Check for updates

DATA NOTE

The genome sequence of a leaf beetle, *Chrysomela saliceti* (Weise, 1884) (Coleoptera: Chrysomelidae)

[version 1; peer review: 3 approved, 1 approved with reservations]

Michael F. Geiser[1], Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

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

v1 **First published:** 22 Dec 2025, **10**:688
https://doi.org/10.12688/wellcomeopenres.25214.1
Latest published: 22 Dec 2025, **10**:688
https://doi.org/10.12688/wellcomeopenres.25214.1

Abstract

We present a genome assembly from an individual male *Chrysomela saliceti* (leaf beetle; Arthropoda; Insecta; Coleoptera; Chrysomelidae). The assembly contains two haplotypes with total lengths of 738.63 megabases and 703.32 megabases. Most of haplotype 1 (98.68%) is scaffolded into 17 chromosomal pseudomolecules, including the X sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 21.9 kilobases. Gene annotation of this assembly on Ensembl identified 14 722 protein-coding genes. 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

Chrysomela saliceti; leaf beetle; genome sequence; chromosomal; Coleoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ? ✓

	1	2	3	4
version 1 22 Dec 2025	✓ view	✓ view	? view	✓ view

1. **Nikoletta Andrea Nagy**, University of Debrecen, Debrecen, Hungary
2. **Wai Lok So**, The Chinese University of Hong Kong, Hong Kong, Hong Kong
The Hang Seng University of Hong Kong (Ringgold ID: 272025), Hong Kong, Hong Kong
3. **Yu-Feng Huang**, Yuan Ze University (Ringgold ID: 34895), Zhongli District, Taiwan
4. **Somjit Homchan**, Naresuan University, Tha Pho, Thailand

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: Geiser MF: Investigation, Resources;

Competing interests: No competing interests were disclosed.

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

Copyright: © 2025 Geiser MF *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: Geiser MF, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of a leaf beetle, *Chrysomela saliceti* (Weise, 1884) (Coleoptera: Chrysomelidae) [version 1; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10**:688 https://doi.org/10.12688/wellcomeopenres.25214.1

First published: 22 Dec 2025, **10**:688 https://doi.org/10.12688/wellcomeopenres.25214.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Chrysomeloidea; Chrysomelidae; Chrysomelinae; Chrysomelini; *Chrysomela*; *Chrysomela saliceti* (Weise, 1884) (NCBI: txid1587176)

Background

We present a chromosome-level genome sequence for *Chrysomela saliceti*. The assembly was produced using the Tree of Life pipeline from a specimen collected from the WWT London Wetland Centre, Barnes, London, England, UK (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Darwin Tree of Life Project Consortium, 2022).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Chrysomela saliceti* (specimen ID NHMUK014560511, ToLID icChrSali1; Figure 1), collected from WWT London Wetland Centre, Barnes, London, England, UK (latitude 51.48, longitude –0.24) on 2022-05-02. The specimen was collected and identified by Michael Geiser (Natural History Museum). Details of the sampling and metadata procedures, which followed recommended standards, are described in Lawniczak *et al.* (2022).

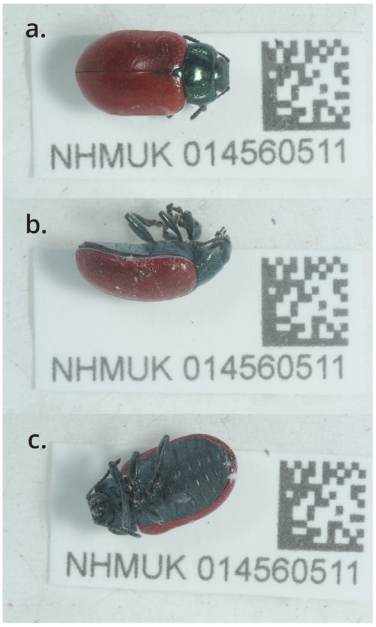

Figure 1. Photograph of the *Chrysomela saliceti* (icChrSali1) specimen used for genome sequencing.

