



DATA NOTE

The genome sequence of a longhorn beetle, *Stictoleptura scutellata* (Fabricius, 1781) (Coleoptera: Cerambycidae)

[version 1; peer review: 2 approved, 1 approved with reservations]

Steve Crellin¹, 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

¹Independent researcher, Ramsey, UK

V1 First published: 05 Nov 2025, 10:615
<https://doi.org/10.12688/wellcomeopenres.25051.1>
Latest published: 05 Nov 2025, 10:615
<https://doi.org/10.12688/wellcomeopenres.25051.1>

Abstract

We present a genome assembly from an individual female *Stictoleptura scutellata* (longhorn beetle; Arthropoda; Insecta; Coleoptera; Cerambycidae). The assembly contains two haplotypes with total lengths of 1 471.09 megabases and 1 480.10 megabases. Most of haplotype 1 (99.33%) is scaffolded into 10 chromosomal pseudomolecules, including the X sex chromosome, while haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 17.97 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

Stictoleptura scutellata; longhorn beetle; genome sequence; chromosomal; Coleoptera



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

Open Peer Review

Approval Status

✓ ? ✓

	1	2	3
version 1	✓	?	✓
05 Nov 2025	view	view	view

1. **Peng-Yu Jin** , University of the Chinese Academy of Sciences, Beijing, China
Anhui Normal University School of Life Sciences (Ringgold ID: 644804), Wuhu, China
2. **Christopher B. Cunningham**, University of Georgia, Georgia, USA
3. **Min Li**, Taiyuan Normal University, Jinzhong, China

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: Crellin S: 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 Crellin S *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: Crellin S, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of a longhorn beetle, *Stictoleptura scutellata* (Fabricius, 1781) (Coleoptera: Cerambycidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, 10:615
<https://doi.org/10.12688/wellcomeopenres.25051.1>

First published: 05 Nov 2025, 10:615 <https://doi.org/10.12688/wellcomeopenres.25051.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Chrysomeloidea; Cerambycidae; Lepturinae; Lepturini; *Stictoleptura*; *Stictoleptura scutellata* (Fabricius, 1781) (NCBI:txid 879004)

Background

We present a chromosome-level genome sequence for *Stictoleptura scutellata*. This longhorn beetle is considered scarce in Britain and depends on ancient broadleaved woodland and wood pasture, with records concentrated in the New Forest, North Downs and the Epping Forest area (Falk, 2025). Across its wider range it is a western Palaearctic, thermophilic species associated with beech forests and other deciduous woodland (Hoskovec *et al.*, 2025). In Britain and Ireland, records are largely from south-east England (NBN Atlas Partnership, 2025).

Stictoleptura scutellata is ecologically linked to deciduous trees, especially beech (*Fagus sylvatica*), but it is polyphagous, developing in the dead wood of a range of broadleaved hosts including *Fagus*, *Carpinus*, *Quercus*, *Alnus*, *Betula*, *Corylus avellana*, *Fraxinus*, *Castanea*, *Populus* and others (Ellis, 2025; Hoskovec *et al.*, 2025). Larvae feed in dead wood of moderate to large diameter in standing trunks, fallen timber, trunk cavities and dead branches; pupation occurs in the wood. It has a prolonged life cycle (typically lasting 2 to 3 years, up to 4 years in some observations), and adults are active from late spring to late summer (May to September), often visiting flowers such as hawthorn and bramble (Falk, 2025; Hoskovec *et al.*, 2025).

Adult *S. scutellata* beetles are medium-sized (12–21 mm), blackish, strongly punctate, and show a partially yellow scutellum, aiding field recognition (Falk, 2025; Hoskovec *et al.*, 2025).

A total of 5 495 GBIF occurrence records with coordinates are available for this species, with observations primarily from Europe and smaller numbers from Asia and North Africa; records are most frequent from France, Sweden, the United Kingdom, Germany, Austria and the Netherlands (GBIF Secretariat, 2025). This assembly was generated as part of the Darwin Tree of Life Project, which aims to deliver high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation and sustainable use of biodiversity (Blaxter *et al.*, 2022). This assembly is the first genome for the genus *Stictoleptura* as of September 2025 (NCBI datasets, O'Leary *et al.*, 2024).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Stictoleptura scutellata* (specimen ID NHMUK015 059141, ToLID icStiScut1; Figure 1), collected from Catfield Fen, England, United Kingdom (latitude 52.73, longitude −1.51)



Figure 1. Photograph of the *Stictoleptura scutellata* (icStiScut1) specimen used for genome sequencing.

on 2022-07-04. The specimen was collected by Steve Crellin and identified by Chris Spilling. The same specimen was used for RNA 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.

