




DATA NOTE

The genome sequence of the barkfly, *Loensia variegata* (Latreille, 1799) (Psocodea: Psocidae)

[version 1; peer review: 1 approved]

James McCulloch¹, Liam M. Crowley ²,
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

¹University of Oxford, Oxford, England, UK²Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 16 Apr 2026, 11:224
<https://doi.org/10.12688/wellcomeopenres.26324.1>
Latest published: 16 Apr 2026, 11:224
<https://doi.org/10.12688/wellcomeopenres.26324.1>

Abstract

We present a genome assembly from an individual male *Loensia variegata* (barkfly; Arthropoda; Insecta; Psocodea; Psocidae). The assembly contains two haplotypes with total lengths of 531.42 megabases and 488.90 megabases. Most of haplotype 1 (97.63%) is scaffolded into 9 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 15.94 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

Loensia variegata, barkfly, genome sequence, chromosomal, Psocodea



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

Open Peer Review

Approval Status

1

version 1

16 Apr 2026

[view](#)

1. **Mireille Steck** , University of Hawai'i at Mānoa, Honolulu, USA

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: **McCulloch J:** Investigation, Resources; **Crowley LM:** Investigation, Resources;

Competing interests: No competing interests were disclosed.

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

Copyright: © 2026 McCulloch J *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: McCulloch J, Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the barkfly, *Loensia variegata* (Latreille, 1799) (Psocodea: Psocidae) [version 1; peer review: 1 approved]** Wellcome Open Research 2026, 11:224 <https://doi.org/10.12688/wellcomeopenres.26324.1>

First published: 16 Apr 2026, 11:224 <https://doi.org/10.12688/wellcomeopenres.26324.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Paraneoptera; Psocodea; Psocomorpha; Psocetae; Psocidae; *Loensia*; *Loensia variegata* (Latreille, 1799) (NCBI:txid209941).

Background

Barkflies (order Psocoptera, family Psocidae) are small, soft-bodied insects that live outdoors on the bark and branches of trees, where they feed on algae, lichens, fungi, and other surface organic matter ([National Barkfly Recording Scheme, 2025a](#)). The name “barkfly” was coined in 2003 to replace the older term “barklice”, distinguishing the arboreal outdoor species from the closely related booklice found indoors ([National Barkfly Recording Scheme, 2025a](#)). They range in size from 1.5 to 7 mm and typically hold their wings tent-wise over the abdomen, giving them a superficial resemblance to miniature lacewings ([National Barkfly Recording Scheme, 2025a](#)).

Loensia variegata (Latreille, 1799) is a moderately large barkfly (4–5 mm), belonging to the ‘picture-winged’ group ([National Barkfly Recording Scheme, 2025b](#)). The forewings are heavily speckled, giving the insect a mottled appearance, typically brown or yellowish against white ([National Barkfly Recording Scheme, 2025b](#)). The sexes can be distinguished by close examination of the abdomen tip: females have a subgenital plate bearing two large oval side lobes with short dark lines, while males have a short rounded appendage on the left-hand side of the hypandrium ([National Barkfly Recording Scheme, 2025b](#)). Reliable identification to species level requires examination of these genital structures, as the species is easily confused with the closely related *L. pearmani* ([NatureSpot, 2025](#)).

The species is Western Palaearctic in distribution, occurring across much of Europe from the British Isles and Scandinavia to Iberia and beyond ([GBIF Secretariat, 2025](#)). In Britain and Ireland it is considered uncommon ([National Barkfly Recording Scheme, 2025c](#)). Adults have been recorded from June to January in the south of England, and from July to September in the north, with the main activity period running through July to September ([National Barkfly Recording Scheme, 2025c](#); [NatureSpot, 2025](#)).

Loensia variegata is found on the trunks and branches of a range of deciduous and evergreen trees and shrubs ([National Barkfly Recording Scheme, 2025c](#)). Recorded host plants include ash, beech, birch, hawthorn, holly, oak, apple, pear, plum, and yew ([National Barkfly Recording Scheme, 2025c](#)). It is most readily recorded by beating branches over a tray or brushing tree trunks with a soft brush ([National Barkfly Recording Scheme, 2025a](#)).

We present a chromosome-level genome sequence for *Loensia variegata*, generated using the Tree of Life pipeline from a specimen collected from Wytham Woods, Oxfordshire, 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](#)).



