DATA NOTE

The genome sequence of the Acorn Weevil, *Curculio glandium* (T.Marsham, 1802) (Coleoptera: Curculionidae)

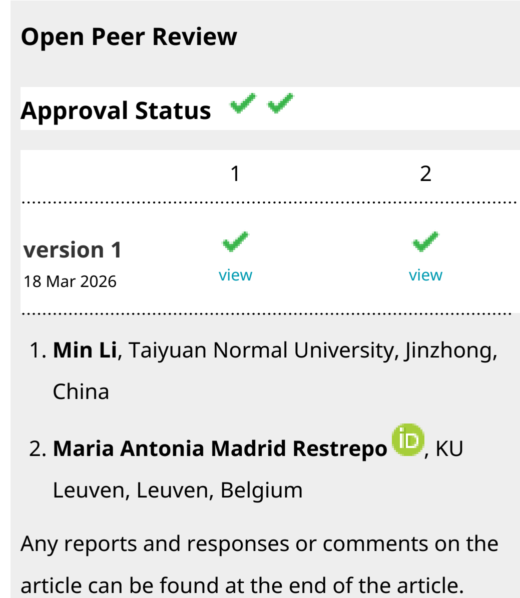
[version 1; peer review: 2 approved]

Douglas Boyes[1+], Liam M. Crowley[2], James McCulloch[2,3], Clare Boyes[4], 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]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK
[3]Wellcome Sanger Institute, Hinxton, England, UK
[4]Independent researcher, Welshpool, Wales, USA

[+] Deceased author

v1 **First published:** 18 Mar 2026, **11**:178
https://doi.org/10.12688/wellcomeopenres.26157.1
Latest published: 18 Mar 2026, **11**:178
https://doi.org/10.12688/wellcomeopenres.26157.1



Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 18 Mar 2026	✓ view	✓ view

1. **Min Li**, Taiyuan Normal University, Jinzhong, China
2. **Maria Antonia Madrid Restrepo**, KU Leuven, Leuven, Belgium

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

We present a genome assembly from an individual female *Curculio glandium* (Acorn Weevil; Arthropoda; Insecta; Coleoptera; Curculionidae). The genome sequence has a total length of 1 121.34 megabases. Most of the assembly (97.77%) is scaffolded into 13 chromosomal pseudomolecules, including the X sex chromosome. The mitochondrial genome has also been assembled, with a length of 21.61 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

Curculio glandium, Acorn Weevil, genome sequence, chromosomal, Coleoptera

This article is included in the Tree of Life gateway.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Boyes D: Investigation, Resources; **Crowley LM**: Investigation, Resources; **McCulloch J**: Investigation, Resources; **Boyes C**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Boyes D *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Boyes D, Crowley LM, McCulloch J *et al.* **The genome sequence of the Acorn Weevil, *Curculio glandium* (T.Marsham, 1802) (Coleoptera: Curculionidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:178 https://doi.org/10.12688/wellcomeopenres.26157.1

First published: 18 Mar 2026, **11**:178 https://doi.org/10.12688/wellcomeopenres.26157.1

Species taxonomy
Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Curculionoidea; Curculionidae; Curculioninae; Curculionini; *Curculio*; *Curculio glandium* (T.Marsham, 1802) (NCBI:txid197013).

Background
Curculio glandium is one of the acorn weevils in the family Curculionidae. It occurs throughout Europe, and is locally common in England and Wales (GBIF Secretariat, 2025). There are also records from eastern Canada and the United States (Udaka & Sinclair, 2014).

Curculio glandium is 4–8 mm long and brownish in colour, with a long rostrum and ovipositor. As its common name suggests, this weevil relies on acorns to feed and reproduce. *C. glandium* appears to prefer larger acorns than other acorn weevil species, and consequently breeds later in the season than similar species. The species is unusual in that the female makes galleries in acorns, and then lays eggs in these excavations. A female can lay up to eight eggs per acorn and has been found to make more galleries than needed; and despite its long rostrum, these galleries are shallow (Reut *et al.*, 2021). As the females do not appear to produce oviposition-deterring pheromones, high densities can occur in some acorns. However, at high densities, it appears that the larvae cannibalise others as a strategy to reduce competition (Reut *et al.*, 2023).

