Check for updates

DATA NOTE

The genome sequence of the Brassy Long-horn, *Nemophora metallica* (Poda, 1761) (Lepidoptera: Adelidae)

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

Jaakko Pohjoismaki[1], Marko Mutanen[2], Charlotte J. Wright[3], Joana I. Meier[3], Mark L. Blaxter[3],
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, Project Psyche Community

[1]University of Eastern Finland, Joensuu, North Karelia, Finland
[2]University of Oulu, Oulu, Finland
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 15 Aug 2025, **10**:445
https://doi.org/10.12688/wellcomeopenres.24685.1
Latest published: 15 Aug 2025, **10**:445
https://doi.org/10.12688/wellcomeopenres.24685.1

Abstract

We present a genome assembly from a male specimen of *Nemophora metallica* (Brassy Long-horn; Arthropoda; Insecta; Lepidoptera; Adelidae). The assembly contains two haplotypes with total lengths of 710.84 megabases and 709.72 megabases. Most of haplotype 1 (99.89%) is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.09 kilobases.

Keywords

Nemophora metallica, Brassy Long-horn, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 1 15 Aug 2025	✓ view	? view	? view

1. **Arjen Van 't Hof**, Biology Centre of the Czech Academy of Sciences, Budějovice, Czech Republic
2. **Frederique Hilliou**, Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Provence-Alpes-Cote d'Azur Site de Sophia (Ringgold ID: 60158), Sophia Antipolis, France
3. **Christopher B. Cunningham**, University of Georgia, Georgia, USA

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

Corresponding author: Charlotte J. Wright (cw22@sanger.ac.uk)

Author roles: Pohjoismaki J: Investigation, Resources, Writing – Original Draft Preparation; **Mutanen M**: Resources, Supervision; **Wright CJ**: Funding Acquisition, Supervision; **Meier JI**: Funding Acquisition, Project Administration, Supervision; **Blaxter ML**: Funding Acquisition, Project Administration, Supervision;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Pohjoismaki 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: Pohjoismaki J, Mutanen M, Wright CJ *et al.* **The genome sequence of the Brassy Long-horn, *Nemophora metallica* (Poda, 1761) (Lepidoptera: Adelidae) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2025, **10**:445 https://doi.org/10.12688/wellcomeopenres.24685.1

First published: 15 Aug 2025, **10**:445 https://doi.org/10.12688/wellcomeopenres.24685.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Incurvarioidea; Adelidae; *Nemophora*; *Nemophora metallica* (Poda, 1761) (NCBI:txid1594367)

Background

Nemophora metallica is a day-flying moth in the family Adelidae. These are commonly referred to as the longhorn moths due to the extraordinarily long antennae of males, which can exceed three times their body length. These small moths have a wingspan of 15–20 mm and metallic bronze forewings that shimmer in sunlight, creating an iridescent effect (Bengtsson *et al.*, 2008; Bryner, 2020). The males' distinctive thread-like antennae are distally white.

This species is widely distributed across Europe (Bengtsson *et al.*, 2008; Bryner, 2020) particularly in temperate regions, inhabiting dry meadows and other open habitats rich in wildflowers. Adults are most active on sunny days, flying from late May to early July. They feed on nectar and can often be seen visiting flowers. Females lay their eggs on flower heads of Scabious (*Knautia* spp.) (Bengtsson *et al.*, 2008; Bryner, 2020); in Finland the sole host plant is Field Scabious (*Knautia arvensis*). The larva initially mines the developing seeds of its hostplant and later drops on ground, feeding on the plant's lower leaves and detritus. Larvae construct portable cases from silk and plant fragments, which they carry as they develop, a behaviour typical of the Adelidae family. Pupation occurs within these cases in spring, after overwintering as a last instar larva.

The populations of *N. metallica* are sensitive to habitat degradation due to changes in land management practices and natural succession. Conservation of this species depends on maintaining diverse, flower-rich habitats. The species is classified as endangered (EN) in Finland (Nupponen *et al.*, 2019). The species is considered to be genetically heterogenous, with possible cryptic variation, especially in Southern France (Bryner, 2020).

