



DATA NOTE

The genome sequence of a picture-winged fly, *Melieria picta* (Meigen, 1826) (Diptera: Ulidiidae)

[version 1; peer review: 2 approved]

Ryan Mitchell¹, 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, Sligo Town, Sligo, Ireland

V1 First published: 15 Sep 2025, 10:515
<https://doi.org/10.12688/wellcomeopenres.24887.1>
Latest published: 15 Sep 2025, 10:515
<https://doi.org/10.12688/wellcomeopenres.24887.1>

Abstract

We present a genome assembly from an individual male *Melieria picta* (picture-winged fly; Arthropoda; Insecta; Diptera; Ulidiidae). The assembly contains two haplotypes with total lengths of 423.82 megabases and 429.48 megabases. Most of haplotype 1 (98.49%) is scaffolded into 4 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 15.78 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

Melieria picta; picture-winged fly; genome sequence; chromosomal; Diptera



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

Open Peer Review

Approval Status

	1	2
version 1		
15 Sep 2025	view	view

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

2. **Teodora Staykova**, Plovdiv University, Plovdiv, Bulgaria

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: Mitchell R: 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 Mitchell R *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mitchell R, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of a picture-winged fly, *Melieria picta* (Meigen, 1826) (Diptera: Ulidiidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:515 <https://doi.org/10.12688/wellcomeopenres.24887.1>

First published: 15 Sep 2025, 10:515 <https://doi.org/10.12688/wellcomeopenres.24887.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Eremoneura; Cyclorrhapha; Schizophora; Acalyptratae; Tephritoidea; Ulidiidae; Otitinae; *Melieria*; *Melieria picta* (Meigen, 1826) (NCBI:txid3037405)

Background

Melieria is a genus of picture-winged flies in the family Ulidiidae. Four species occur in north-western Europe, including *Melieria picta* (Meigen, 1826), a small species (body length 4–5 mm) distinguished by particularly large wing spots – formed by the fusion of two apical spots common in other *Melieria* species – and by dark bands along the posterior margins of the tergites. The only other British *Melieria* species with a banded abdomen is the larger *M. crassipennis*, which differs in having separate apical spots and an inland distribution (Falk, 2025).

In Britain, *M. picta* is a coastal species with a distribution concentrated in south and east England – particularly the Solent, Thames Estuary, Wash, and Humber Estuary – with scattered western records from Somerset to Cumbria. It appears to be more thermophilic than *M. cana*, a closely related species that shares similar habitats and may co-occur with *M. picta* (Falk, 2025). The NBN Atlas also shows a predominantly coastal distribution in southern England and Wales (NBN Atlas Partnership, 2025).

Melieria picta typically inhabits upper saltmarsh and brackish coastal marsh, particularly at the strandline and in the zones dominated by Saltmarsh Rush (*Juncus gerardii*) and Saltmarsh Grass (*Puccinellia maritima*). It is also associated with brackish ditches in coastal grazing marshes, often where Sea Clubrush (*Bolboschoenus maritimus*) is present (Falk, 2025).

The species has been recorded in large numbers on *Elytrigia atherica* (sea couch) in Dutch coastal marshes, suggesting a close association with this plant. Although the larval biology is unconfirmed, observations of mating pairs and other adult behaviours exclusively on this plant suggest it may serve as the larval host (Smit, 2010). Earlier accounts suggested an association with *Artemisia maritima* (sea wormwood), but this has not been supported by field observations (Kabos & Van Aartsen, 1984; Smit, 2010).

Adults are typically observed from May to August (Falk, 2025). According to GBIF, all known occurrence records with coordinates are from Europe, with the majority from the United Kingdom (203) and the Netherlands (65), and smaller numbers from France, Germany, Ukraine, Austria, Hungary, Belgium and Denmark (GBIF Secretariat, 2025, accessed August 2025).

