



DATA NOTE

The genome sequence of a leaf beetle, *Galeruca laticollis*

Sahlberg, C.R., 1838

[version 1; peer review: 2 approved]

Roger Booth¹, 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

¹Natural History Museum, London, England, UK

V1 First published: 15 Oct 2024, 9:594
<https://doi.org/10.12688/wellcomeopenres.23195.1>
Latest published: 15 Oct 2024, 9:594
<https://doi.org/10.12688/wellcomeopenres.23195.1>

Abstract

We present a genome assembly from an individual leaf beetle, *Galeruca laticollis* (Arthropoda; Insecta; Coleoptera; Chrysomelidae). The genome sequence has a total length of 2,154.60 megabases. Most of the assembly (99.92%) is scaffolded into 12 chromosomal pseudomolecules, including the X and Y sex chromosomes. The mitochondrial genome has also been assembled and is 19.98 kilobases in length. Gene annotation of this assembly on Ensembl identified 32,229 protein-coding genes.

Keywords

Galeruca laticollis, leaf beetle, genome sequence, chromosomal, Coleoptera




This article is included in the [Tree of Life gateway](#).

Open Peer Review

Approval Status

	1	2
version 1		
15 Oct 2024	view	view

1. **Željko Tomanović**, Serbia University of Belgrade, Belgrade, Serbia
2. **Jose Manuel Rico Porras** , Universidad de Jaen, Andalusia, Spain

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, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, <https://doi.org/10.35802/206194>] and the Darwin Tree of Life Discretionary Award [218328, <https://doi.org/10.35802/218328>]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2024 Booth R *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: Booth R, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of a leaf beetle, *Galeruca laticollis* Sahlberg, C.R., 1838 [version 1; peer review: 2 approved]** Wellcome Open Research 2024, 9:594 <https://doi.org/10.12688/wellcomeopenres.23195.1>

First published: 15 Oct 2024, 9:594 <https://doi.org/10.12688/wellcomeopenres.23195.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Chrysomeloidea; Chrysomelidae; Galerucinae; Galerucini; Galerucites; *Galeruca*; *Galeruca laticollis* Sahlberg, C.R., 1838 (NCBI:txid2017130).

Background

Galeruca laticollis Sahlberg, C.R., 1838 is a leaf beetle in the family Chrysomelidae. It is ovate in shape with a strongly transverse pronotum, 6–9 mm in length, and can be recognised in the field by its black head, antennae, and legs on a brownish body rather than having the wholly black body of its congener, *G. tanacetii* (Linnaeus) (Duff, 2016).

In recent years, adults have been found on a variety of plants, including thistles, growing in fenland. Larvae in the UK have been found on the leaves of Common Meadow-rue *Thalictrum flavum* (Duff, 2016).

Galeruca laticollis was regarded formerly as very rare in the UK and listed as RDBI – Endangered by Hyman and Parsons (1992) and possibly extinct, having not been seen since 1919 when it was found in Dorset. However, in the mid-1990s it was rediscovered at the same site from which the present material was collected in East Norfolk, where there is an apparently stable population. Outside the UK, it occurs in most of Europe and extends eastwards into temperate parts of Asia.

Here we present the first genome sequence for *G. laticollis*, based on an adult male specimen from Wheatfen Broad, England, United Kingdom, sequenced as part of the Darwin Tree of Life Project.

Genome sequence report

The genome of an adult male *Galeruca laticollis* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 74.49 Gb (gigabases) from 7.04 million reads, providing approximately 35-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 135.01 Gb from 894.11 million reads, yielding an approximate coverage of 63-fold. Specimen and sequencing information is summarised in Table 1.

Manual assembly curation corrected 35 missing joins or misjoins, reducing the scaffold number by 30.3%. The final assembly has a total length of 2,154.60 Mb in 45 sequence scaffolds, with 597 gaps, and a scaffold N50 of 231.3 Mb (Table 2). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.92%) of the assembly sequence was assigned to 12 chromosomal-level scaffolds, representing 10 autosomes and the X and Y sex chromosomes.