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 icChrSali1 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. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 31.9 ng/μL and a yield of 1 499.30 ng, with a fragment size of 16.1 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 icChrSali1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools were created. 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 X.

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

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 49 breaks and 128 joins. These changes reduced the scaffold count by 28.1% and increased the scaffold N50 by 17.7%. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for 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 *Chrysomela saliceti* specimen generated 19.08 Gb (gigabases) from 1.72 million reads, which

were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 726.46 Mb, with a heterozygosity of 0.89% and repeat content of 49.89% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 26× coverage. Hi-C sequencing produced 115.88 Gb from 767.45 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

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

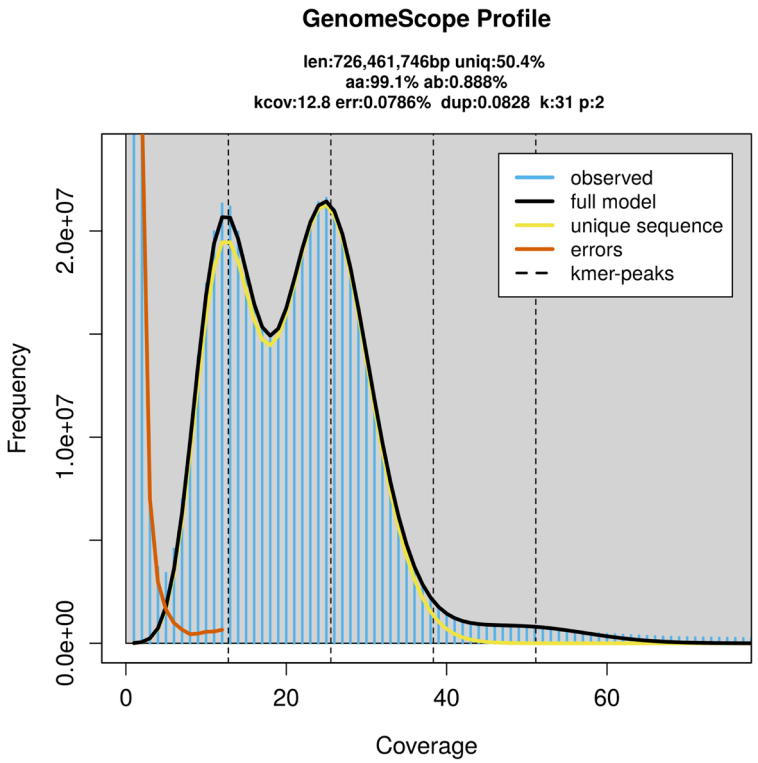

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB79783.

Platform	PacBio HiFi	Hi-C
ToLID	icChrSali1	icChrSali1
Specimen ID	NHMUK014560511	NHMUK014560511
BioSample (source individual)	SAMEA114806286	SAMEA114806286
BioSample (tissue)	SAMEA114806422	SAMEA114806422
Tissue	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13650041	ERR13654251
Read count total	1.72 million	767.45 million
Base count total	19.08 Gb	115.88 Gb

has a total length of 738.63 Mb in 122 scaffolds, with 587 gaps, and a scaffold N50 of 45.39 Mb (Table 2).

Most of the assembly sequence (98.68%) was assigned to 17 chromosomal-level scaffolds, representing 16 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Chromosome X was identified by BUSCO gene painting with ancestral elements and by copy number in the diploid assembly. No Chromosome Y could be identified.

Table 2. Genome assembly statistics.