Only four genomes are available for the family Ulidiidae. The present assembly is one of three genomes for the genus *Melieria* as of August 2025 (data obtained via NCBI datasets, O'Leary *et al.*, 2024). We present a chromosome-level genome sequence for *Melieria picta*, produced using the Tree

of Life pipeline from a specimen collected in Morston Creek, Blakeney Estuary, England, United Kingdom (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Melieria picta* (specimen ID NHMUK015059604, ToLID idMelPict1; Figure 1), collected from Morston Creek, Blakeney Estuary, England, United Kingdom (latitude 52.96, longitude −0.98) on 2022-07-07. The specimen was collected and identified by Ryan Mitchell. For the Darwin Tree of Life sampling and metadata approach, refer to 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

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 idMelPict1 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. We used centrifuge-mediated fragmentation to produce DNA fragments



Figure 1. Photograph of the *Melieria picta* (idMelPict1) specimen used for genome sequencing.

in the 8–10 kb range, following the [Covaris g-TUBE](#) protocol for ultra-low input (ULI). 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.

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 whole organism tissue of the idMelPict1 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 6000.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mercounts (*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 MERQUERY.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 30 breaks, 96 joins, and removal of 43 haplotypic duplications. The curation process is documented 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 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 *Melieria picta* specimen generated 28.73 Gb (gigabases) from 2.60 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 360.71 Mb, with a heterozygosity of 1.25% and repeat content of 29.29% ([Figure 2](#)). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 74× coverage. Hi-C sequencing produced 117.23 Gb from 776.38 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 423.82 Mb in 142 scaffolds, with 487 gaps, and a scaffold N50 of 134.71 Mb ([Table 2](#)).

Most of the assembly sequence (98.49%) was assigned to 4 chromosomal-level scaffolds, representing 3 autosomes and the X sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size

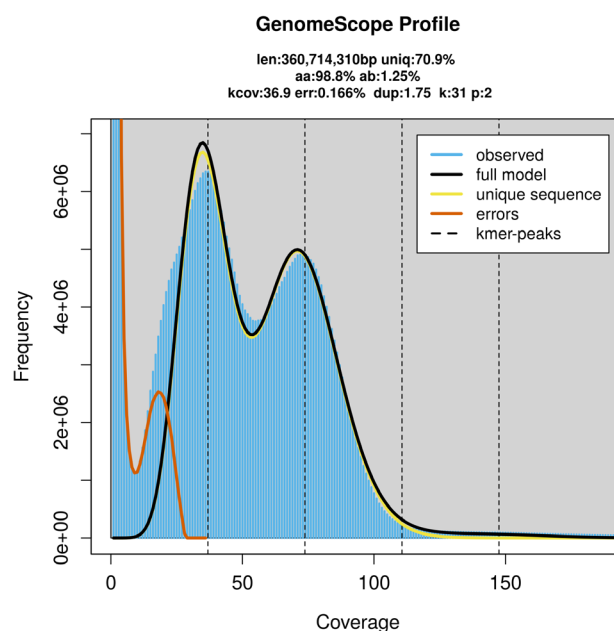


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB73950.

Platform	PacBio HiFi	Hi-C
ToLID	idMelPict1	idMelPict1
Specimen ID	NHМУK015059604	NHМУK015059604
BioSample (source individual)	SAMEA112964202	SAMEA112964202
BioSample (tissue)	SAMEA112975363	SAMEA112975363
Tissue	whole organism	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000
Run accessions	ERR12760835	ERR12765206
Read count total	2.60 million	776.38 million
Base count total	28.73 Gb	117.23 Gb

Table 2. Genome assembly statistics.

Assembly name	idMelPict1.hap1.1	idMelPict1.hap2.1
Assembly accession	GCA_964056695.1	GCA_964056685.1
Assembly level	chromosome	scaffold
Span (Mb)	423.82	429.48
Number of chromosomes	4	N/A
Number of contigs	629	566
Contig N50	1.58 Mb	1.55 Mb
Number of scaffolds	142	141
Scaffold N50	134.71 Mb	132.26 Mb
Longest scaffold length (Mb)	146.35	N/A
Sex chromosomes	X	N/A
Organelles	Mitochondrion: 15.78 kb	N/A

(Figure 3; Table 3). The X chromosome was identified by read coverage, but no Y chromosome could be found – the specimen is likely to be an XO male.