Sample metadata were collected in line with the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

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 icStiScut1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the head and thorax was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted 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 43 ng/μL and a yield of 2 021.00 ng, with a fragment size of 11.7 kb. The 260/280 spectrophotometric ratio was 1.94, and the 260/230 ratio was 2.03.

RNA was also extracted from abdomen tissue of *icStiScut1* in the Tree of Life Laboratory at the WSI using the [RNA Extraction: Automated MagMax™ mirVana protocol](#). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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 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 head and thorax tissue of the *icStiScut1* 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 (ThermoFisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using

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

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

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X to generate 150-bp paired-end reads.###
Genome assembly

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

The HiFi reads were assembled using Hifiasm 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 22 breaks, 46 joins, and removal of 9 haplotypic duplications. The curation process is described at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextView 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 *Stictoleptura scutellata* specimen generated 116.52 Gb (gigabases) from 10.95 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 476.23 Mb, with a heterozygosity of 0.98% and repeat content of 39.25% (Figure 2). These estimates guided expectations for the assembly. Based

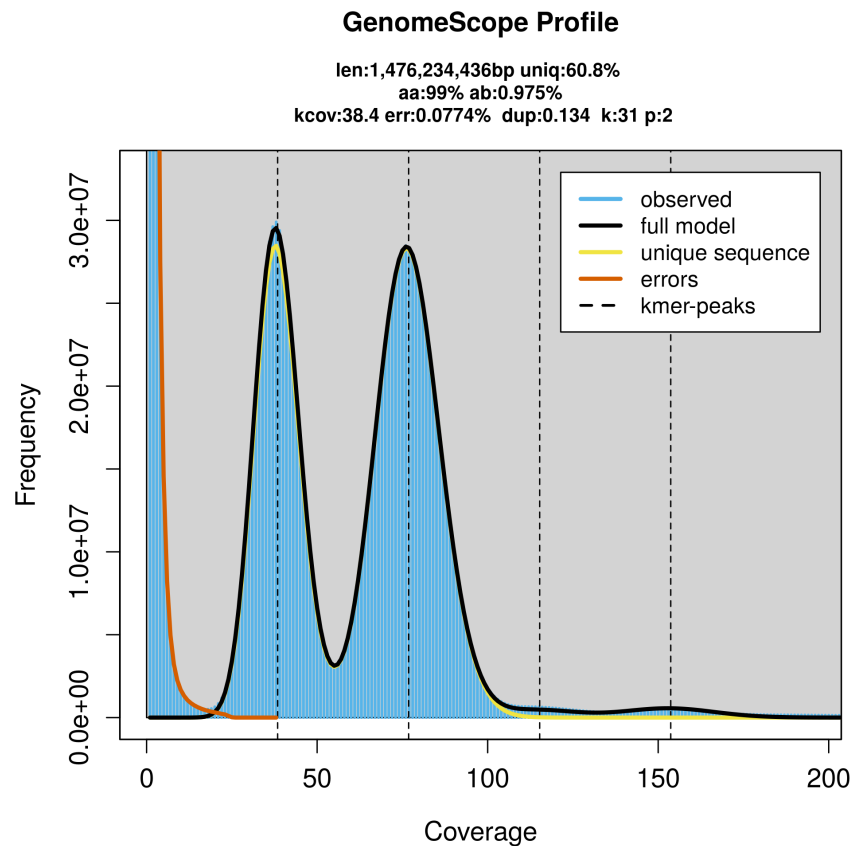


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

on the estimated genome size, the sequencing data provided approximately 77× coverage. Hi-C sequencing produced 108.64 Gb from 719.46 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The 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 1 471.09 Mb in 230 scaffolds, with 272 gaps, and a scaffold N50 of 172.53 Mb (Table 2).