Assembly name	icChrSali1.hap1.1	icChrSali1.hap2.1
Assembly accession	GCA_965121785.1	GCA_965122525.1
Assembly level	chromosome	scaffold
Span (Mb)	738.63	703.32
Number of chromosomes	17	scaffold-level
Number of contigs	709	798
Contig N50	2.15 Mb	2.04 Mb
Number of scaffolds	122	242
Scaffold N50	45.39 Mb	43.77 Mb
Longest scaffold length (Mb)	66.92	-
Sex chromosomes	X	-
Organelles	Mitochondrion: 21.9 kb	-

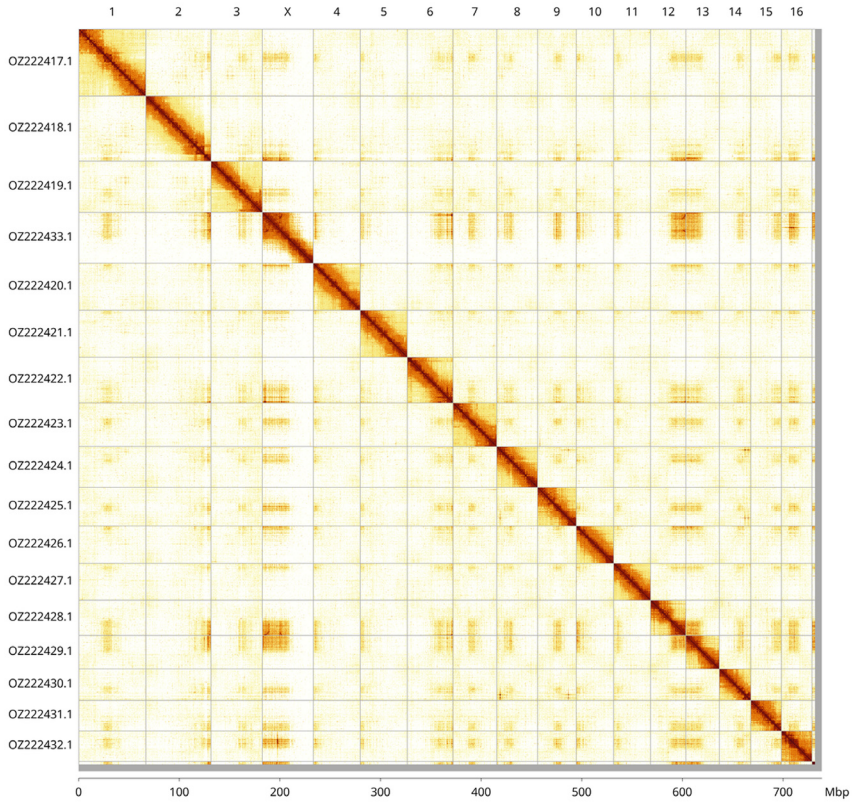

Figure 3. Hi-C contact map of the *Chrysomela saliceti* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Chrysomela saliceti* icChrSali1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ222417.1	1	66.92	35.50
OZ222418.1	2	64.84	35.50
OZ222419.1	3	50.83	35.50
OZ222420.1	4	46.81	35.50
OZ222421.1	5	46.69	35.50
OZ222422.1	6	45.39	36
OZ222423.1	7	43.49	35.50
OZ222424.1	8	40.56	35.50
OZ222425.1	9	38.56	35.50
OZ222426.1	10	37.08	35.50
OZ222427.1	11	36.81	35.50
OZ222428.1	12	34.91	36
OZ222429.1	13	33.28	36
OZ222430.1	14	31.33	36
OZ222431.1	15	30.44	36
OZ222432.1	16	30.37	36.50
OZ222433.1	X	50.58	36.50

The mitochondrial genome was also assembled (length 21.9 kb, OZ222434.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 60.6, and for haplotype 2, 60.3. When the two haplotypes are combined, the assembly achieves an estimated QV of 60.5. The *k*-mer completeness is 82.53% for haplotype 1, 78.73% for haplotype 2, and 99.01% for the combined haplotypes (Figure 4).

BUSCO analysis using the endopterygota_odb10 reference set (n = 2 124) identified 99.0% of the expected gene set (single = 98.2%, duplicated = 0.8%) 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.C.Q60**, meeting the recommended reference standard.