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.

For haplotype 1, the estimated QV is 60.7, and for haplotype 2, 60.8. When the two haplotypes are combined, the assembly achieves an estimated QV of 60.8. The *k*-mer completeness is 81.11% for haplotype 1, 81.50% for haplotype 2, and 98.63% for the combined haplotypes (Figure 4).

BUSCO analysis using the diptera_odb10 reference set (*n* = 3 285) identified 98.6% of the expected gene set (single = 98.3%, duplicated = 0.3%) 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.Q60**, meeting the recommended reference standard.

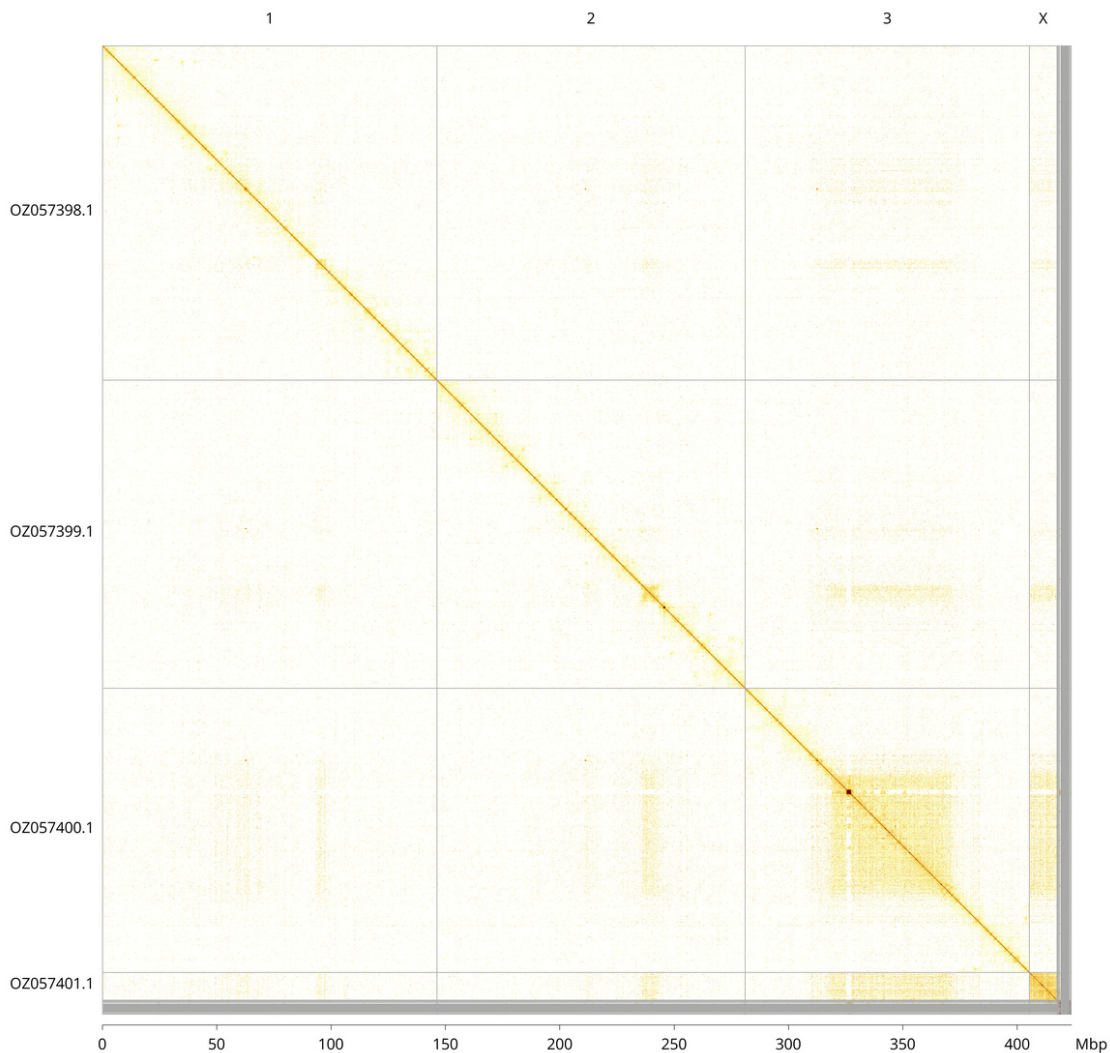


Figure 3. Hi-C contact map of the *Melieria picta* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Melieria picta* idMelPict1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ057398.1	1	146.35	38
OZ057399.1	2	134.71	37.50
OZ057400.1	3	124.28	38.50
OZ057401.1	X	12.10	39.50