The genome assembly for *Nemophora metallica* is a significant addition to the resources available for micromoths, providing new opportunities for studying the genetics of iridescence, sexual dimorphism, and larval case-building behaviours. Comparative analyses with other Adelidae genomes may shed light on the evolutionary history of this ancient lineage and their ecological adaptations. Furthermore, the genomic data will support population genetics studies, informing conservation efforts for this species.

We present a chromosome-level, haplotype-resolved genome sequence of *Nemophora metallica*, sequenced as part of Project Psyche. The sequence data were derived from a male specimen (Figure 1) collected from Hovattavaara, Liperi, North Karelia, Finland.

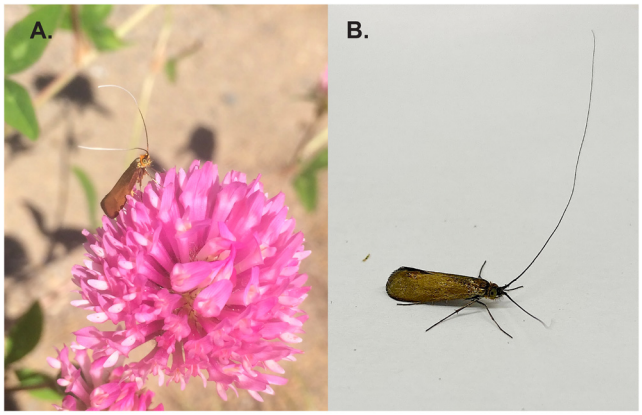

Figure 1. (**A**) A male *Nemophora metallica* feeding on Red Clover (*Trifolium pratense*), Liperi, Finland. (**B**) The specimen used for the generation of the genome assembly (ilNemMetl5).

Methods

Sample acquisition

The specimen used for genome sequencing was an adult male *Nemophora metallica* (specimen ID SAN28000009, ToLID ilNemMetl5; Figure 1), collected from Hovattavaara, Liperi, North Karelia, Finland (latitude 62.6206, longitude 29.4757) on 06/07/2023. The specimen was collected and identified by Jaakko Pohjoismaki (University of Eastern Finland).

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 ilNemMetl5 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 21.42 ng/μL and a yield of 1 006.74 ng, with a fragment size of 16.4 kb. The 260/280 spectrophotometric ratio was 1.81, and the 260/230 ratio was 3.2.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Samples with an average fragment size greater than 8 kb and total mass exceeding 400 ng

were eligible for the low-input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and required sequencing depth. Libraries were prepared using the SMRTbell Prep Kit 3.0 according to 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.

Specimen details, sequencing platforms, and data yields are summarised in Table 1.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the ilNemMetl5 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. 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.

Specimen details, sequencing platforms, and data yields are summarised in Table 1.

Table 1. Specimen and sequencing data for BioProject.

Platform	PacBio HiFi	Hi-C
ToLID	ilNemMetl5	ilNemMetl5
Specimen ID	SAN28000009	SAN28000009
BioSample (source individual)	SAMEA114539588	SAMEA114539588
BioSample (tissue)	SAMEA114539622	SAMEA114539622
Tissue	whole organism	whole organism
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13510350	ERR13621458; ERR13621459
Read count total	2.01 million	1 709.09 million
Base count total	22.76 Gb	258.07 Gb

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 7 breaks and 15 joins. 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

Chromosomal painting was performed using lep_busco_painter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths from the genome index produced using SAMtools faidx (Danecek *et al.*, 2021). Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used 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 BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoaT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO 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 package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Nemophora metallica* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 22.76 Gb (gigabases) from 2.01 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 708.66 Mb, with a heterozygosity of 0.62% and repeat content of 38.78%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 31× coverage. Hi-C sequencing produced 258.07 Gb from 1 709.09 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 710.84 Mb in 61 scaffolds, with 210 gaps, and a scaffold N50 of 25.22 Mb (Table 2).

Most of the assembly sequence (99.89%) was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. The Z chromosome was identified based on synteny with *Nematopogon swammerdamellus* (GCA_946902875.1) (Langdon *et al.*, 2023). These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; Table 3). Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 3).

The mitochondrial genome was also assembled. 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 66.3, and for haplotype 2, 66.6. When the two haplotypes are combined, the assembly

Table 2. Genome assembly statistics.