Most of the assembly sequence (99.33%) was assigned to 10 chromosomal-level scaffolds, representing 9 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 assigned based on alignment to that of *Lep-tura quadrifasciata* (GCA_963675555.1). During curation, we observed a haplotypic inversion on Chromosome 5 in the region ~85.47–88.39 Mb.

The mitochondrial genome was also assembled (length 17.97 kb, OZ125674.1). 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 64.6, and for haplotype 2, 64.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.6. The *k*-mer completeness is 79.54% for haplotype 1, 79.68% for haplotype 2, and 99.60% for the combined haplotypes (Figure 4).

Table 1. Specimen and sequencing data for BioProject PRJEB77316.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	icStiScut1	icStiScut1	icStiScut1
Specimen ID	NHMUK015059141	NHMUK015059141	NHMUK015059141
BioSample (source individual)	SAMEA112964455	SAMEA112964455	SAMEA112964455
BioSample (tissue)	SAMEA112975689	SAMEA112975689	SAMEA112975653
Tissue	head and thorax	head and thorax	abdomen
Instrument	Sequel IIE	Illumina NovaSeq 6000	Illumina NovaSeq X
Run accessions	ERR13362602; ERR13362603	ERR13350325	ERR14792835
Read count total	10.95 million	719.46 million	91.56 million
Base count total	116.52 Gb	108.64 Gb	13.82 Gb

Table 2. Genome assembly statistics.

Assembly name	icStiScut1.hap1.1	icStiScut1.hap2.1
Assembly accession	GCA_964212005.1	GCA_964212015.1
Assembly level	chromosome	scaffold
Span (Mb)	1 471.09	1 480.10
Number of chromosomes	10	Scaffold-level
Number of contigs	502	438
Contig N50	10.87 Mb	10.13 Mb
Number of scaffolds	230	201
Scaffold N50	172.53 Mb	170.6 Mb
Longest scaffold length (Mb)	286.49	-
Sex chromosomes	X	-
Organelles	Mitochondrion: 17.97 kb	-