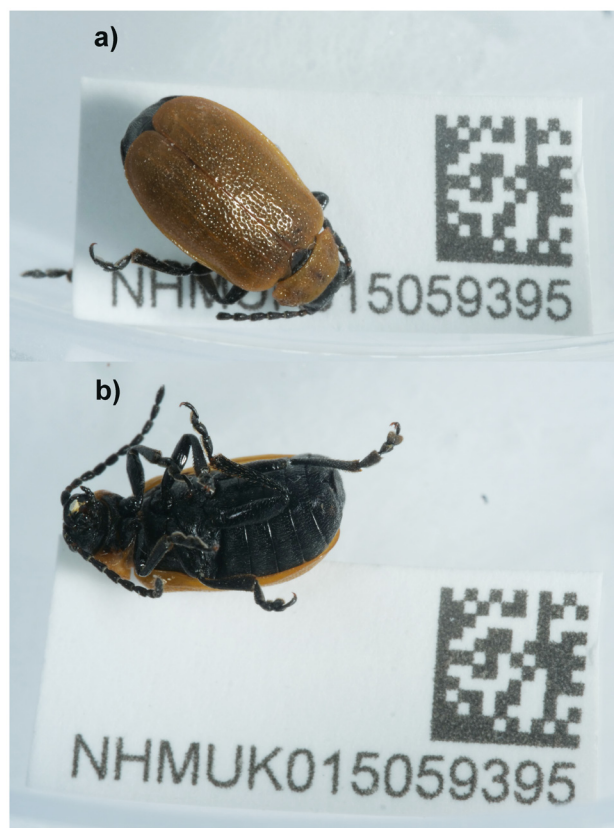


Figure 1. Photographs of the *Galeruca laticollis* (icGallati1) specimen used for genome sequencing. **a)** Dorsal view, **b)** Ventral view.

Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 3). Chromosome X was identified by homology with the *Lochmaea crataegi* assembly (GCA_947563755.1) (Crowley *et al.*, 2024), and chromosome Y was identified based on coverage. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 58.3 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.4.3 completeness of 98.7% (single = 97.9%, duplicated = 0.8%), using the endopterygota_odb10 reference set ($n = 2,124$).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at <https://links.tol.sanger.ac.uk/species/2017130>

Genome annotation report

The *Galeruca laticollis* genome assembly (GCA_963921935.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes

Table 1. Specimen and sequencing data for *Galeruca laticollis*.

Project information			
Study title	<i>Galeruca laticollis</i>		
Umbrella BioProject	PRJEB65735		
BioSample	SAMEA112964282		
NCBI taxonomy ID	2017130		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	icGalLati1	SAMEA112975451	Whole organism
Hi-C sequencing	icGalLati1	SAMEA112975451	Whole organism
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR12035317	8.94e+08	135.01
PacBio Revio	ERR12015776	7.04e+06	74.49

47,793 transcribed mRNAs from 32,229 protein-coding and 830 non-coding genes (Table 2; https://rapid.ensembl.org/Galeruca_laticollis_GCA_963921935.1/Info/Index). The average transcript length is 31,279.23. There are 1.45 coding transcripts per gene and 4.60 exons per transcript.

Methods

Sample acquisition and DNA barcoding

An adult male *Galeruca laticollis* (specimen ID NHMUK015059395, ToLID icGalLati1) was collected from Wheatfen Broad, England, United Kingdom (latitude 52.6, longitude 1.44) on 2022-07-03. The specimen was collected and identified by Roger Booth (Natural History Museum) and preserved by dry freezing at -80°C .

The initial identification was verified by an additional DNA barcoding process according to the framework developed by [Twyford et al. \(2024\)](#). A small sample was dissected from the specimens and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification ([Crowley et al., 2023](#)). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI ([Twyford et al., 2024](#)). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io ([Beasley et al., 2023](#)).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core

Laboratory includes a sequence of core procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io ([Denton et al., 2023b](#)). The icGalLati1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice ([Jay et al., 2023](#)), and tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor ([Denton et al., 2023a](#)).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol ([Oatley et al., 2023](#)). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system ([Bates et al., 2023](#)). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA ([Strickland et al., 2023](#)). 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.

Hi-C preparation

The icGalLati1 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, frozen tissue (stored at -80°C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde. After crosslinking, the tissue was homogenised using the Diagenode Power Masher-II and BioMasher-II tubes and pestles. Following the kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out

Table 2. Genome assembly data for *Galeruca laticollis*, icGalLati1.1.