Figure 1. Photograph of the *Loensia variegata* (iuLoeVari1) specimen used for genome sequencing.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Loensia variegata* (specimen ID OX002633, ToLID iuLoeVari1; [Figure 1](#)), collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338) on 2022-08-03. The specimen was collected by James McCulloch and Liam Crowley and identified by James McCulloch (University of Oxford). Sample metadata were collected in line with the Darwin Tree of Life project standards described by [Lawniczak et al. \(2022\)](#).

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

Detailed protocols for nucleic acid extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on [protocols.io](#) ([Howard et al., 2025](#)). The iuLoeVari1 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.

High molecular weight (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. 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 [manual SPRI](#) (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 3.06 ng/μL and a yield of 137.70 ng, with a fragment size of 12.6 kb. The Genomic Quality Number (GQN) was 6.5.

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 iuLoeVari1 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 Diagenode Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of k -mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the k -mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 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 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 Cointons 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 174 breaks, 337 joins, and removal of 263 haplotypic duplications. This reduced the scaffold count by 55.4%, increased the scaffold N50 by 5.7%, and reduced the total assembly length by 9.6%. The curation process is described at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot 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 both haplotypes using the k -mer database ($k = 31$) 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 *Loensia variegata* specimen generated 24.84 Gb (gigabases) from 2.46 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 612.95 Mb, with a

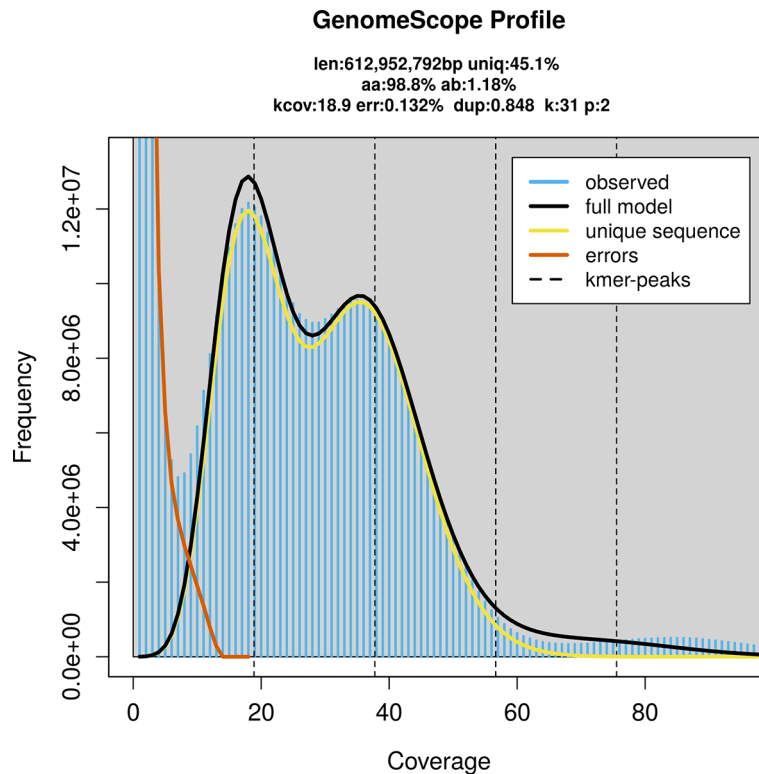


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB77452.

Platform	PacBio HiFi	Hi-C
ToLID	iuLoeVari1	iuLoeVari1
Specimen ID	Ox002633	Ox002633
BioSample (source individual)	SAMEA112232811	SAMEA112232811
BioSample (tissue)	SAMEA112233317	SAMEA112233317
Tissue	whole organism	whole organism
Instrument	Sequel Iie	Illumina NovaSeq 6000
Run accessions	ERR13362642	ERR13363417
Read count total	2.46 million	329.59 million
Base count total	24.84 Gb	99.54 Gb

Table 2. Genome assembly statistics.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	iuLoeVari1.hap1.1	iuLoeVari1.hap2.1
Assembly accession	GCA_964261705.1	GCA_964261695.1
Assembly level	chromosome	scaffold
Span (Mb)	531.42	488.90
Number of chromosomes	9	-
Number of contigs	2 350	2 022
Contig N50	0.4 Mb	0.38 Mb
Number of scaffolds	403	263
Scaffold N50	62.97 Mb	62.71 Mb
Longest scaffold length (Mb)	72.16	-
Sex chromosomes	X	-
Organelles	Mitochondrion: 15.94 kb	-