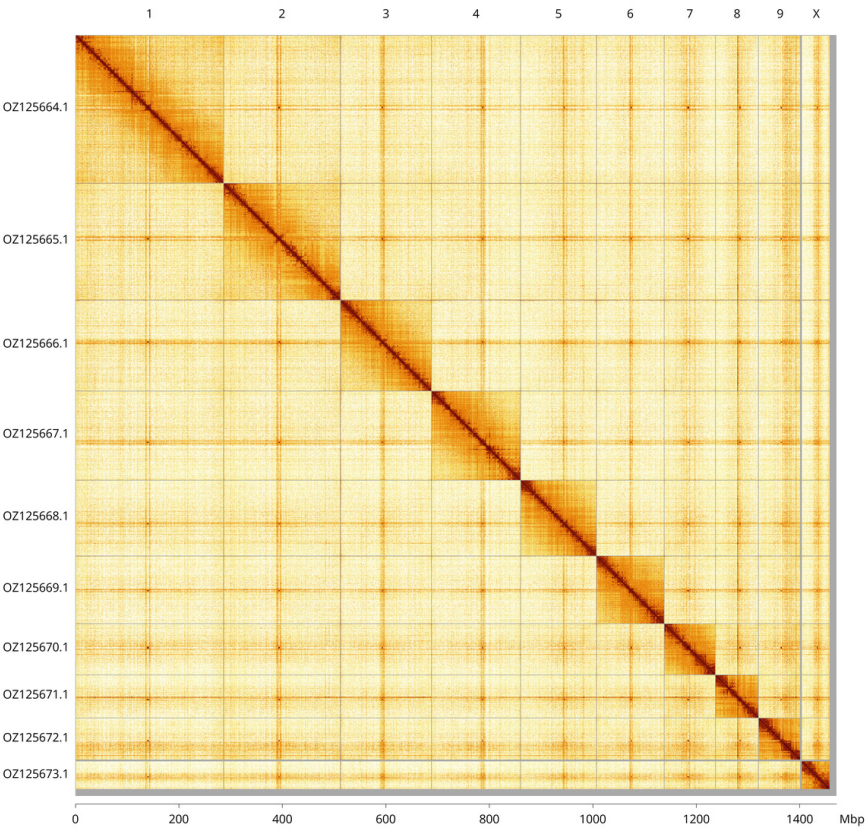


Figure 3. Hi-C contact map of the *Stictoleptura scutellata* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Stictoleptura scutellata* icStiScut1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ125664.1	1	286.49	33.50
OZ125665.1	2	225.97	33.50
OZ125666.1	3	175.87	33.50
OZ125667.1	4	172.53	33.50
OZ125668.1	5	146.22	33.50
OZ125669.1	6	131.20	33.50
OZ125670.1	7	98.91	33.50
OZ125671.1	8	83.23	33.50
OZ125672.1	9	83.17	34
OZ125673.1	X	57.67	34

BUSCO analysis using the endopterygota_odb10 reference set ($n = 2\,124$) identified 99.2% of the expected gene set (single = 96.6%, duplicated = 2.7%) 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 **7.C.Q64**, meeting the recommended reference standard.

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

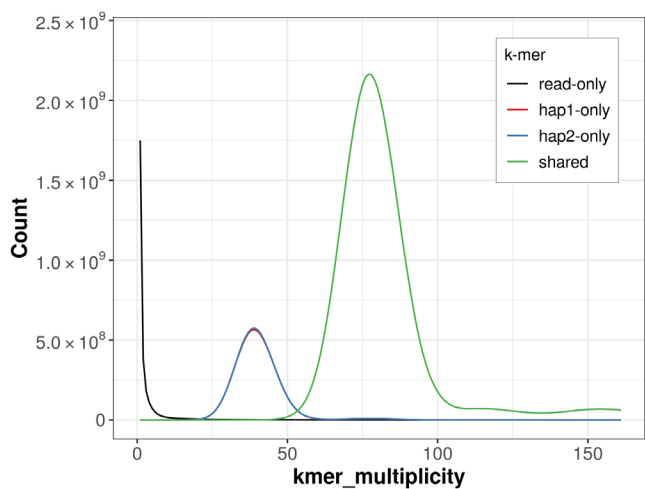


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

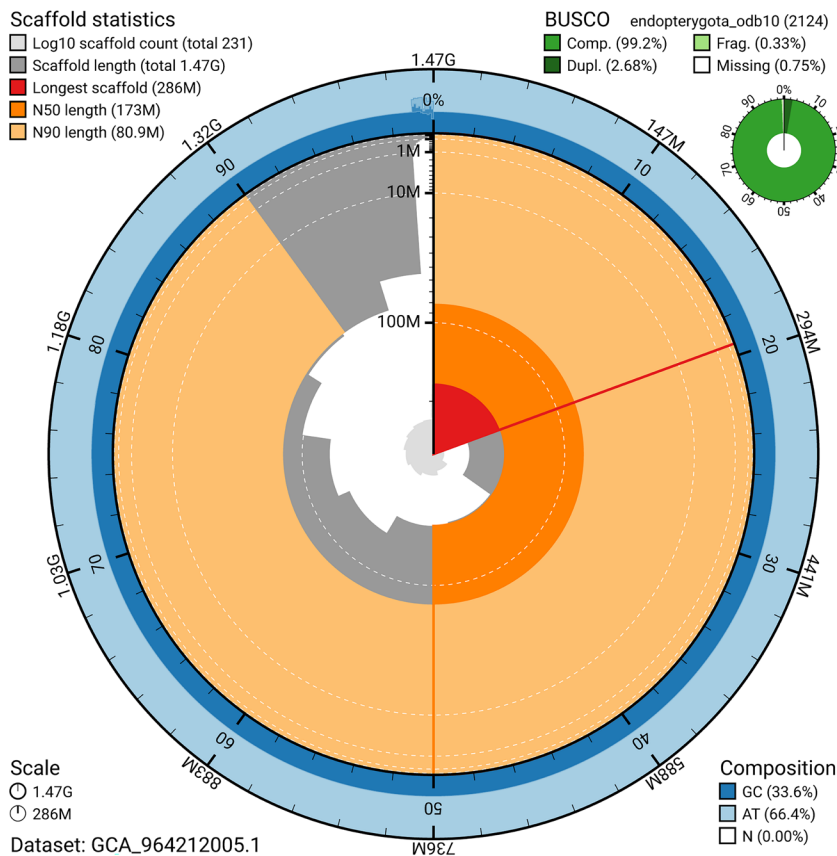


Figure 5. Assembly metrics for icStiScut1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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](#).

