome-level genome assembly for *Kyklioacalles roboris* generated within the Darwin Tree of Life framework. The study follows well-established and robust protocols, and the resulting assembly is of high technical quality, with strong contiguity, high consensus accuracy, and excellent BUSCO completeness. The data are clearly deposited and accessible, and the manuscript is generally well written and easy to follow.
Overall, this genome represents a valuable resource for future evolutionary and comparative genomic studies of cryptorhynchine weevils and Curculionidae more broadly.
However, I have a small number of comments that should be addressed to improve clarity, accuracy, and usability of the resource.
General comments
Although I recognise the intentionally concise nature of Darwin Tree of Life data notes, the manuscript would benefit from a brief element of biological interpretation. For example, it would be helpful to indicate whether the inferred number of pseudo-chromosomes is consistent with expectations for this species or closely related taxa, and whether the reported GC content is typical for cryptorhynchine weevils. In addition, a short statement highlighting potential future applications of this genome (e.g. comparative genomics, phylogenomics, or studies of cryptic diversification) would increase its value to readers.
Specific comments

1. **Nucleic acid extraction**
   The authors state that an appropriate extraction protocol was selected based on sample characteristics. Please clarify which criteria were used to choose this protocol.
2. **Figure 1**
   Please consider adding a physical scale or size reference to the specimen photograph.
3. **Assembly curation section**
   It is not explicit whether any duplication or haplotypic purging step was applied. If purging was handled implicitly by hifiasm (e.g. through default or optional parameters), this should be clearly stated, as haplotypic purging is a standard step in current assembly workflows.
4. **Table 2**
   The line "Longest scaffold length (Mb)" contains formatting issues that should be corrected.
5. **Figure 3 (Hi-C contact map)**
   Chromosome 18 appears to be missing from the top legend. Please update the figure accordingly.
6. **Table 4**
   The entry “Total BUSCO groups searched (2124)” is incorrectly formatted and should be corrected.
7. **Figure 6**
   The Y-axis label appears to be inverted or incorrectly oriented. Please check and correct.

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: Evolutionary Biology, Population Genomics, Bioinformatics

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

Reviewer Report 28 January 2026

https://doi.org/10.21956/wellcomeopenres.28002.r144321

Nikoletta Andrea Nagy
University of Debrecen, Debrecen, Hungary

The manuscript is well written, and the genome presented is of high quality and can contribute to future research on weevils. I have only a few suggestions for the authors.
Background
“In Britain it is the least often encountered of the three similar species in this group” – please list the other two species
Figure 1
Please put a scale on the photo
Assembly statistics
I would recommend including whether the assembled mitochondrial genome is an incomplete linear or a complete circular sequence.
Table 2
Longest scaffold length – the “-“ character should be placed in the “icKykRobo1.hap2.1” column

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: arthropod genomics, transcriptomics, behavioural genetics

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.