Genome assembly		
Assembly name	icGalLati1.1	
Assembly accession	GCA_963921935.1	
Accession of alternate haplotype	GCA_963921905.1	
Span (Mb)	2,154.60	
Number of contigs	643	
Contig N50 length (Mb)	6.4	
Number of scaffolds	45	
Scaffold N50 length (Mb)	231.3	
Longest scaffold (Mb)	320.04	
Assembly metrics*		Benchmark
Consensus quality (QV)	58.3	≥ 50
k-mer completeness	99.99%	≥ 95%
BUSCO**	C:98.7%[S:97.9%,D:0.8%], F:0.3%,M:1.0%,n:2,124	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.92%	≥ 95%
Sex chromosomes	XY	localised homologous pairs
Organelles	Mitochondrial genome: 19.98 kb	complete single alleles
Genome annotation of assembly GCA_963921935.1 at Ensembl		
Number of protein-coding genes	32,229	
Number of non-coding genes	830	
Number of gene transcripts	47,793	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from [Rhie et al. \(2021\)](#).

** BUSCO scores based on the endopterygota_odb10 BUSCO set using version 5.4.3. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/Galeruca_laticollis/dataset/GCA_963921935.1/busco.

for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation.

Library preparation and sequencing

Library preparation and DNA sequencing were performed at the WSI Scientific Operations core. Pacific Biosciences HiFi circular consensus DNA sequencing libraries were prepared using the PacBio Express Template Preparation Kit v2.0 (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for removal of single-strand overhangs, DNA damage repair, end repair/A-tailing, adapter ligation, and nuclease treatment. Library

preparation also included a library purification step using 0.8X AMPure PB beads (Pacific Biosciences, California, USA) and size selection step to remove templates < 5 kb using AMPure PB modified SPRI. Samples were sequenced on the Revio system (Pacific Biosciences, California, USA).

For Hi-C library preparation, DNA was fragmented to a size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit following manufacturers' instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