The genome of *Curculio glandium* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all the named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *Curculio glandium* based on one specimen from Wytham Woods, Oxfordshire, UK.

Methods
Sample acquisition and DNA barcoding
The specimen used for genome sequencing was an adult female *Curculio glandium* (specimen ID Ox000996, ToLID icCurGlan1; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude −1.338) on 2020-11-21. The specimen was collected by Douglas Boyes and identified by Liam Crowley. A second specimen was used for Hi-C and RNA sequencing (specimen ID Ox002538, ToLID icCurGlan2). It was collected from the same location on 2022-07-22. The specimen was collected by James McCulloch and Liam Crowley, and identified by Liam Crowley.

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 icCurGlan1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the thorax and abdomen was homogenised by powermashing 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 8.04 ng/μL and a yield of 3 135.60 ng.

RNA was extracted from whole organism tissue of icCurGlan2 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.

Figure 1. Photograph of the *Curculio glandium* (icCurGlan1) specimen used for genome sequencing.

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 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 tissue from the icCurGlan2 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.

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 6000, generating 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. 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 MerquryFK (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 74 breaks and 139 joins. This reduced the scaffold count by 6.1%, increased the scaffold N50 by 12.7%, and reduced the total assembly length by 1.0%. 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 MerquryFK 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

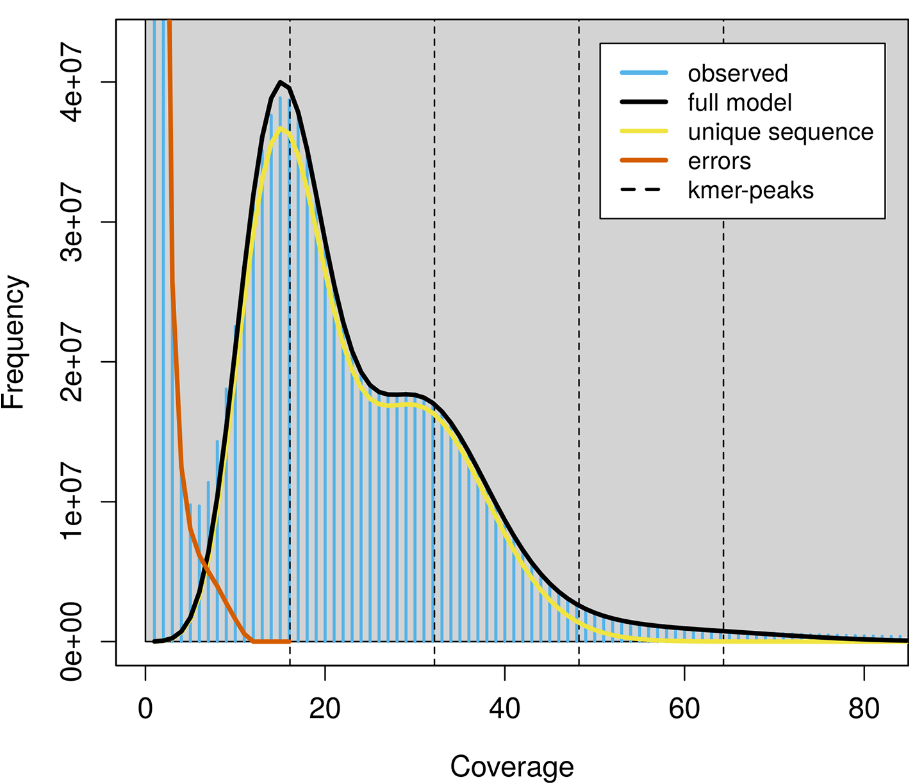


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB80665.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	icCurGlan1	icCurGlan2	icCurGlan2
Specimen ID	Ox000996	Ox002538	Ox002538
BioSample (source individual)	SAMEA8603219	SAMEA112232732	SAMEA112232732
BioSample (tissue)	SAMEA8603796	SAMEA112233224	SAMEA112233224
Tissue	thorax and abdomen	whole organism	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR13762703; ERR13762704	ERR13766872	ERR13766873
Read count total	3.06 million	567.06 million	91.40 million
Base count total	37.02 Gb	85.63 Gb	13.80 Gb

Table 2. Genome assembly statistics.