Genome annotation report

The *Chrysomela saliceti* genome assembly (GCA_965121785.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI), using the Ensembl non-vertebrate genome annotation method. This annotation includes 27 910 transcribed mRNAs from 14 722 protein-coding and 4 217 non-coding

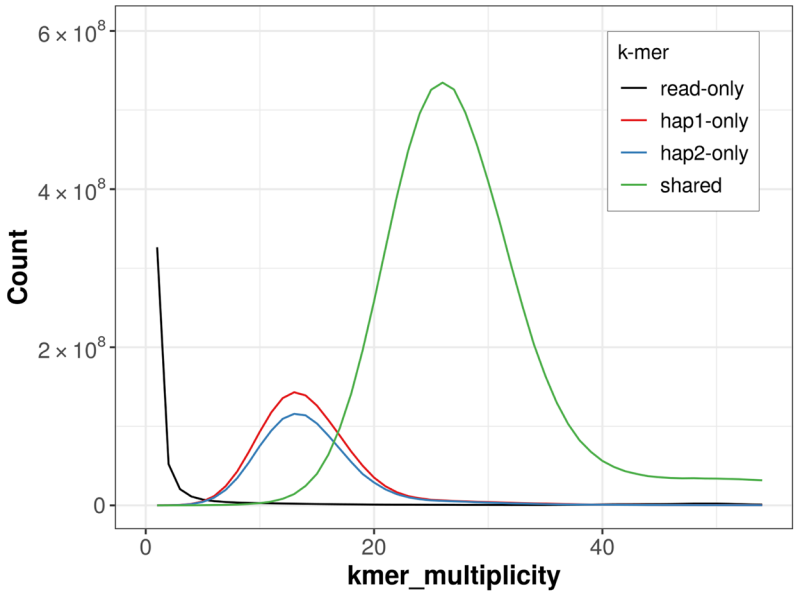

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