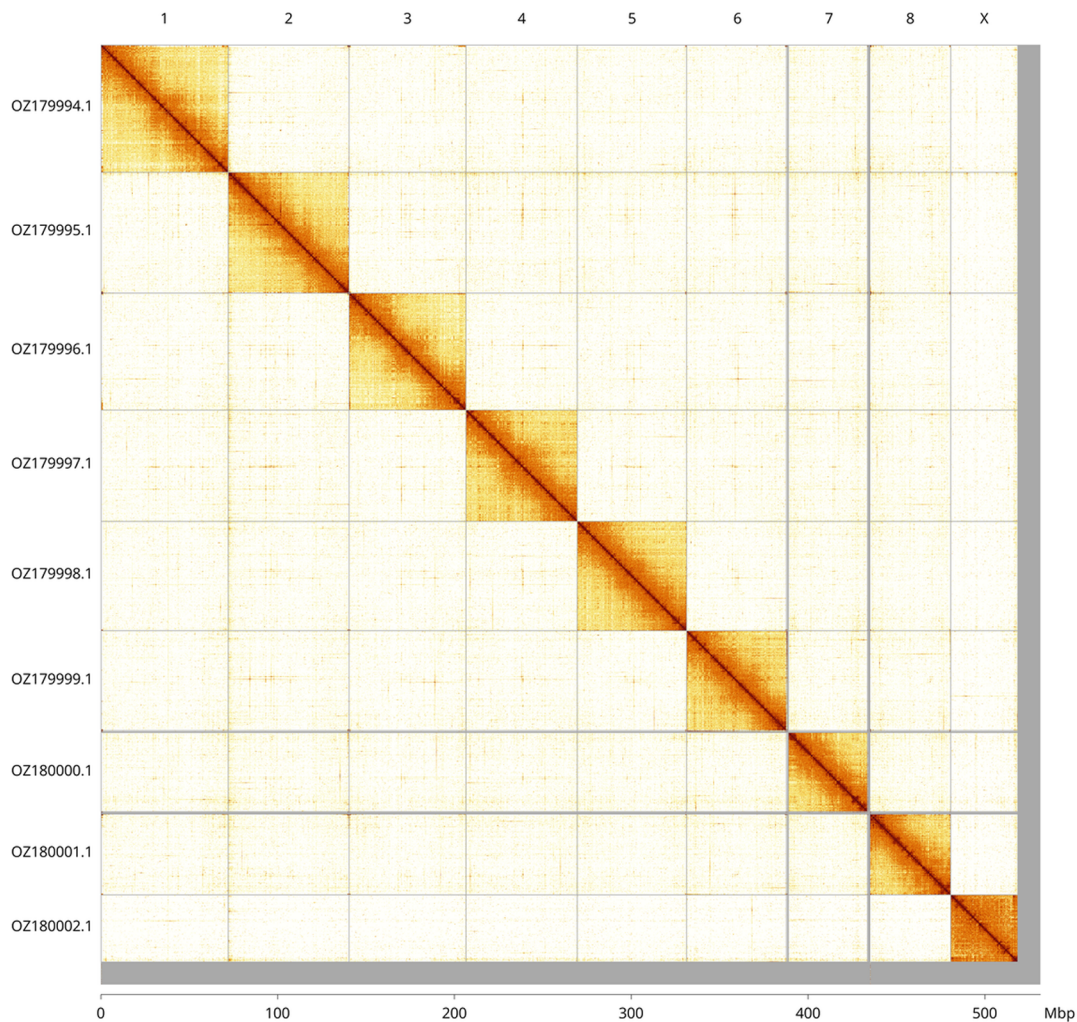


Figure 3. Hi-C contact map of the *Loensia variegata* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

heterozygosity of 1.18% and repeat content of 55.67% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 38× coverage. Hi-C sequencing produced 99.54 Gb from 329.59 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 has a total length of 531.42 Mb in 403 scaffolds, with 1 947 gaps, and a scaffold N50 of 62.97 Mb (Table 2).