Assembly name	ilNemMetl5.hap1.1	ilNemMetl5.hap2.1
Assembly accession	GCA_964261965.1	GCA_964261955.1
Assembly level	chromosome	scaffold
Span (Mb)	710.84	709.72
Number of chromosomes	31	N/A
Number of contigs	271	259
Contig N50	4.78 Mb	4.63 Mb
Number of scaffolds	61	61
Scaffold N50	25.22 Mb	25.06 Mb
Longest scaffold length (Mb)	36.77	N/A
Sex chromosomes	Z	N/A
Organelles	Mitochondrial genome: 15.09 kb	N/A

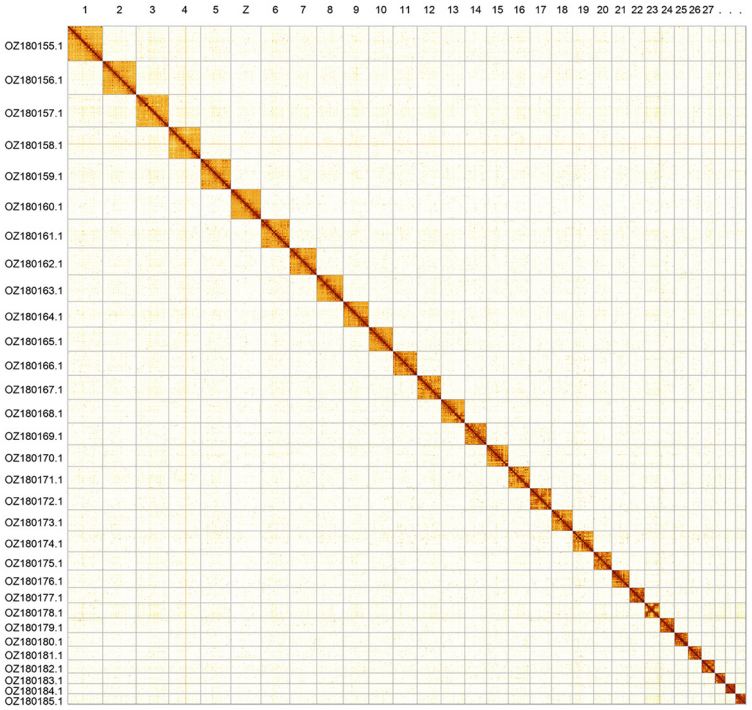

Figure 2. Hi-C contact map of the *Nemophora metallica* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Nemophora metallica* ilNemMetl5.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ180155.1	1	36.77	37.50	M2
OZ180156.1	2	35.20	37.50	M1
OZ180157.1	3	33.98	37.50	M17;M20
OZ180158.1	4	33.32	37.50	M3
OZ180159.1	5	31.80	37.50	M5
OZ180161.1	6	30.08	37.50	M8
OZ180162.1	7	28.32	37	M9
OZ180163.1	8	27.84	37.50	M12
OZ180164.1	9	26.71	37.50	M4
OZ180165.1	10	25.39	37.50	M7
OZ180166.1	11	25.22	37	M16
OZ180167.1	12	25.06	37.50	M6
OZ180168.1	13	24.79	37	M18
OZ180169.1	14	22.84	37	M22
OZ180170.1	15	22.84	37.50	M15
OZ180171.1	16	22.60	37.50	M11
OZ180172.1	17	22.48	37	M21
OZ180173.1	18	22.23	37.50	M10
OZ180174.1	19	21.90	38	M23
OZ180175.1	20	19.18	38	M14
OZ180176.1	21	18.35	37.50	M13
OZ180177.1	22	16.01	37	M24
OZ180178.1	23	15.80	39	M31
OZ180179.1	24	15.23	38	M19
OZ180180.1	25	14.37	37.50	M26
OZ180181.1	26	14.23	38	M28
OZ180182.1	27	13.70	38	M27
OZ180183.1	28	11.06	38	M29
OZ180184.1	29	10.68	38.50	M25
OZ180185.1	30	10.67	39.50	M30
OZ180160.1	Z	31.40	37	MZ