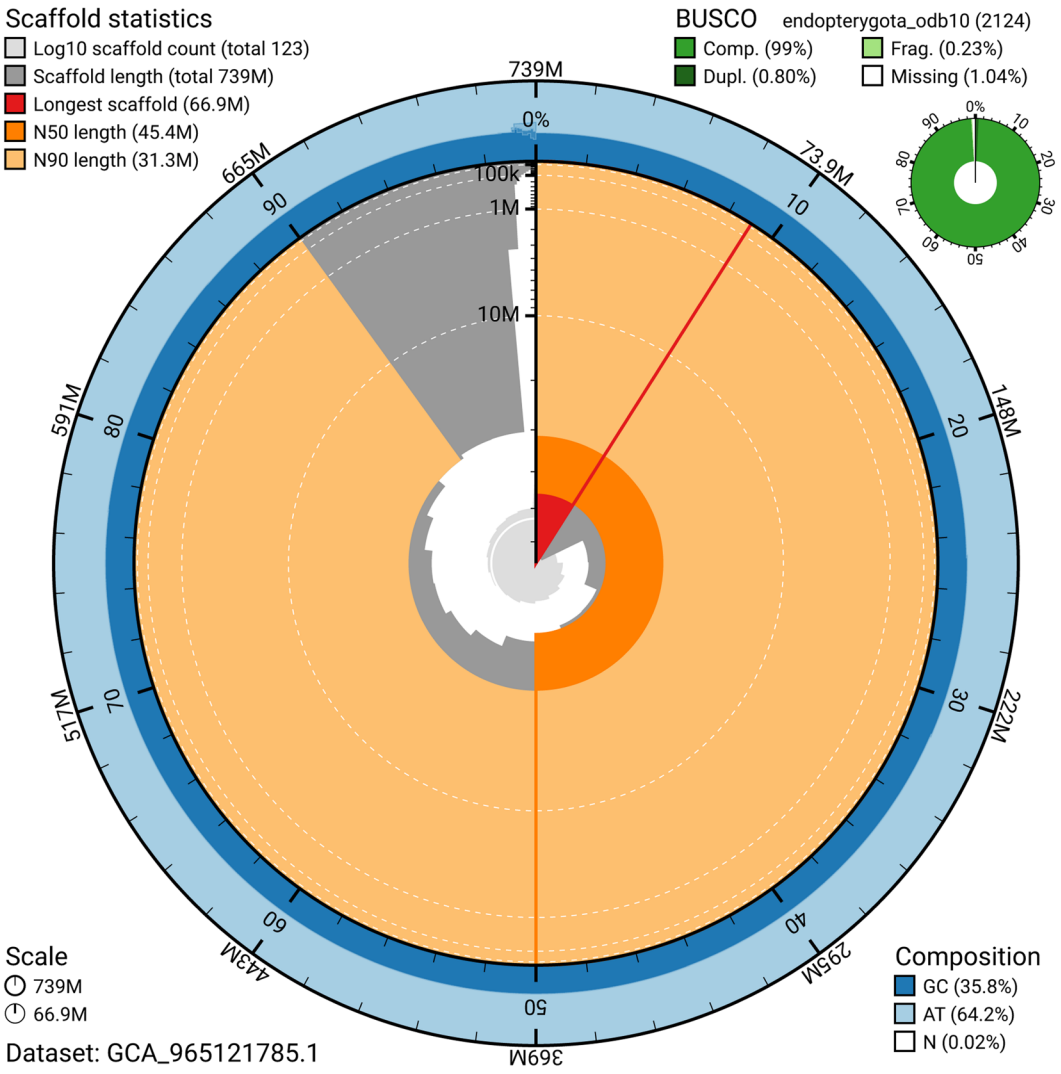

Figure 5. Assembly metrics for icChrSali1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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.

genes. The average transcript length is 15 492.29 bp, with an average of 1.47 coding transcripts per gene and 4.84 exons per transcript. For further information about the annotation, please refer to the annotation page on Ensembl.