Most of the haplotype 1 assembly sequence (97.63%) was assigned to 9 chromosomal-level scaffolds, representing 8 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The X chromosome was assigned based on read coverage, but no Y chromosome was found. Closely related species are known to have the XO karyotype.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Loensia variegata* iuLoeVari1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ179994.1	1	72.16	39
OZ179995.1	2	68.28	39
OZ179996.1	3	66.18	39
OZ179997.1	4	62.97	39
OZ179998.1	5	61.78	39
OZ179999.1	6	57.52	39
OZ180000.1	7	46.07	39.50
OZ180001.1	8	45.77	39.50
OZ180002.1	X	38.10	39.50

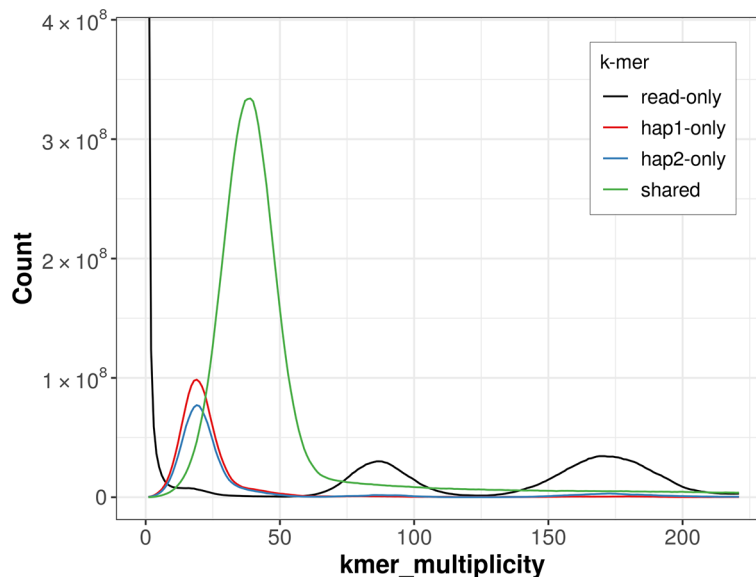


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

The mitochondrial genome was also assembled (length 15.94 kb, OZ180003.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 58.0, and for haplotype 2, 58.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 58.1. The *k*-mer completeness is 76.32% for haplotype 1, 72.28% for haplotype 2, and 91.68% for the combined haplotypes (Figure 4).

BUSCO analysis using the insecta_odb10 reference set ($n = 1\ 367$) identified 96.7% of the expected gene set (single = 94.8%, duplicated = 1.9%) in haplotype 1. For haplotype 2, BUSCO v.5.5.0 analysis identified 94.8% of the expected gene set (single = 93.3%, duplicated = 1.5%). The snail plot in Figure 5 summarises the scaffold length