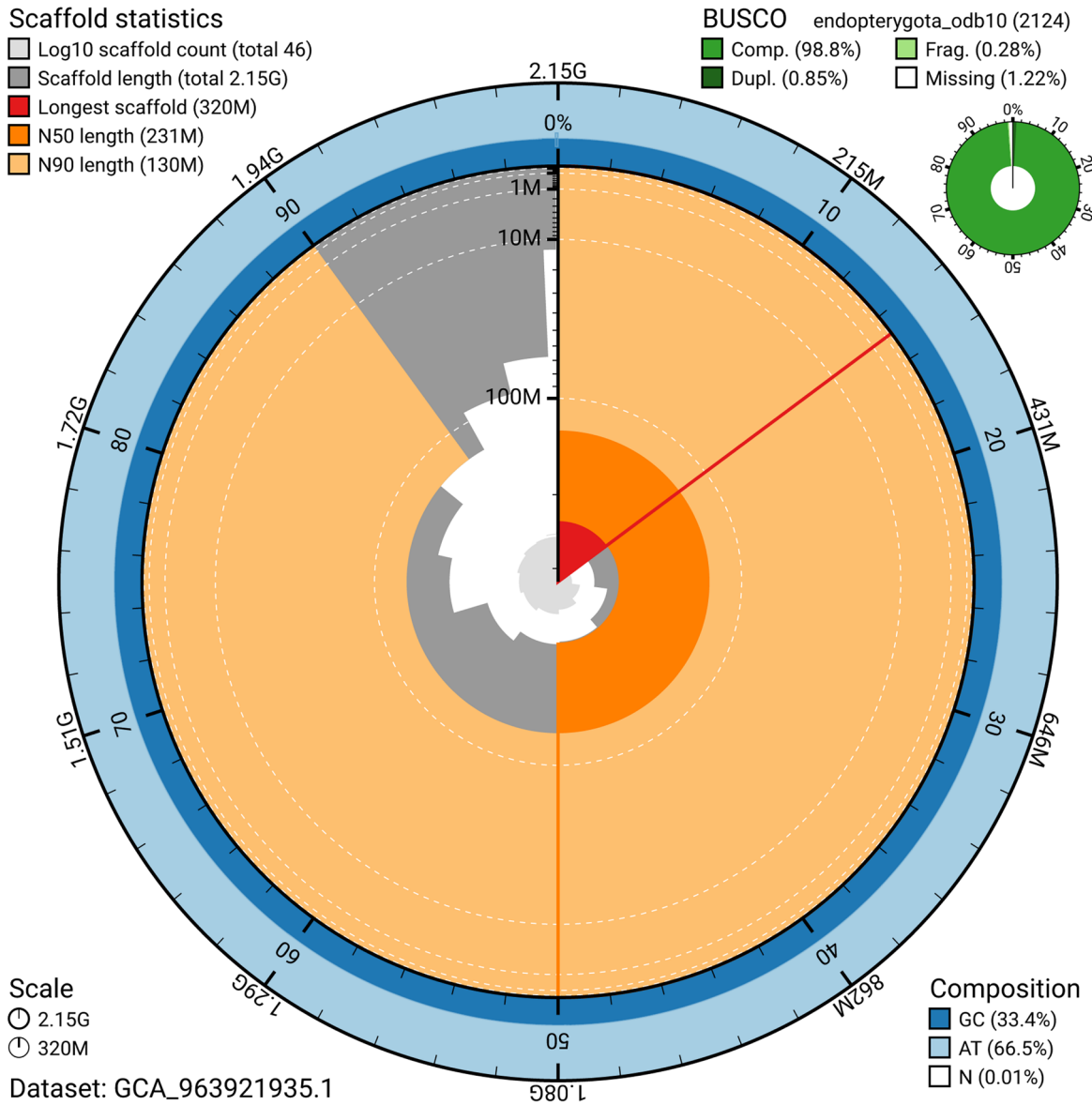


Figure 2. Genome assembly of *Galeruca laticollis*, icGalLati1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 2,154,579,676 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (320,044,732 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (231,279,273 and 129,513,413 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the endopterygota_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963921935.1/dataset/GCA_963921935.1/snail.

Genome assembly, curation and evaluation

Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads were mapped to the primary contigs