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

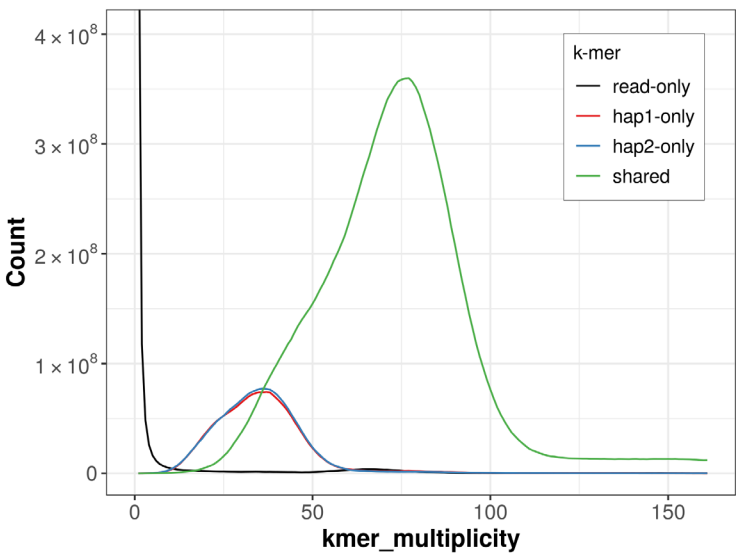


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

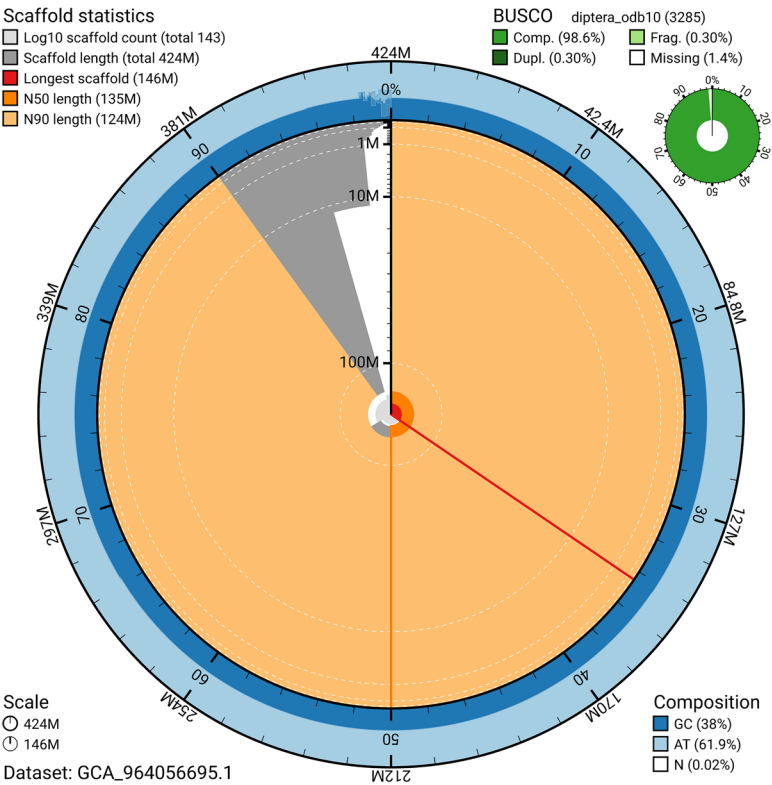


Figure 5. Assembly metrics for idMelPict1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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](#).