Assembly name	icCurGlan1.1
Assembly accession	GCA_965648435.1
Alternate haplotype accession	GCA_965648465.1
Assembly level	chromosome
Span (Mb)	1 121.34
Number of chromosomes	13
Number of contigs	1 177
Contig N50	2.42 Mb
Number of scaffolds	275
Scaffold N50	99.31 Mb
Sex chromosomes	X
Organelles	Mitochondrion: 21.61 kb

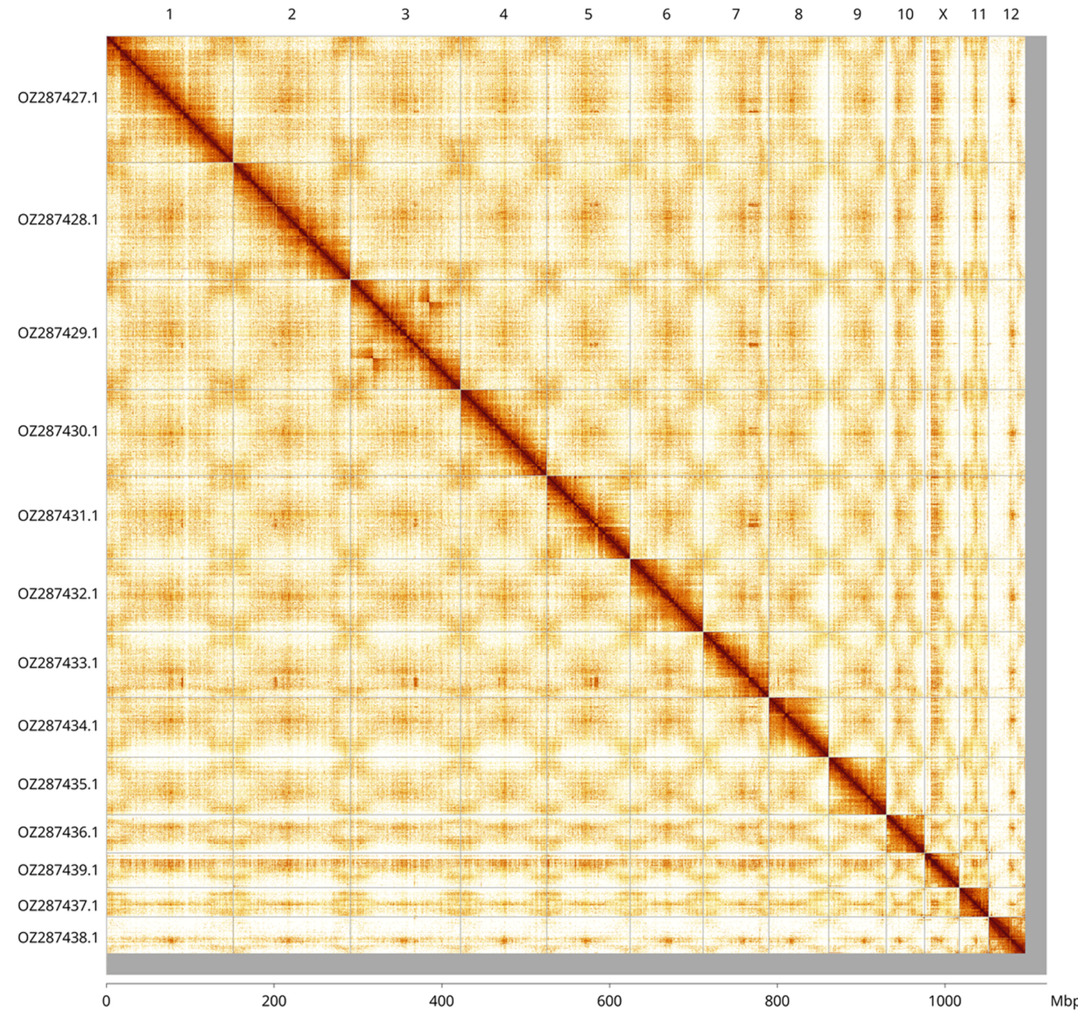


Figure 3. Hi-C contact map of the *Curculio glandium* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