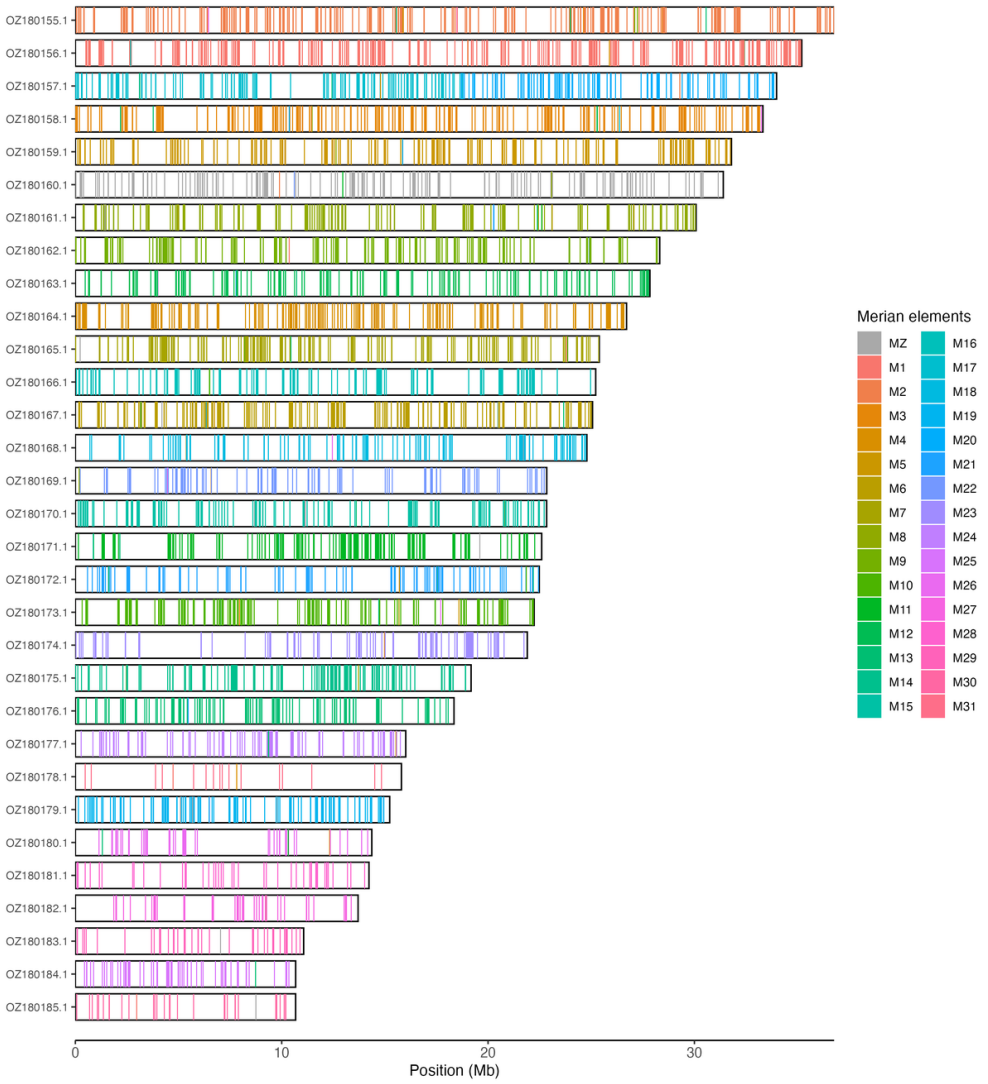

Figure 3. Merian elements painted across chromosomes in the ilNemMetl5.hap1.1 assembly of *Nemophora metallica*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

achieves an estimated QV of 66.5. The *k*-mer completeness is 85.13% for haplotype 1, 85.09% for haplotype 2, and 99.17% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 91.9% of the expected gene set (single = 90.9%, duplicated = 1.1%) 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) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q66**, 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 Tree of Life collaborator. 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:

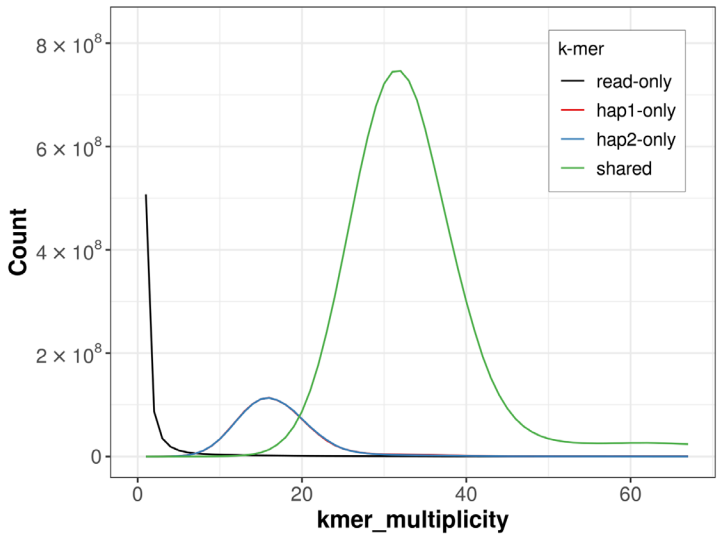

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

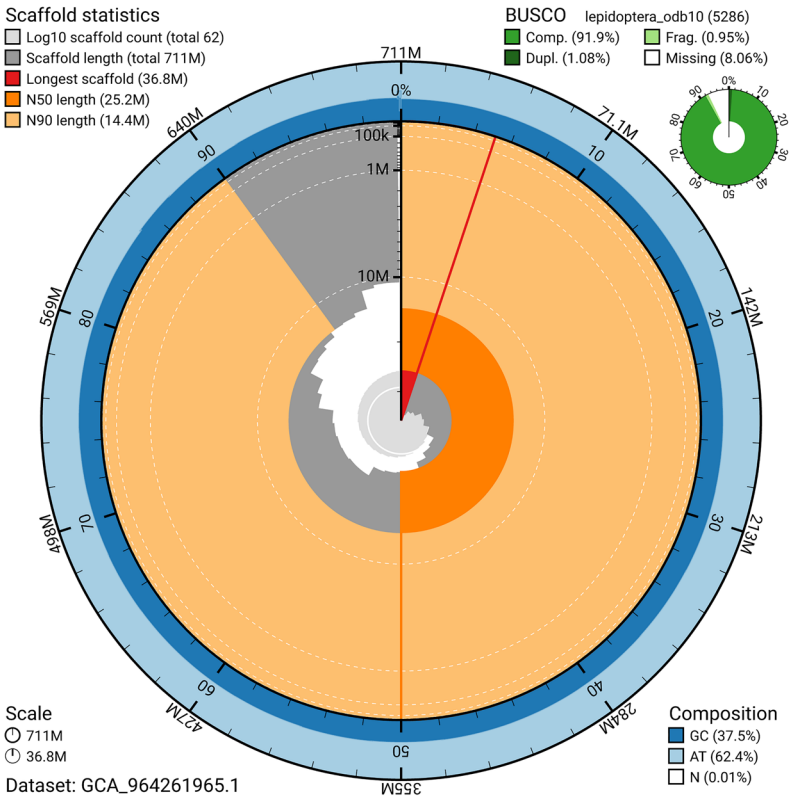

Figure 5. Assembly metrics for ilNemMetl5.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