using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

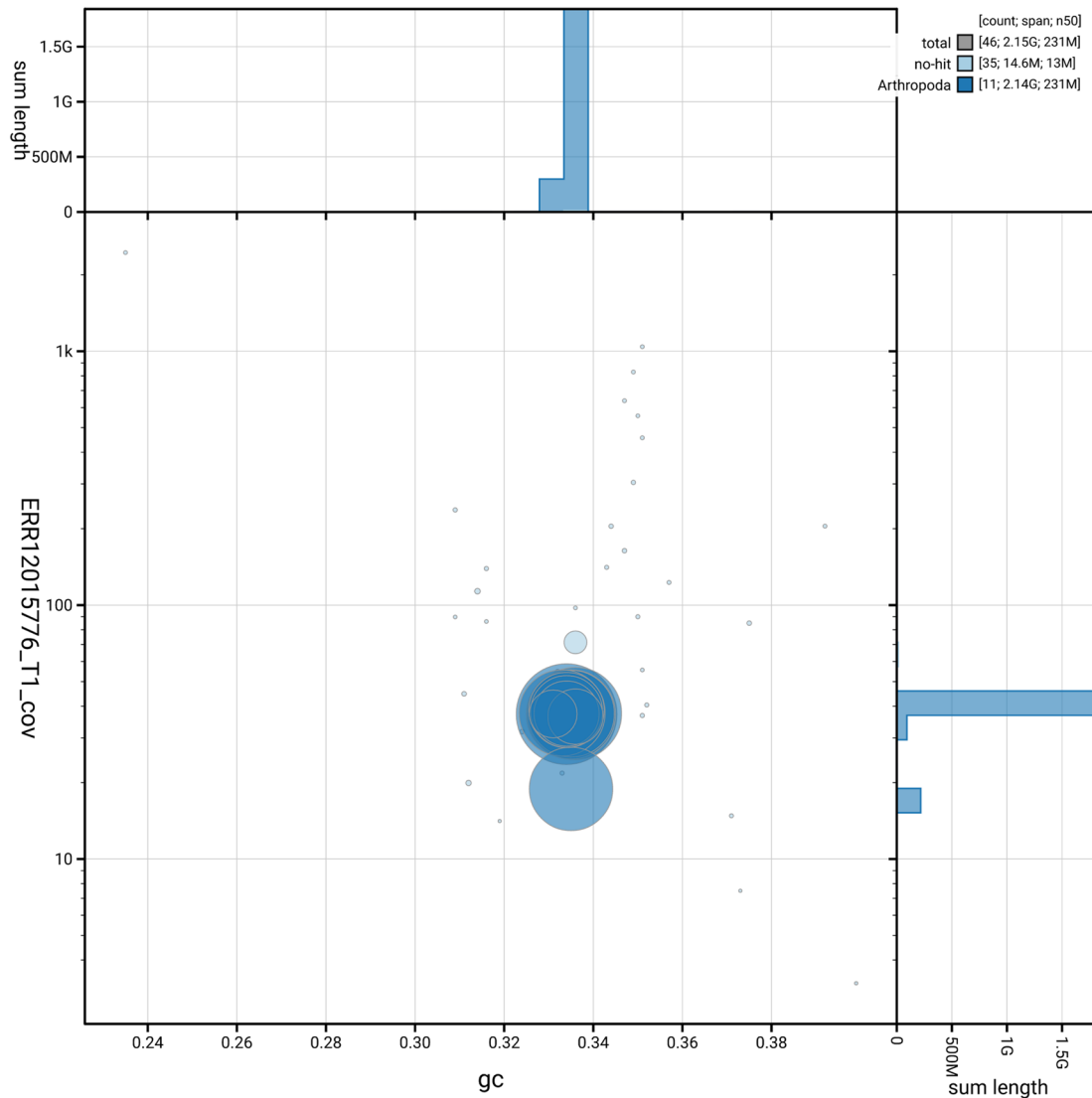


Figure 3. Genome assembly of *Galeruca laticollis*: Blot plot of base coverage in the raw data against GC proportion for sequences in icGalLati1.1. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963921935.1/dataset/GCA_963921935.1/blob.

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023)

and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and misjoins were corrected, and duplicate sequences were tagged and removed. Sex chromosomes were identified using both read coverage and synteny analysis. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation> (article in preparation).

Evaluation of the final assembly

The final assembly was post-processed and evaluated using the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines: sanger-tol/readmapping (Surana *et al.*, 2023a),