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 *Curculio glandium* specimen generated 37.02 Gb (gigabases) from 3.06 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 052.56 Mb, with a

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Curculio glandium* icCurGlan1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ287427.1	1	151.40	35
OZ287428.1	2	139.72	35
OZ287429.1	3	131.33	34.50
OZ287430.1	4	102.82	35
OZ287431.1	5	99.31	35
OZ287432.1	6	87.17	35
OZ287433.1	7	78.20	35
OZ287434.1	8	71.46	35.50
OZ287435.1	9	68.65	35.50
OZ287436.1	10	45.93	35.50
OZ287437.1	11	41.23	35.50
OZ287438.1	12	34.93	36.50
OZ287439.1	X	44.13	35

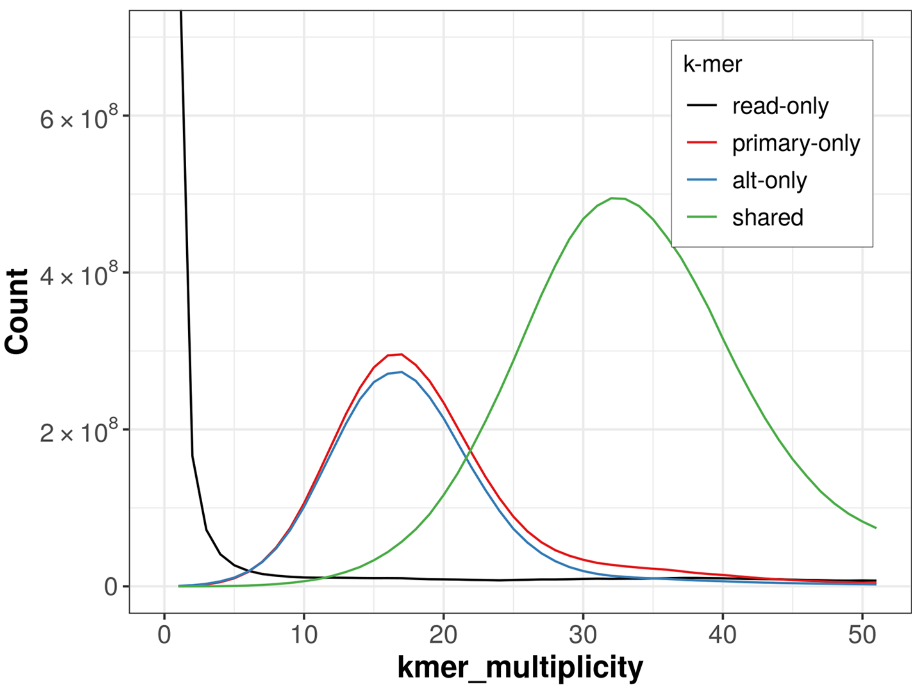


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

heterozygosity of 1.88% and repeat content of 52.50% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 32× coverage. Hi-C sequencing produced 85.63 Gb from 567.06 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.

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 1 121.34 Mb in 275 scaffolds, with 902 gaps, and a scaffold N50 of 99.31 Mb (Table 2).