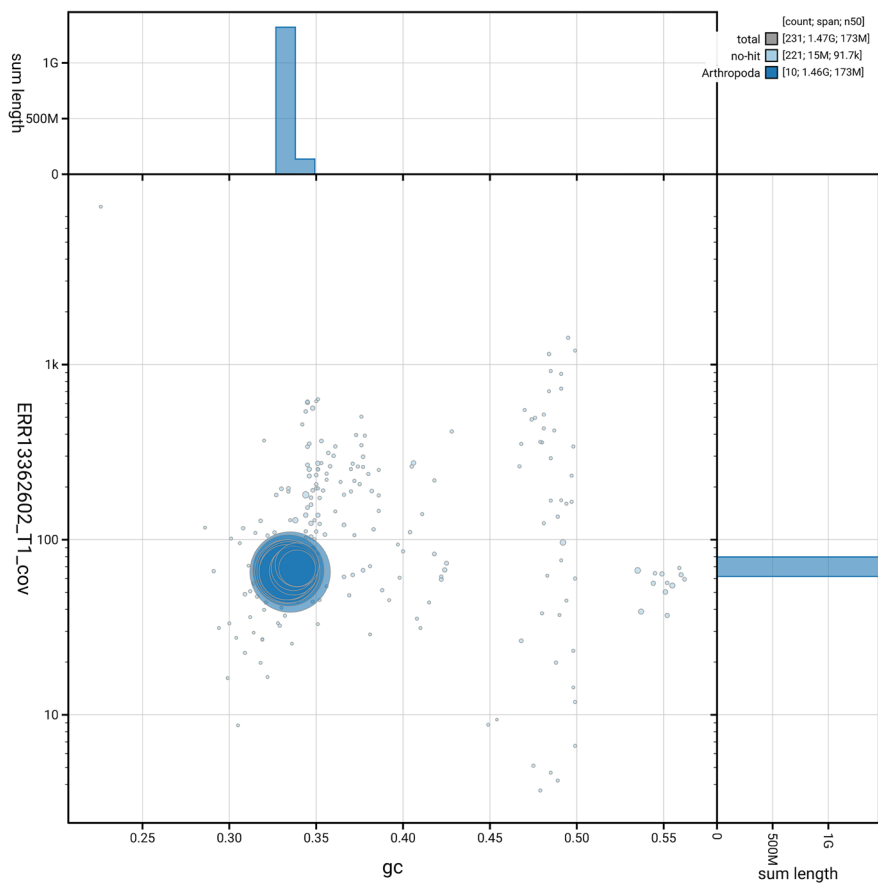


Figure 6. BlobToolKit GC-coverage plot for icStiScut1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Stictoleptura scutellata* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.C.Q64	6.C.Q40
Contig N50 length	10.87 Mb	≥ 1 Mb
Scaffold N50 length	172.53 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.6; haplotype 2: 64.5; combined: 64.6	≥ 40
k-mer completeness	Haplotype 1: 79.54%; Haplotype 2: 79.68%; combined: 99.60%	≥ 95%
BUSCO	C:99.2% [S:96.6%; D:2.7%]; F:0.3%; M:0.4%; n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.33%	≥ 90%

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: *Stictoleptura scutellata*. Accession number [PRJEB77316](#). The genome sequence is released openly for reuse. The *Stictoleptura scutellata* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#) and [Table 2](#).

Production code used in genome assembly at the WSI Tree of Life is available at <https://github.com/sanger-tol>. [Table 5](#) lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of the [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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhy123/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
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow