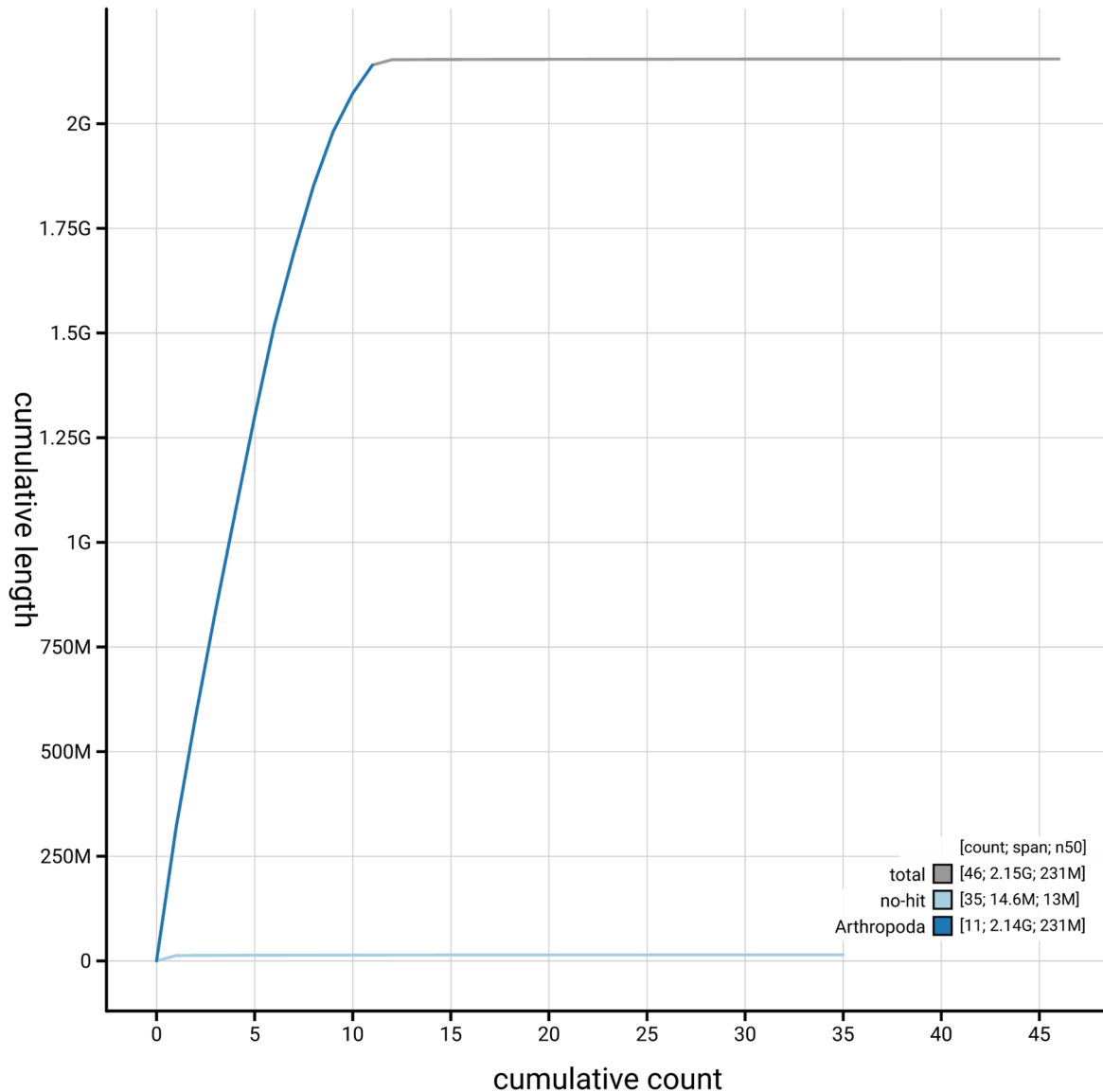


Figure 4. Genome assembly of *Galeruca laticollis* icGallLati1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963921935.1/dataset/GCA_963921935.1/cumulative.

sanger-tol/genomenote (Surana *et al.*, 2023b), and sanger-tol/blobtoolkit (Muffato *et al.*, 2024). The readmapping pipeline aligns the Hi-C reads using bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The genomenote pipeline converts the Hi-C alignments into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018). This pipeline also generates assembly statistics using the NCBI datasets report (Sayers *et al.*, 2024), computes *k*-mer completeness and QV consensus quality values with

FastK and MERQURY.FK, and runs BUSCO (Manni *et al.*, 2021) to assess completeness.

The blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes

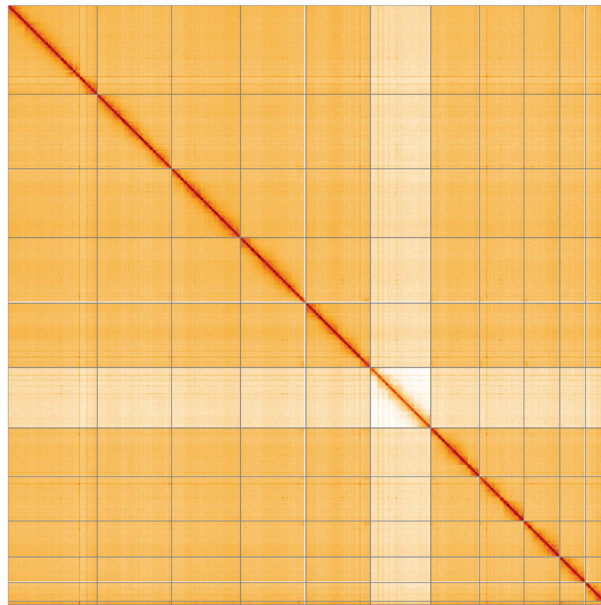


Figure 5. Genome assembly of *Galeruca laticollis* icGalLati1.1: Hi-C contact map of the icGalLati1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=WQfcah4sQICI-kvqtLjt0A>.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Galeruca laticollis*, icGalLati1.

INSDC accession	Name	Length (Mb)	GC%
OY998175.1	1	320.04	33.5
OY998176.1	2	266.52	33.5
OY998177.1	3	248.29	33.5
OY998178.1	4	234.71	33.5
OY998179.1	5	231.28	33.5
OY998181.1	6	174.9	33.5
OY998182.1	7	159.02	33.5
OY998183.1	8	129.51	33.5
OY998184.1	9	91.58	33.5
OY998185.1	10	67.33	33.0
OY998180.1	X	216.78	33.5
OY998186.1	Y	12.97	33.5
OY998187.1	MT	0.02	23.5

database (Bateman *et al.*, 2023) with DIAMOND (Buchfink *et al.*, 2021) blastp. The genome is also split into chunks according to the density of the BUSCO genes from the

closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Genome sequences without a hit are chunked with seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these outputs into a blobdir for visualisation.