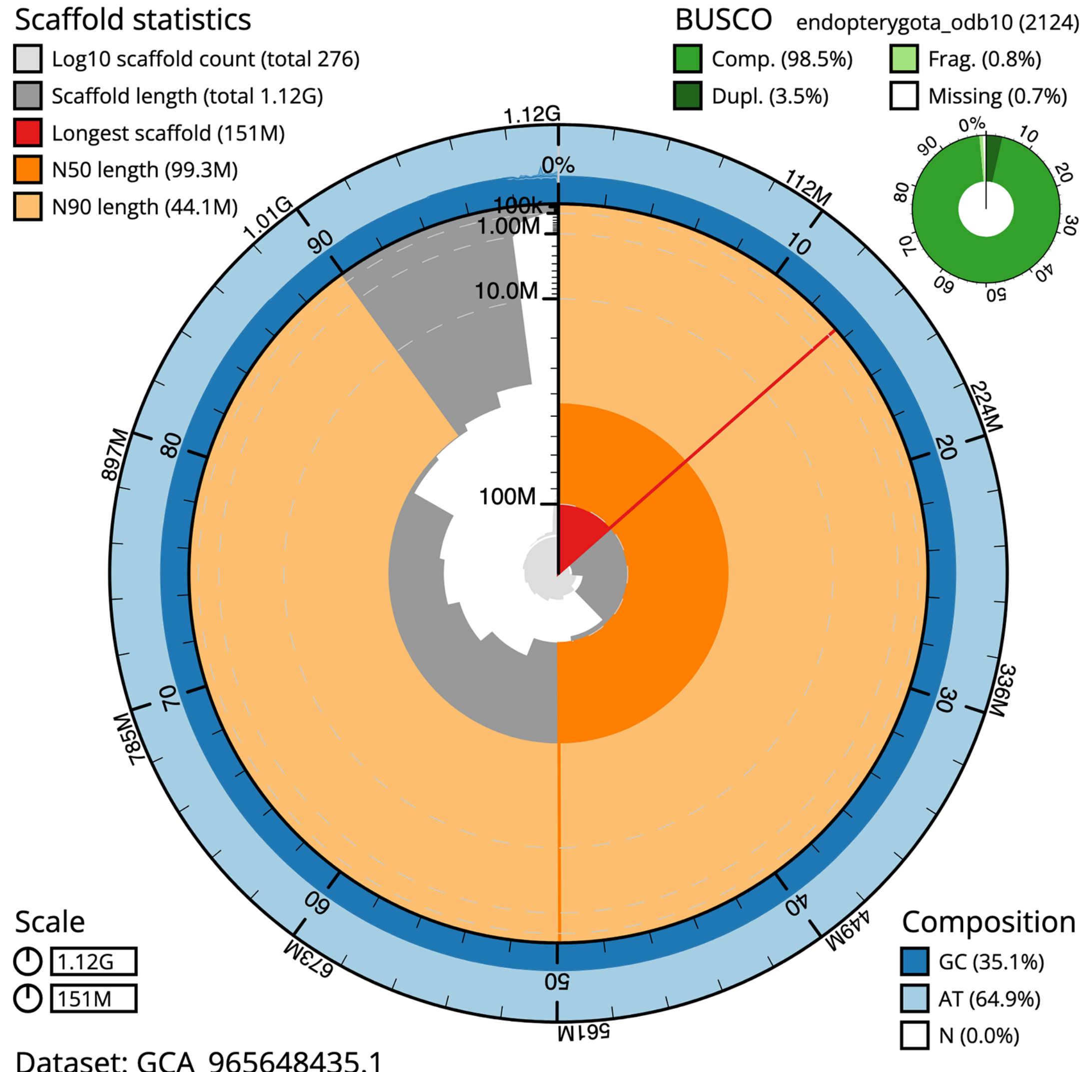


Figure 5. Assembly metrics for icCurGlan1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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.

Most of the assembly sequence (97.77%) 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; Table 3). A haplotypic inversion was observed in the region on chromosome 3 (23.9–94.0 Mbp). The X chromosome was assigned based on the presence of ancestral X BUSCO genes.

The mitochondrial genome was also assembled (length 21.61 kb, OZ287440.1). 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 59.7. The *k*-mer completeness is 71.23% for the primary assembly, 68.56% for the alternate haplotype, and 96.98% for the combined assemblies (Figure 4).

BUSCO v.5.8.3 analysis using the endopterygota_odb10 reference set (n = 2 124) identified 98.5% of the expected gene set (single = 95.0%, duplicated = 3.5%). 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.

