



DATA NOTE

The genome sequence of the Marbled Brown, *Drymonia dodonaea* (Denis & Schiffermüller), 1775 (Lepidoptera: Notodontidae)

[version 1; peer review: 3 approved]

Denise C. Wawman ¹,

University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,

Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹Department of Biology, University of Oxford, Oxford, England, UK

V1 First published: 09 Sep 2025, 10:498
<https://doi.org/10.12688/wellcomeopenres.24797.1>
Latest published: 09 Sep 2025, 10:498
<https://doi.org/10.12688/wellcomeopenres.24797.1>

Abstract

We present a genome assembly from an individual *Drymonia dodonaea* (Marbled Brown; Arthropoda; Insecta; Lepidoptera; Notodontidae). The assembly contains two haplotypes with total lengths of 330.51 megabases and 331.17 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 16.37 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

Drymonia dodonaea; Marbled Brown; genome sequence; chromosomal; Lepidoptera

Open Peer Review

Approval Status

	1	2	3
version 1			
09 Sep 2025	view	view	view

1. **Alexandra Cerqueira De Araujo** ,
University College London, London, UK
2. **Kashif Zahoor** , Government College
University Faisalabad, Faisalabad, Pakistan
3. **Sarah Inwood** , University of Otago,
Dunedin, New Zealand

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



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

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Wawman DC: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

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 Wawman DC *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: Wawman DC, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of the Marbled Brown, *Drymonia dodonaea* (Denis & Schiffermüller), 1775 (Lepidoptera: Notodontidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, 10:498 <https://doi.org/10.12688/wellcomeopenres.24797.1>

First published: 09 Sep 2025, 10:498 <https://doi.org/10.12688/wellcomeopenres.24797.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Notodontidae; Notodontinae; *Drymonia*; *Drymonia dodonaea* (Denis & Schiffermüller, 1775 (NCBI:txid988176))

Background

The Marbled Brown, *Drymonia dodonaea* is a moth in the family Notodontidae. As its common name suggests, it is brown, with patterning similar to that seen in marble. It has a paler central band across its 16–20mm long forewings. The similar-looking Lunar Marbled Brown *Drymonia ruficornis* (Hufnagel) has a more distinct pale central band which is usually white or whitish with a distinctive black comma-shaped marking (Waring *et al.*, 2017).

The larvae are bluish-green with a brighter green dorsal stripe that is bordered by yellow lines. There is a yellowish-green lateral stipe, edged superiorly with a faint red line, following the line of the spiracles which have red dots. Below this line the body colour is more yellowish than above it (Henwood *et al.*, 2020).

In the United Kingdom, *D. dodonaea* has a single generation on the wing in May and June and into early July (Waring *et al.*, 2017). The larva feed on both Sessile Oak *Quercus petraea* and Pedunculate Oak *Quercus robur*, from late June to early September and pupate in a cocoon in the ground to overwinter as a pupa (Henwood *et al.*, 2020).

Drymonia dodonaea can be found across much of Europe between latitudes 40°N and 60°N (GBIF Secretariat, 2023). In the UK, it can be locally common in oak woodlands as are north as western Scotland (Skinner & Wilson, 2009; Waring *et al.*, 2017), although like UK many moths, the population of Marbled Brown is decreasing (Conrad *et al.*, 2006).

We present a chromosomally complete genome sequence for *Drymonia dodonaea*, the Marbled Brown. The specimen used for sequencing was attracted to light from a mercury vapour lamp, in a rural garden, in the hamlet of Bratton, near Minehead in Somerset.

The assembly was produced using the Tree of Life pipeline from a specimen collected in Bratton, Somerset, United Kingdom (Figure 1), as part of the Darwin Tree of Life project.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult *Drymonia dodonaea* (specimen ID Ox003328, ToLID iLDryDodo1; Figure 1), collected from Bratton, Somerset, United Kingdom (latitude 51.205, longitude -3.509) on 2022-06-03. The specimen was collected and identified by Denise Wawman



Figure 1. Photograph of the *Drymonia dodonaea* (iLDryDodo1) specimen from which samples were taken for genome sequencing.