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

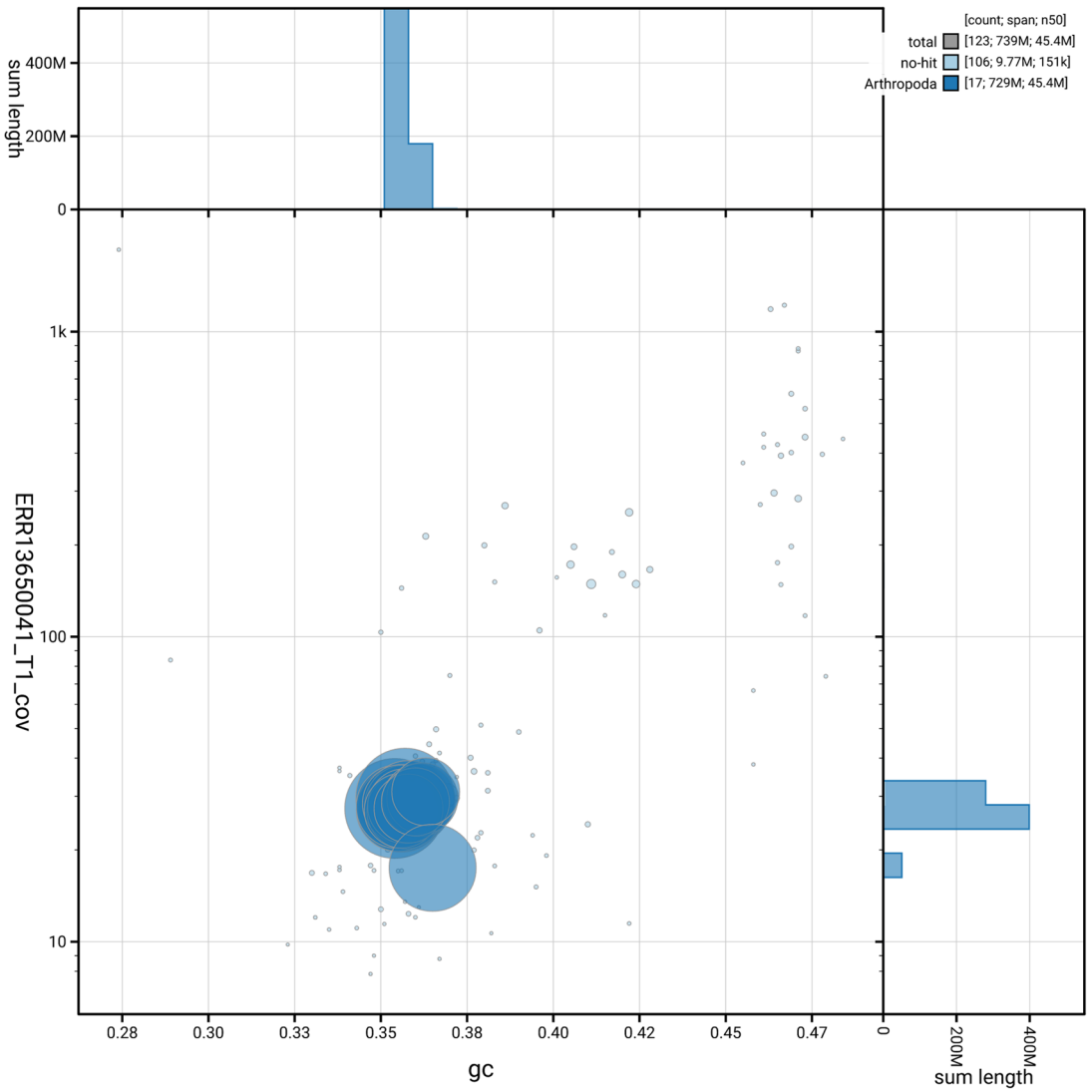

Figure 6. BlobToolKit GC-coverage plot for icChrSali1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Chrysomela saliceti* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q60	6.C.Q40
Contig N50 length	2.15 Mb	≥ 1 Mb
Scaffold N50 length	45.39 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 60.6; haplotype 2: 60.3; combined: 60.5	≥ 40
k-mer completeness	Haplotype 1: 82.53%; Haplotype 2: 78.73%; combined: 99.01%	≥ 95%
BUSCO	C:99.0% [S:98.2%; D:0.8%]; F:0.2%; M:0.8%; n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.68%	≥ 90%

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: Chrysomela saliceti. Accession number PRJEB79783. The genome sequence is released openly for reuse. The *Chrysomela saliceti* 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. 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 Natural History Museum Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
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	0.2.5	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

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

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

Darwin Tree of Life Project Consortium: **Sequence locally, think globally: the Darwin Tree of Life project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

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**

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 16 February 2026

https://doi.org/10.21956/wellcomeopenres.27790.r145670

© **2026 Homchan S.** 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.

✔ **Somjit Homchan**
Naresuan University, Tha Pho, Thailand

Overall, this manuscript is well-written and presents a high-quality genome assembly. I have only a few suggestions and questions below.
1) The background is too brief. Please provide more details on this species and its importance. Additionally, this section reads like an abstract that incorporates methods, leading to redundancy with later sections.
2) The authors should address why this species' complete mitochondrial genome is unusually large compared to other beetles in the same family.
3) Please explain while K-mer completeness of haplotype 1(82.53%) and haplotype 2(78.73%), when the combined both haplotypes it shows K-mer completeness 99.01%.

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: Genetic markers, Population Genetics, Genomics

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 09 February 2026

https://doi.org/10.21956/wellcomeopenres.27790.r145662

© 2026 Huang Y. 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.

?

Yu-Feng Huang
Department of Computer Science and Engineering, Yuan Ze University (Ringgold ID: 34895), Zhongli District, Taoyuan City, Taiwan