Software	Version	Source
PretextSnapshot	-	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
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	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](#)
- Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **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](#)
- 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](#)
- Ellis WN: ***Stictoleptura scutellata* (Fabricius, 1781) (large black longhorn) in Plant Parasites of Europe.** 2025.
[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](#)
- Falk S: ***Stictoleptura scutellata* (Large Black Longhorn) on Flickr.** 2025.
[Reference Source](#)
- 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: ***Stictoleptura scutellata* (Fabricius, 1781) in GBIF Backbone Taxonomy.** 2025.
[Reference Source](#)
- Hoskovec M, Navrátil D, Jelínek P, *et al.*: ***Stictoleptura scutellata* (Fabricius, 1781) in Cerambycidae: Longhorn beetles (Cerambycidae, Coleoptera) of the West Palaearctic region, neighboring territories and countries of the former Soviet Union.** 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): g1aa153.
[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, 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.
[Reference Source](#)
- NBN Atlas Partnership: ***Stictoleptura scutellata* (Fabricius, 1781).** 2025.
[Reference Source](#)
- O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and**

metadata for species across the Tree of Life with NCBI Datasets. *Sci Data*. 2024; **11**(1): 732.

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

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun*. 2020; **11**(1): 1432.

[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](#)

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, 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.** 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 05 January 2026

<https://doi.org/10.21956/wellcomeopenres.27617.r140244>

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



Min Li

Taiyuan Normal University, Jinzhong, China

This paper presents a chromosome-level genome assembly of *Stictoleptura scutellata* (Coleoptera: Cerambycidae), a key output of the Darwin Tree of Life Project. As the first genome assembly for the genus *Stictoleptura*, this study fills a critical gap in genomic resources for longhorn beetles and contributes valuable data to eukaryotic biodiversity research in Britain and Ireland. The manuscript is well-structured, methodologically rigorous, and comprehensive in presenting experimental details, results, and data availability. The assembly quality meets high standards (e.g., BUSCO completeness of 99.2%, QV score of 64.6) and adheres to the specifications of data note articles in Wellcome Open Research. Overall, the work is scientifically sound and suitable for Indexing with minor revisions.

I suggest only a few minor changes:

1. Specify the expected release channel for annotation files (e.g., Ensembl, NCBI) in the "Data Availability" section to guide comparative genomics and functional studies.
2. Before DNA extraction, were the specimens undergone aseptic treatment?
3. Expand on population status variations across the species' range (e.g., endangered status in different regions, major threats to habitats) to emphasize the relevance of genomic data for conservation.
4. Page 6: For the sentence "Chromosome X was assigned based on alignment to that of *Leptura quadrifasciata* (GCA_963675555.1)", change "aligment" for "alignment".
5. Ensure consistency in software version reporting (e.g., specify PretextSnapshot version if available, or note "unversioned" if applicable).

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: Insect Molecular Ecology

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 29 December 2025

<https://doi.org/10.21956/wellcomeopenres.27617.r140240>

© 2025 Cunningham C. 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.



Christopher B. Cunningham

University of Georgia, Georgia, USA

The project continues to produce quality resources for the insect genomics community. The methods are sound and relatively well documented. The article needs has enough details to be reasonably reproducible, especially now with the inclusion of the pipeline documentation. The data supporting the article are publicly available.

Keywords. These are repeated form the title and abstract so add no value. Google indexes off these too so they should reflect some broader value.

Background. Background. The stated motivation for this study in the introduction is fine – part of a very large survey.

Methods. Version number for the software used is not reported. I think its reasonably reproducible but reporting that is the field standard.

BUSCO completeness needs to be assessed with the updated OrthoDB version 12. Difference from 10 to 12 have made a big difference for several insects I work with and colleagues have reported the same. The update has been out for 13 months.

I find it odd to produce RNA-seq data and then not do anything with it. But its generation is described reasonably enough that it could be useful for someone following up on this report.

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: Insects, Behavior, Evolution, Genetics, Genomics, Epigenetics, Reproduction.

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 22 November 2025

<https://doi.org/10.21956/wellcomeopenres.27617.r138767>

© 2025 Jin P. 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.



Peng-Yu Jin 

¹ University of the Chinese Academy of Sciences, Beijing, Beijing, China

² Anhui Normal University School of Life Sciences (Ringgold ID: 644804), Wuhu, Anhui, China

The manuscript is excellent in describing the species, sample source, sequencing and analysis methods, results, and data availability. I have no concerns; it provides a very good reference workflow for genome sequencing and assembly. One minor point is that, as in other papers, GFF annotation results are rarely included. Providing these would be useful for future downstream analyses or comparative studies.

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: Comparative genomics; Single-cell RNA sequencing; Animal behavior

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.