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

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Nucleic acid extraction

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 iLDryDodo1 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 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 41.19 ng/μL and a yield of 1 935.93 ng, with a fragment size of 16.1 kb. The 260/280 spectrophotometric ratio was 2, and the 260/230 ratio was 2.92.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the head and thorax of the iIDryDodo1 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 to 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.

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. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). 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 MERQUY.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). 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 Merquy.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 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 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

PacBio sequencing of the *Drymonia dodonaea* specimen generated 25.37 Gb (gigabases) from 2.61 million reads, which were used to assemble the genome. GenomeScope2.0 analysis

estimated the haploid genome size at 331.01 Mb, with a heterozygosity of 0.15% and repeat content of 25.72% ([Figure 2](#)). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 75× coverage. Hi-C sequencing produced 151.63 Gb from 1 004.20 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 330.51 Mb in 46 scaffolds, with gaps, and a scaffold N50 of 12.0 Mb ([Table 2](#)).

Most of the assembly sequence (99.89%) was assigned to 31 chromosomal-level scaffolds, representing autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to ([Figure 3](#); [Table 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.

For haplotype 1, the estimated QV is 62.2, and for haplotype 2, 62.1. When the two haplotypes are combined, the assembly

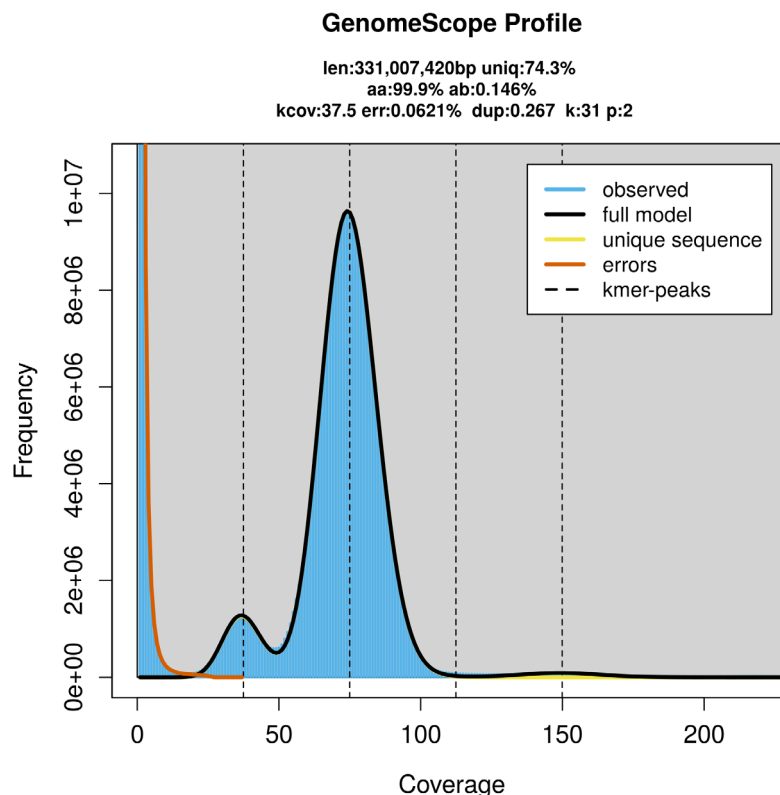


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB85001.

Platform	PacBio HiFi	Hi-C
ToLID	ilDryDodo1	ilDryDodo1
Specimen ID	Ox003328	Ox003328
BioSample (source individual)	SAMEA114644434	SAMEA114644434
BioSample (tissue)	SAMEA114645038	SAMEA114645038
Tissue	head and thorax	head and thorax
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR14209121; ERR14209122	ERR14224608
Read count total	2.61 million	1 004.20 million
Base count total	25.37 Gb	151.63 Gb

Table 2. Genome assembly statistics.