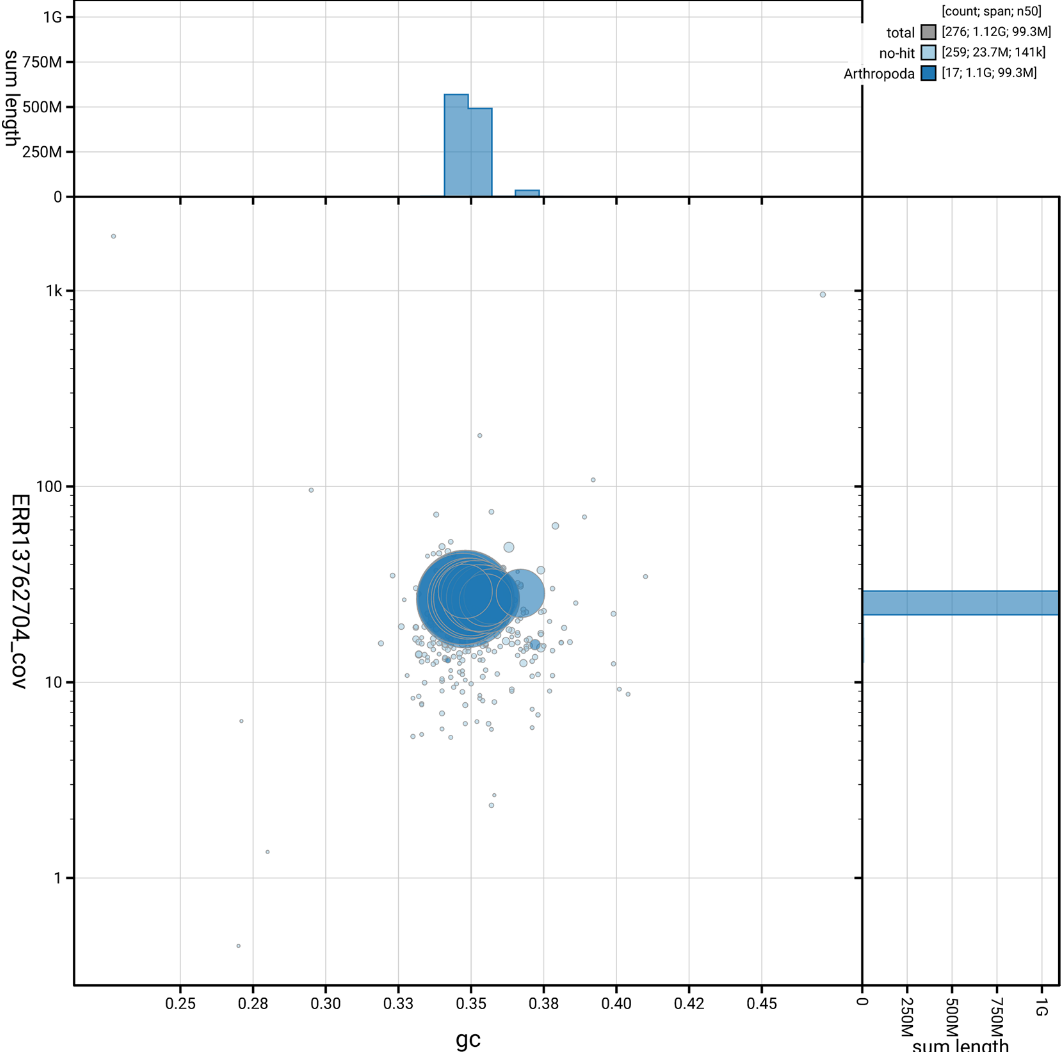


Figure 6. BlobToolKit blob plot for icCurGlan1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total 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 *Curculio glandium* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q59	6.C.Q40
Contig N50 length	2.42 Mb	≥ 1 Mb
Scaffold N50 length	99.31 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.2; alternate: 60.0; combined: 59.7	≥ 40
k-mer completeness	Primary: 71.23%; alternate: 68.56%; combined: 96.98%	≥ 95%
BUSCO	C:98.5% [S:95.0%, D:3.5%], F:0.8%, M:0.7%, n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.77%	≥ 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.

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.C.Q59**, meeting the recommended reference standard.