This Data Note reports a high-quality chromosome-level assembly for a male *Chrysomela saliceti* generated using PacBio HiFi with Illumina Hi-C for scaffolding and phasing. The assembly is supported by strong algorithmic QC, including a k-mer–based QV (~60), high BUSCO completeness (99.0% for haplotype 1), Hi-C contact maps consistent with chromosome-scale scaffolds, and documented contamination screening and manual curation. Overall, the assembly meets (and exceeds) the stated Earth BioGenome Project benchmark summary (6.C.Q60).
However, although the wet/dry workflows follow established DToL practices and the headline metrics are strong, several points require clarification to avoid misinterpretation - particularly regarding how phased haplotypes relate to completeness metrics and the structural validity of the mitochondrial genome. Specifically, the manuscript should (1) explain why haplotype-level k-mer completeness (~79–83%) is lower than the combined completeness (99%) in a phased assembly and clearly state which metric is used for EBP-style reporting; (2) briefly rationalize the small difference between the GenomeScope-estimated haploid genome size (~726 Mb) and the haplotype 1 assembly span (~739 Mb); and (3) explicitly confirm whether the 21.9 kb mitochondrial genome is circularized/complete and provide a brief explanation for its relatively large size.
Adding a minimal annotation QC indicator and a short species-background paragraph would further strengthen the note.
Major Comments

1. **Clarify k-mer completeness interpretation for phased assemblies**
   The manuscript reports haplotype-specific k-mer completeness of 82.53% (haplotype 1) and 78.73% (haplotype 2), while the combined haplotypes reach 99.01%. This pattern is generally expected for phased diploid assemblies (each haplotype captures only a subset of the total k-mer set), but it may confuse readers - especially since Table 4 juxtaposes “haplotype 1 summary” metrics with a combined-haplotype completeness benchmark. Please add a short explanation and explicitly indicate which completeness value is intended to support completeness claims and EBP-style reporting.
2. **Assembly span versus estimated genome size: briefly rationalize differences**
   GenomeScope2 estimates a haploid genome size of 726.46 Mb, whereas haplotype 1 spans 738.63 Mb. The ~12 Mb difference may be benign (e.g., unresolved repeats, small retained alternate segments, or minor residual duplication), but readers often interpret such discrepancies as evidence of duplication or incomplete haplotig purging. A brief statement clarifying plausible causes (and whether any residual duplication was assessed) would improve interpretability.
3. **Mitochondrial genome: confirm circularity/completeness and address unusually large size**

The mitochondrial genome is reported as 21.9 kb, which is relatively large for many insects and raises the possibility of an expanded control region, duplicated segments, or a linear/incomplete assembly. Please explicitly state whether circularity was verified (e.g., by identifying overlapping ends and/or standardizing the circular start position). If available, a minimal validation (gene content presence or a brief note on the control region) would help alleviate concerns—particularly since the current text does not provide mitogenome annotation details.

4. **Gene annotation: add a minimal quality indicator beyond gene counts**
The note reports 14,722 protein-coding genes, but does not provide a concise annotation QC indicator (even a single value). To strengthen claims of gene-level completeness/correctness, please include a minimal metric such as BUSCO on predicted proteins/transcripts or a brief statement about the evidence sources used in the annotation (e.g., homology and/or transcript support).

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

Are the protocols appropriate and is the work technically sound?
Partly

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: Bioinformatics, Cancer Biology, Mitogenomics, Genome assembly and annotation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 06 February 2026

https://doi.org/10.21956/wellcomeopenres.27790.r145663

© 2026 So W. 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.

Wai Lok So
[1] The Chinese University of Hong Kong, Hong Kong, Hong Kong
[2]

The Hang Seng University of Hong Kong (Ringgold ID: 272025), Hong Kong, Hong Kong

This manuscript presents a high-quality genome assembly of *Chrysomela saliceti*. The sequencing and assembly methodologies are technically sound and logical. A well assembled haplotype 1, including the X sex chromosome was yield, together with a complete mitochondrial genome.The methods and processing pipelines are appropriate and transparent, which demonstrates replicability. The resources generated would be useful for research on Coleoptera and insect evolution, and related genomic studies. I would like to request for a scale bar on Figure 1, otherwise I have no major concerns of the manuscript.

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: Evolution, Genomics, Insects, Myriapods

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

Reviewer Report 27 January 2026

https://doi.org/10.21956/wellcomeopenres.27790.r144143

© 2026 Nagy N. 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.

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 leaf beetles. I have only a few suggestions for the authors.
Background
The background seems to be too short for me. I would recommend providing an additional paragraph which briefly describes the 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.

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.