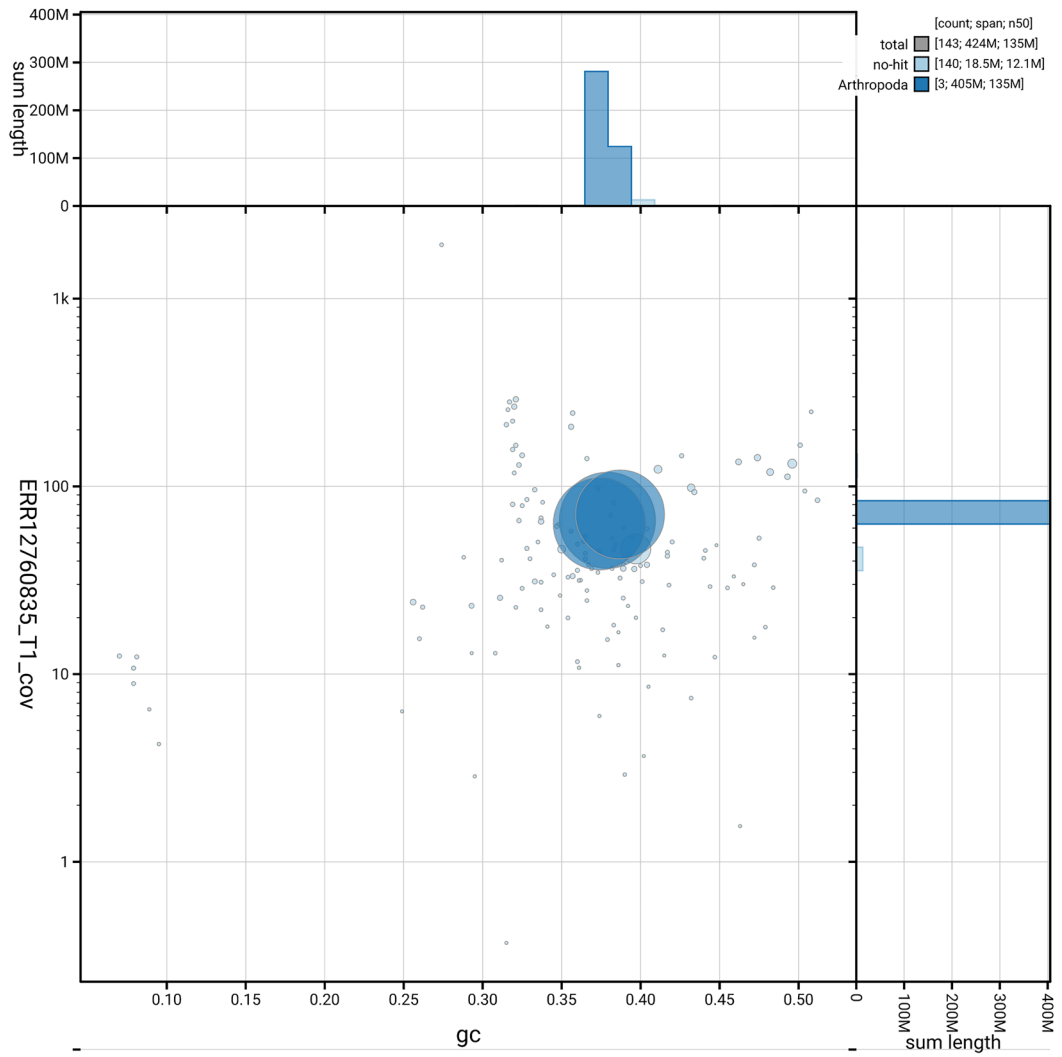


Figure 6. BlobToolKit GC-coverage plot for idMelPict1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Meliera picta* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q60	6.C.Q40
Contig N50 length	1.58 Mb	≥ 1 Mb
Scaffold N50 length	134.71 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 60.7; haplotype 2: 60.8; combined: 60.8	≥ 40
k-mer completeness	Haplotype 1: 81.11%; Haplotype 2: 81.50%; combined: 98.63%	≥ 95%
BUSCO	C:98.6% [S:98.3%; D:0.3%]; F:0.3%; M:1.1%; n:3 285	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.49%	≥ 90%

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: *Melieria picta*. Accession number [PRJEB73950](#). The genome sequence is released openly for reuse. The *Melieria picta* 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/chhyllp123/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

Software	Version	Source
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
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

- Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–35. [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](#)
- 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: **Meliera picta (Saltmarsh Spotwing).** 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: **GBIF occurrence data for *Meliera picta* (Meigen, 1826).** 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): gaa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kabos WJ, Van Aartsen B: **De nederlandse boorvliegen (Tephritidae) en prachtvliegen (Otitidae).** Hoogwoud, The Netherlands: Koninklijke Nederlandse Natuurhistorische Vereniging, 1984; **163**. [Reference Source](#)
- 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](#)
- Kurtz 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, Seppay 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: **Species distribution map for *Meliera picta* (Meigen, 1826).** 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.*: **Merqury: reference-free quality,**

completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020; **21**(1): 245.

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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.

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

Smit JT: **De prachtvlieg *Meliera picta* in grote aantallen op strandkweek *Elytrigia atherica* op schiermonnikoog (Diptera: Ulidiidae).** *Nederlandse Faunistische Mededelingen.* 2010; **33**: 1–8.

[Reference Source](#)

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.

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

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