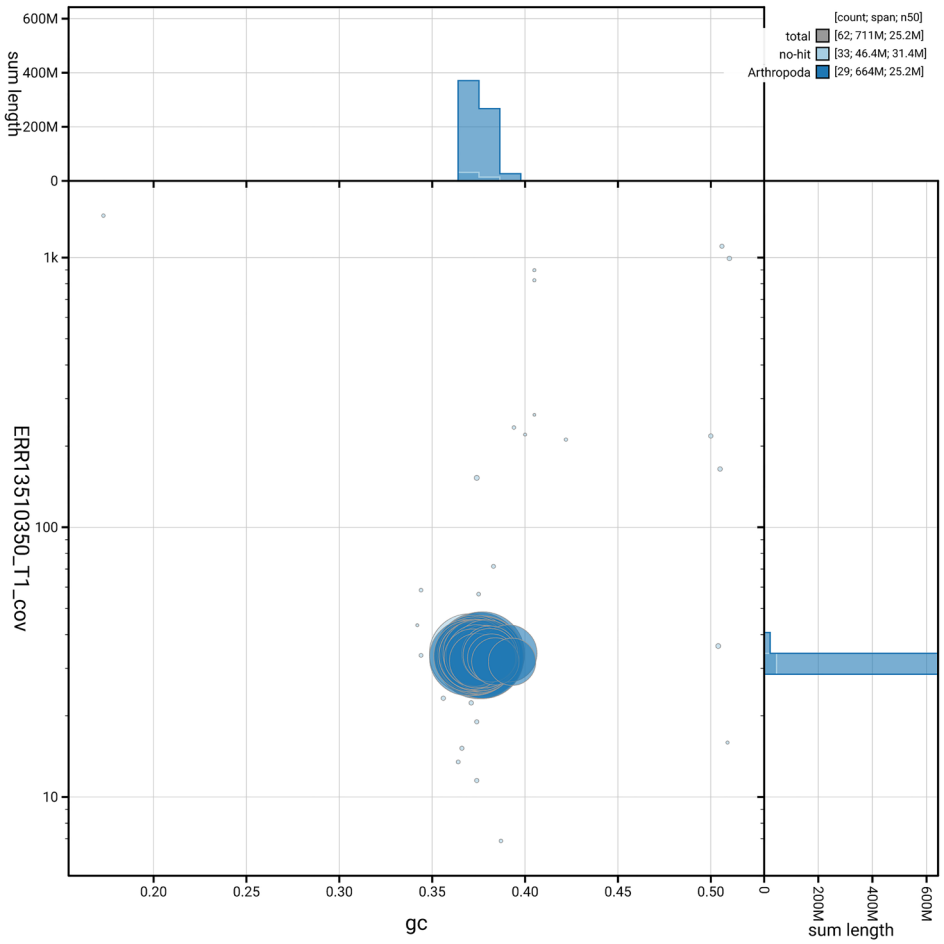

Figure 6. BlobToolKit GC-coverage plot for ilNemMetl5.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Nemophora metallica* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q66	6.C.Q40
Contig N50 length	4.78 Mb	≥ 1 Mb
Scaffold N50 length	25.22 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 66.3; haplotype 2: 66.6; combined: 66.5	≥ 40
k-mer completeness	Haplotype 1: 85.13%; Haplotype 2: 85.09%; combined: 99.17%	≥ 95%
BUSCO	C:91.9%[S:90.9%,D:1.1%], F:0.9%,M:7.1%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.89%	≥ 90%

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international).

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Tree of Life collaborators.

Data availability

European Nucleotide Archive: Nemophora metallica (brassy long-horn). Accession number PRJEB78967. The genome sequence is released openly for reuse. The *Nemophora metallica* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). 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 Ensembl at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Pipelines used for genome assembly at the WSI Tree of Life are available at https://pipelines.tol.sanger.ac.uk/pipelines. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- 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
- Project Psyche Community.

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/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
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot

Software	Version	Source
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

Bengtsson BA, Johansson R, Palmqvist G: **Nationalnyckeln till Sveriges flora och fauna. Fjärilar: Käkmalar—säckspinnare / Lepidoptera: Micropterigidae—Psychidae.** Uppsala: ArtDatabanken, SLU, 2008.
Reference Source

Bryner R: **Adelidae (Lepidoptera) – Beitrag zur Kenntnis der Biologie und Bestimmungshilfe für die europäischen Arten.** *Contributions to Natural History.* 2020; **38**: 1–475.
Reference Source

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

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

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

Kriventseva EV, Kuznetsov D, Tegenfeldt F, *et al.*: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.** *Nucleic Acids Res.* 2019; **47**(D1): D807–D811.
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

Langdon WBV, Baumberg C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.*: **The genome sequence of the Large Longhorn, *Nematopogon swammerdamella* (Linnaeus, 1758) [version 1; peer review: 2 approved, 1 approved with reservations].** *Wellcome Open Res.* 2023; **8**: 531.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Manni M, Berkeley MR, 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

Nupponen K, Nieminen M, Kaitila JP, *et al.*: **Perhoset – Butterflies and Moths Lepidoptera.** In: E. Hyvärinen, A. Juslén, E. Kemppainen, *et al.* (eds), *Suomen Lajien Uhanalaisuus – Punainen Kirja 2019 [Red Book of Finnish Species 2019, in Finnish with English Summaries].* Helsinki: Ympäristöministeriö & Suomen ympäristökeskus, 2019; 470–508.