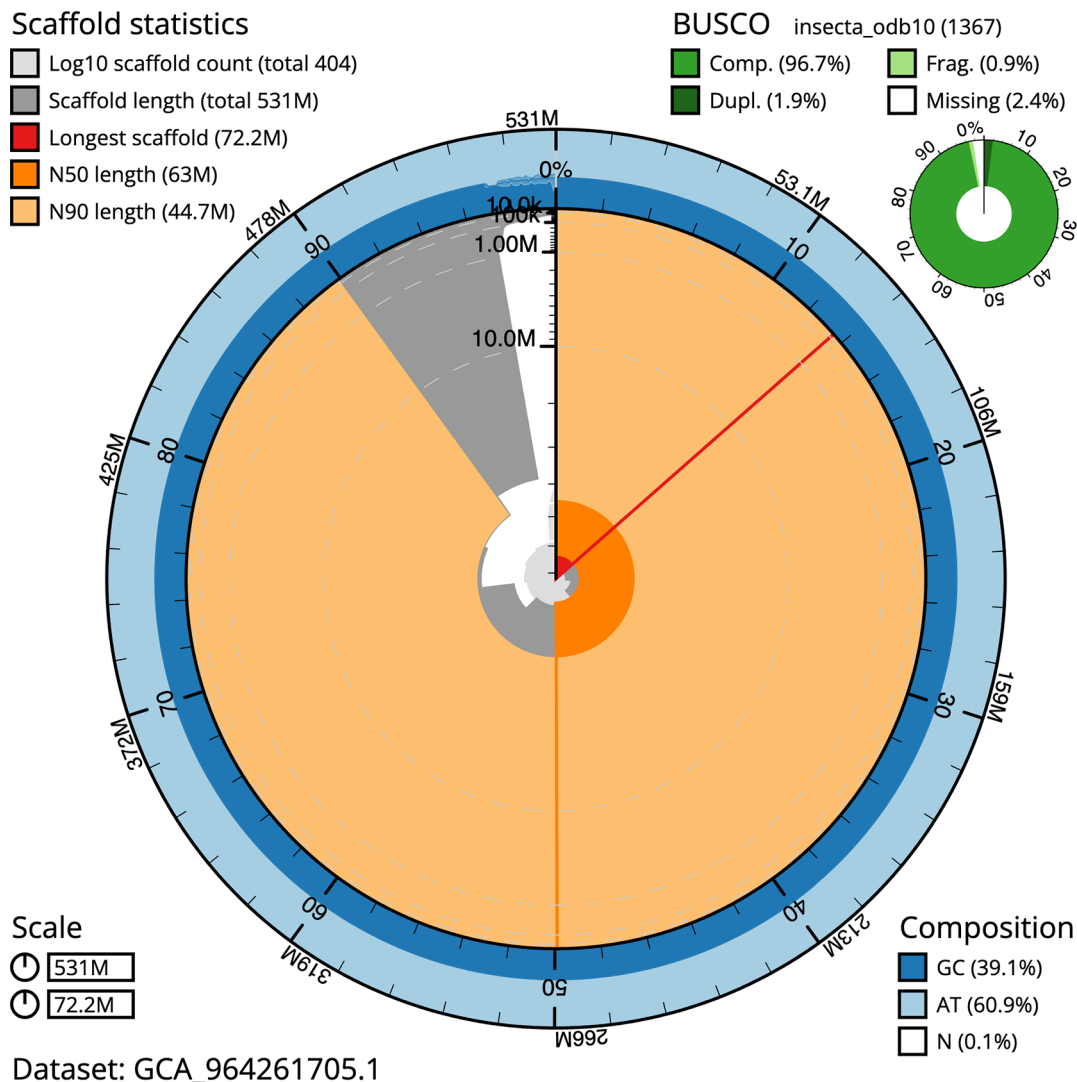


Figure 5. Assembly metrics for iuLoeVari1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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](#).