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

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.

[Publisher Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.

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

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 25 October 2025

<https://doi.org/10.21956/wellcomeopenres.27414.r134654>

© 2025 Staykova T. 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.



Teodora Staykova

Plovdiv University, Plovdiv, Bulgaria

The genome of a picture-winged fly *Melieria picta* (Diptera: Ulidiidae) was studied based on a male specimen collected from Morston Creek, Blakeney Estuary, England, United Kingdom. The methods used and statistical processing of the data are adequate to the purpose of the study. Results were obtained and are presented in 6 figures and 4 tables. Two haplotypes were identified. The haplotype 1 is scaffolded in 4 chromosomal pseudomolecules, of which 3 are autosomes and one X sex chromosome. No second sex chromosome was identified during the study, leading to the conclusion that male heterogamety of the XO type is observed in this species. The presented study is interesting and significant, as it fills in the missing information regarding the genome of this species. This study could serve as a basis for future population genetic studies within the species.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetics, Population Genetics

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 19 September 2025

<https://doi.org/10.21956/wellcomeopenres.27414.r133375>

© 2025 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 well-documented summary of a novel genomic assembly for a male *Meliera picta*. The methods used are appropriate and the statistics and summaries for the final genome are thorough. The quality of the genome and writing are both high, measures were taken during all steps of this experiment to generate a well curated final product. Documentation of all of these measures can be found in the protocols.io and githubs, which are well linked and easily accessible, with the exception of the link to "triaged" leading to the **Sanger Tree of Life Sample Preparation: Triage and Dissection**. The attachments in this protocol were not accessible at the time of this review for those without a protocols.io account. A similar protocol can be accessed through the Tree of Life Core Laboratory; however the results are significantly less detailed for decision making based on unique tissue types.

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, short read sequencing, biogeography and evolution

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.