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

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

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–790.
PubMed Abstract | Publisher Full Text | Free 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 07 October 2025

https://doi.org/10.21956/wellcomeopenres.27201.r132335

© **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 report on the Brassy Long-horn, *Nemophora m*etallica, is no different. 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 publication are publicly available.

Abstract. There is no justification within the abstract, just results. At least a sentence that this was part of a large survey of insects is needed. And one sentence for its possible value.

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. It is short and too the point, gives some possible uses, and is very appropriate to the article type.

Results. Table 1 is not really needed. They are not details that the reader needs to under the work and they can all be found under the NCBI BioProject ID or are repeated from the methods paragraph.

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 paper should not be indexed until this is updated. The new version has been for 6 months.

Figure 4. Fig 4 has very little information beyond what is found elsewhere in the manuscript and can easily be summarized as a sentence in the results.

Figure 3 & 5. Suggest some assembly dead-ends. But overall it seems very reasonable.

No genome annotation limits the use of the resource, but is not strictly needed.

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, genetics, genomics, epigenetics, behavior, 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 30 September 2025

https://doi.org/10.21956/wellcomeopenres.27201.r131299

© 2025 Hilliou F. 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.

?

Frederique Hilliou
SPE, Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Provence-Alpes-Cote d'Azur Site de Sophia (Ringgold ID: 60158), Sophia Antipolis, Provence-Alpes-Côte d'Azur, France

The publication of the genome of *Nemophora metallica* presents the life cycle and the host plant of this Adelidae in Finland, a country where this species is classified as endangered. *N metallica* like other members of Adelidae family has developed larval case-building behaviour. A chromosome-level genome containing 2 haplotypes was obtained from a male adult DNA extraction. It is scaffolded into 31 chromosomes including the Z-sex chromosome with an estimated haploid size of 708.66Mb. The manuscript also provides Merian elements on each chromosome which could be useful for comparative genomic studies within Lepidoptera. The used of the 32 Merian-elements could be proposed as a standard procedure in lepidopteran genome chromosome numbering. For instance, two Merian elements (M17 and M20) were fused on molecule 3 (OZ180157.1) as found in all Ditrysian (Wright et al 2024). (Ref 1)

Tools are being developed to use proteins/CDS/transcript/RNAseq from closed species to obtain genome annotation. Unfortunately, no annotation is yet provided with this genome of *Nemophora metallica*, which will limit its use by biologist.
Thy is the reason I choose "Approved with Reservations "

References

1. Wright C, Stevens L, Mackintosh A, Lawniczak M, et al.: Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nature Ecology & Evolution*. 2024; **8** (4): 777-790
Publisher Full Text

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

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

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: plant insect interactions; insect detoxification

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 24 September 2025

https://doi.org/10.21956/wellcomeopenres.27201.r130134

© 2025 Hof A. 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.

Arjen Van 't Hof
Biology Centre of the Czech Academy of Sciences, Budějovice, Czech Republic

The genome assembly Brassy Long-horn is of outstanding quality, which we have become accustomed to. It is a bit challenging to come up with original comments after having done several of these genome assembly reviews. The quality is always close to perfect, and the methods seem so streamlined by now that it is hard to find any issues to comment on. Last week there was a very interesting article about dancing spiders on the BBC website, which ended by disclosing the

aim to decode 10,000 genomes next year and complete all 1.8 million living species in the next decade. This would mean that, in case every genome is assessed by 2 reviewers, the editor would need to send out at least 20,000 review requests next year, and 3.6 million in the next decade. It could be that finding reviewers will become more challenging than producing genome assemblies at some point.
The chromosome numbers are assigned based on Merian elements representing 32 ancestral chromosomes. I have not seen this in previous Lepidoptera genome assembly notes, which was surprising. I hope that standardizing the chromosome numbers based on ancestry as is done here for the Brassy Long-horn will become the norm for future assemblies.
One minor comment:
I assume the Nanodrop spectrophotometer was not used to assess the concentration of the sheared and purified DNA, but instead to establish the purity of the samples based on the 260/280 and 260/230 ratios.

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: Lepidoptera 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.