The genome assembly and evaluation pipelines were developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), relying on the Conda package manager, the Bioconda initiative (Grünig *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

Genome annotation

The Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Galeruca laticollis* assembly (GCA_963921935.1) in Ensembl Rapid Release at the EBI. Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

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

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/ark5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/
BlobToolKit	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Goat CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
Mercury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/genomenote	1.1.1	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.2.1	https://github.com/sanger-tol/readmapping
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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 [here](https://www.darwintreeoflife.org/). 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: *Galeruca laticollis*. Accession number PRJEB65735; <https://identifiers.org/ena.embl/PRJEB65735> (Wellcome Sanger Institute, 2024). The genome

sequence is released openly for reuse. The *Galeruca laticollis* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in [Table 1](#) and [Table 2](#).

Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12159242>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.4783558>.

References

- Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays**. *Bioinformatics*. 2020; **36**(1): 311–316. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aken BL, Ayling S, Barrell D, et al.: **The ensembl gene annotation system**. *Database (Oxford)*. 2016; **2016**: baw093. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Allio R, Schomaker-Bastos A, Romiguiet J, et al.: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics**. *Mol Ecol Resour*. 2020; **20**(4): 892–905. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altschul SF, Gish W, Miller W, et al.: **Basic local alignment search tool**. *J Mol Biol*. 1990; **215**(3): 403–410. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, et al.: **UniProt: the universal protein knowledgebase in 2023**. *Nucleic Acids Res*. 2023; **51**(D1): D523–D531. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor#3 for LI PacBio**. *protocols.io*. 2023. [Publisher Full Text](#)
- Beasley J, Uhl R, Forrest LL, et al.: **DNA barcoding SOPs for the Darwin Tree of Life project**. *protocols.io*. 2023. [Accessed 25 June 2024]. [Publisher Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at tree-of-life scale using DIAMOND**. *Nat Methods* 2021; **18**(4): 366–368. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Kumar S, Sotero-Caio C, et al.: **Genomes on a Tree (GoAT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2023; **8**: 24. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, et al.: **BlobToolKit – interactive quality assessment of genome assemblies**. *G3 (Bethesda)*. 2020; **10**(4): 1361–1374. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, et al.: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm**. *Nat Methods*. 2021; **18**(2): 170–175. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crowley L, Allen H, Barnes I, et al.: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2023; **8**: 123. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crowley LM, Telfer MG, Escalona HE, et al.: **The genome sequence of the hawthorn leaf beetle, *Lochmaea crataegi* (Forster, 1771) [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2024; **9**: 80. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- da Veiga Leprevost F, Grüning BA, Alves Aflitos S, et al.: **BioContainers: an open-source and community-driven framework for software standardization**. *Bioinformatics*. 2017; **33**(16): 2580–2582. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danecek P, Bonfield JK, Liddle J, et al.: **Twelve years of SAMtools and BCFtools**. *GigaScience*. 2021; **10**(2): gjab008. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Denton A, Oatley G, Cornwell C, et al.: **Sanger Tree of Life sample homogenisation: PowerMash**. *protocols.io*. 2023a. [Publisher Full Text](#)
- Denton A, Yatsenko H, Jay J, et al.: **Sanger Tree of Life wet laboratory protocol**

collection V.1. *protocols.io*. 2023b.

[Publisher Full Text](#)

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows**. *Nat Biotechnol*. 2017; **35**(4): 316–319.

[PubMed Abstract](#) | [Publisher Full Text](#)

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation**. *Genome Biol*. 2023; **24**(1): 74.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Duff AG: **Beetles of Britain and Ireland. Cerambycidae to curculionidae**. West Rinton: A.G. Duff publishing, 2016; 4.

[Reference Source](#)

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report**. *Bioinformatics*. 2016; **32**(19): 3047–3048.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines**. *Nat Biotechnol*. 2020; **38**(3): 276–278.

[PubMed Abstract](#) | [Publisher Full Text](#)

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfstats: conversion, evaluation and manipulation of genome sequences using assembly graphs**. *Bioinformatics*. 2022; **38**(17): 4214–4216.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences**. *Nat Methods*. 2018; **15**(7): 475–476.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies**. *Bioinformatics*. 2020; **36**(9): 2896–2898.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Harry E: **PretextView (Paired Read Texture Viewer): a desktop application for viewing pretext contact maps**. 2022.