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

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: Curculio glandium (acorn weevil). Accession number PRJEB80665. The genome sequence is released openly for reuse. The *Curculio glandium* 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	5.8.3	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
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	v0.8.0	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.
Publisher 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.
Publisher 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

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.
Publisher 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.
Publisher Full Text

GBIF Secretariat: ***Curculio glandium* (Marsham, 1802) in GBIF Backbone Taxonomy.** 2025.
Publisher Full Text | 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).
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

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

Reut M, Bonal R, Chrabąszcz M, *et al.*: **Cannibalism as competition strategy in larvae of the acorn weevil *Curculio glandium* (Coleoptera: Curculionidae).** *Diversity.* 2023; **15**(2): 145.
Publisher Full Text

Reut M, Chrabąszcz M, Moniuszko H: **Timing Is everything. Temporal and spatial niche segregation in *Curculio* spp. (Coleoptera: Curculionidae) associated with oak trees.** *Insects.* 2021; **12**(8): 687.
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 Research.* 2024; **9**: 339.
Publisher Full Text

Udaka H, Sinclair BJ: **The overwintering biology of the acorn weevil, *Curculio glandium* in southwestern Ontario.** *J. Therm. Biol.* 2014; **44**: 103–109.
PubMed Abstract | 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 20 April 2026

https://doi.org/10.21956/wellcomeopenres.28805.r151236

© **2026 Madrid Restrepo M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Maria Antonia Madrid Restrepo
KU Leuven, Leuven, Belgium

This data note presents a chromosome-level genome assembly for the acorn weevil *Curculio glandium*, generated within the Darwin Tree of Life project. The authors assembled a 1.12 Gb genome from a female specimen collected in Wytham Woods, with 97.77% of the sequence assigned to 13 chromosomal pseudomolecules, including the X chromosome, and they also recovered the mitochondrial genome. The study combines PacBio HiFi, Hi-C, and RNA-seq data, and reports strong assembly statistics, including a scaffold N50 of 99.31 Mb, BUSCO completeness of 98.5%, and a combined assembly QV of 59.7. Overall, this is a useful and technically solid genomic resource that will be valuable for comparative genomics and future work on Curculionidae. The rationale for creating the dataset is clearly described. The protocols are appropriate and the work appears technically sound. The methods and materials are, in general, described in sufficient detail to allow replication. The manuscript provides specimen identifiers, sequencing accessions, software names and versions, and the main analytical steps used in processing and assembly. Public deposition of raw reads and assembly files also supports reuse. The datasets are clearly presented and accessible.
First, the authors should correct the obvious affiliation error listing “Welshpool, Wales, USA”; this should be changed to “Welshpool, Wales, UK.” This is minor and editorial, but it should be fixed in the revised version. Second, the introduction would benefit from a brief comparison with existing genomic resources for related weevil species, especially any available *Curculio* or closely related Curculionidae assemblies. Third, the statement that a “haplotypic inversion was observed” on chromosome 3 would benefit from a little more explanation. At present, the manuscript notes the inversion interval but does not explain how confidently this represents a biological haplotypic inversion rather than an assembly or phasing feature. Points that should be addressed to make the article scientifically sound: in my view, none of the issues rise to the level of a major scientific flaw, and the article is already scientifically sound as a genome data note. The only point that clearly must be corrected is the affiliation error.

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: Eco-evolutionary genomics, computational biology, population genetics, bioinformatics

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

Reviewer Report 03 April 2026

https://doi.org/10.21956/wellcomeopenres.28805.r151239

© 2026 Li M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Min Li

Taiyuan Normal University, Jinzhong, China

This manuscript is a high-quality data note reporting a chromosome-level genome assembly of the acorn weevil *Curculio glandium*. The work follows standard best practices for eukaryotic genome sequencing and assembly, produces a robust, well-validated genomic resource, and fully meets the publication criteria for Wellcome Open Research. I suggest only a few minor changes:

1. Affiliation 4: Welshpool, Wales, USA → Welshpool, Wales, UK.
2. The introduction lacks a brief summary of existing genomic resources for closely related Curculio species. To add 1–2 sentences comparison would strengthen the novelty of this chromosome-level assembly.

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.

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.