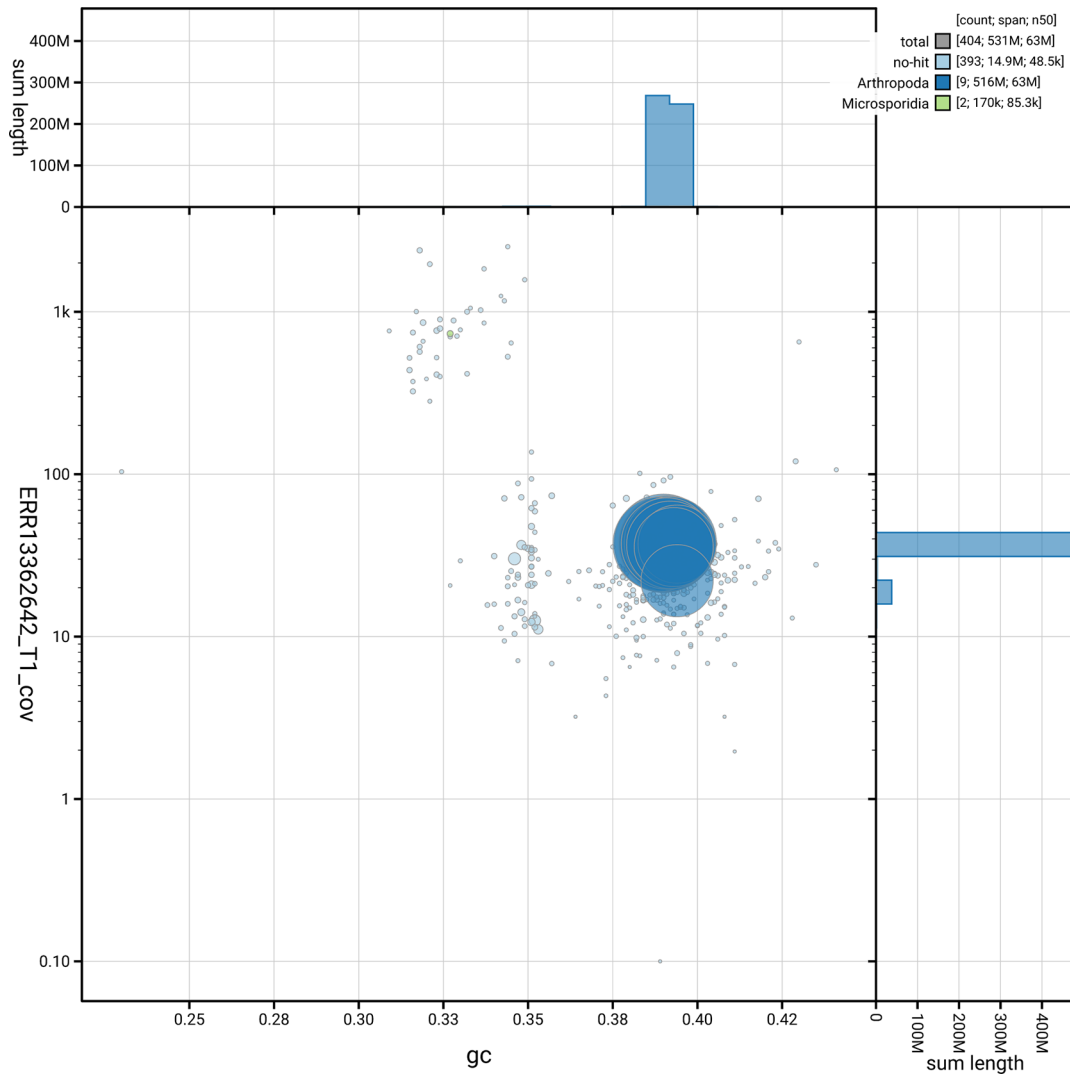


Figure 6. BlobToolKit blob plot for *iuLoeVari1.hap1.1*. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the 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](#).

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 January 2026](#). The EBP metric, calculated for the haplotype 1, is **5.C.Q58**.

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](#)

Table 4. Earth Biogenome Project summary metrics for the *Loensia variegata* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	5.C.Q58	6.C.Q40
Contig N50 length	0.40 Mb	≥1 Mb
Scaffold N50 length	62.97 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 58.0; haplotype 2: 58.2; combined: 58.1	≥40
k-mer completeness	Haplotype 1: 76.32%; Haplotype 2: 72.28%; combined: 91.68%	≥95%
BUSCO	C:96.7% [S:94.8%, D:1.9%], F:0.9%, M:2.4%, n:1 367	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.63%	≥90%

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

- 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: *Loensia variegata*. Accession number [PRJEB77452](#). The genome sequence is released openly for reuse. The *Loensia variegata* 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	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/eziab/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/chhy123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MercuryFK	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
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat. Methods.* 2021; **18**(2): 170–175.
[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.** *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).
[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. USA.* 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](#)
- GBIF Secretariat: **Loensia variegata (Latreille, 1799) in GBIF Backbone Taxonomy.** 2025.
[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.

[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.** *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, 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).

[Publisher Full Text](#)

National Barkfly Recording Scheme: **Introduction to barkflies.** 2025a.

[Reference Source](#)

National Barkfly Recording Scheme: **Identification key — *Loensia variegata*.** 2025b.

[Reference Source](#)

National Barkfly Recording Scheme: **Species account: *Loensia variegata* (Latreille, 1799).** 2025c.

[Reference Source](#)

NatureSpot: ***Loensia variegata*.** 2025.

[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.*: **Mercury: 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.

[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.** *Wellcome Open Res*. 2024; **9**: 339.

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

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; 314–324.

[Publisher Full Text](#)

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

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

Open Peer Review

Current Peer Review Status: 

Version 1

Reviewer Report 25 May 2026

<https://doi.org/10.21956/wellcomeopenres.28991.r155293>

© 2026 Steck 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.



Mireille Steck 

University of Hawai'i at Mānoa, Honolulu, USA

This is a fine addition for a genomic assembly of a barkfly following appropriate genomic assembly methods. Clear cutoffs and assembly statistics have been well reported. The genome assembly looks to be complete and well annotated. This genome will provide important metrics for future insect evolution studies and is a great contribution to the field.

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: Molecular Biology

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.