[Reference Source](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation**. *GigaScience*. 2021; **10**(1): g1aa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hyman PS, Parsons MS: **A review of the scarce and threatened Coleoptera of great Britain: Part 1**. Peterborough: Joint Nature Conservation Committee, 1992.

[Reference Source](#)

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection**. *protocols.io*. 2023.

[Publisher Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Mol Biol Evol*. 2021; **38**(10): 4647–4654.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Merkel D: **Docker: lightweight Linux containers for consistent development**

and deployment. *Linux J*. 2014; **2014**(239): 2. [Accessed 2 April 2024].

[Reference Source](#)

Muffato M, Butt Z, Challis R, *et al.*: **sanger-tol/blobtoolkit: v0.3.0 – poliwig**. 2024.

[Publisher Full Text](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2**. *protocols.io*. 2023.

[Publisher Full Text](#)

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – ancient Atlantis**. 2023.

[Publisher Full Text](#)

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics*. 2010; **26**(6): 841–842.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rhie A, Walenz BP, Koren S, *et al.*: **Mercury: reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biol*. 2020; **21**(1): 245.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sayers EW, Cavanaugh M, Clark K, *et al.*: **GenBank 2024 update**. *Nucleic Acids Res*. 2024; **52**(D1): D134–D137.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI**. *protocols.io*. 2023.

[Publisher Full Text](#)

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0)**. *Zenodo*. 2023a.

[Publisher Full Text](#)

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev)**. *Zenodo*. 2023b.

[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: awaiting peer review]**. *Wellcome Open Res*. 2024; **9**: 339.

[Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads**. *BMC Bioinformatics*. 2023; **24**(1): 288.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge**. *Nucleic Acids Res*. 2019; **47**(D1): D506–D515.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Vasimuddin Md, 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](#)

Wellcome Sanger Institute: **The genome sequence of a leaf beetle, *Galeruca laticollis* Sahlberg, C.R., 1838**. European Nucleotide Archive. [dataset], accession number PRJEB65735, 2023.

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 20 November 2024

<https://doi.org/10.21956/wellcomeopenres.25545.r109784>

© 2024 Porras J. 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.



Jose Manuel Rico Porras 

Universidad de Jaen, Andalusia, Spain

The article accurately describes the steps used for the assembly of the genome of the leaf beetle *Galeruca laticollis*. Some points to consider:

You can include the family and order in the title: The genome sequence of a leaf beetle, *Galeruca laticollis* Sahlberg, C.R., 1838 (Coleoptera, Chrysomelidae)

You can mention in the Background section that there are no cytogenetic data available for this species (Coleoptera Karyotype Database¹), but this data are available for some species of this genus²:

Galeruca angusta: 2n=44M; 21+Xyp

Galeruca artemisiae: 2n=44M; 21+Xyp

Galeruca canigouensis: 2n=38-40M

Galeruca indica: 2n=46M; 22+Xyp

Galeruca miegi: 2n=38M; 18+Xyp

Galeruca monticola: 2n=38M; 18+Xyp

Galeruca taceti: 2n=40M; 19+Xyp

In Table 2, Number of contigs is 643, but in the Genome assembly GCA_963921935.1 NCBI web is 642.

References

1. Blackmon H, Demuth J: Coleoptera Karyotype Database. The Coleopterists Bulletin. [Publisher Full Text](#)
2. Petitpierre, E., Segarra, C., Yadav, J.S., Virkki, N, In: Jolivet, P., Petitpierre, E., Hsiao, T.H.(eds) Biology of Chrysomelidae: Chromosome Numbers and Meioformulae of Chrysomelidae. [Publisher](#)

[Full 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: Cytogenomics, repetitive DNA

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 20 November 2024

<https://doi.org/10.21956/wellcomeopenres.25545.r109779>

© 2024 Tomanović Ž. 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.



Željko Tomanović

Serbia University of Belgrade, Belgrade, Serbia

The authors presented the sequenced genome of the chrysomelid beetle *Galeruca laticollis* using a male specimen collected in Wheatfen Broad, England, United Kingdom. The authors gave a brief overview of the biology of *G. laticollis*. All relevant quantitative parameters of the assembled genome are listed and the methodology presented in detail. The following is just a suggestion.

"Outside the UK, it occurs in most of Europe and extends eastwards into temperate parts of Asia." Please add appropriate references on the distribution of the species.

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: phylogenetics and biodiversity

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.