Assembly name	ilDryDodo1.hap1.1	ilDryDodo1.hap2.1
Assembly accession	GCA_965178025.1	GCA_965177995.1
Assembly level	chromosome	scaffold
Span (Mb)	330.51	331.17
Number of chromosomes	31	-
Number of contigs	115	105
Contig N50	6.16 Mb	5.84 Mb
Number of scaffolds	46	49
Scaffold N50	12.0 Mb	12.1 Mb
Longest scaffold length (Mb)	14.06	-
Sex chromosomes	Z	-
Organelles	Mitochondrion: 16.37 kb	-

achieves an estimated QV of 62.1. The *k*-mer completeness is 95.91% for haplotype 1, 95.91% for haplotype 2, and 99.73% for the combined haplotypes (Figure 4).

BUSCO analysis using the lepidoptera_odb10 reference set (*n* = 5 286) identified 98.7% of the expected gene set (single = 98.3%, duplicated = 0.4%) 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.Q62**, 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

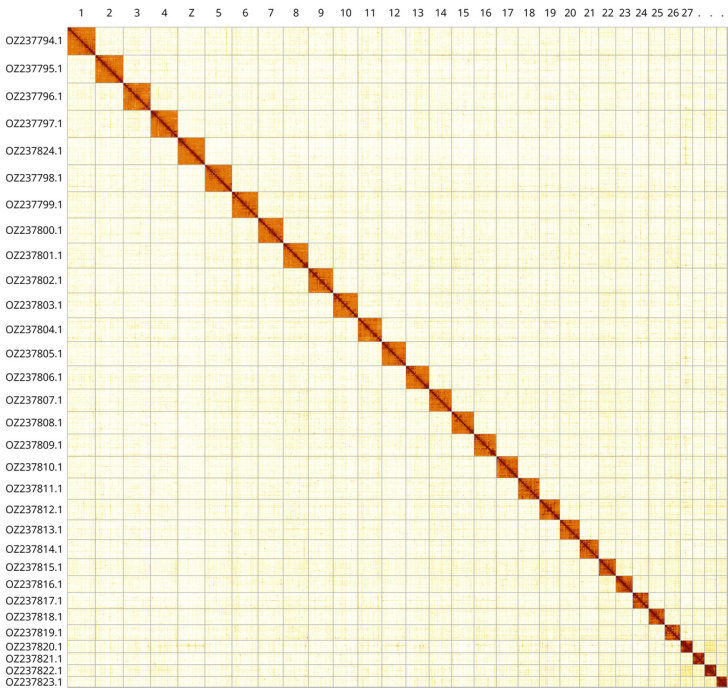


Figure 3. Hi-C contact map of the *Drymonia dodonaea* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Drymonia dodonaea* iIDryDodo1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ237794.1	1	14.06	39
OZ237795.1	2	13.94	39.50
OZ237796.1	3	13.73	39.50
OZ237797.1	4	13.61	39.50
OZ237798.1	5	13.53	38.50
OZ237799.1	6	13.03	39
OZ237800.1	7	12.63	39
OZ237801.1	8	12.53	38.50
OZ237802.1	9	12.38	38.50
OZ237803.1	10	12.37	39
OZ237804.1	11	12.03	39
OZ237805.1	12	12	39.50
OZ237806.1	13	11.64	39.50
OZ237807.1	14	11.26	39

INSDC accession	Molecule	Length (Mb)	GC%
OZ237808.1	15	11.23	39
OZ237809.1	16	11.14	39
OZ237810.1	17	10.87	40
OZ237811.1	18	10.67	40
OZ237812.1	19	10.27	40.50
OZ237813.1	20	9.84	39
OZ237814.1	21	9.60	39.50
OZ237815.1	22	8.50	40.50
OZ237816.1	23	8.45	40
OZ237817.1	24	8.04	40.50
OZ237818.1	25	7.94	41
OZ237819.1	26	7.92	40
OZ237820.1	27	6.16	42.50
OZ237821.1	28	5.92	41
OZ237822.1	29	5.90	41
OZ237823.1	30	5.40	41.50
OZ237824.1	Z	13.57	39.50

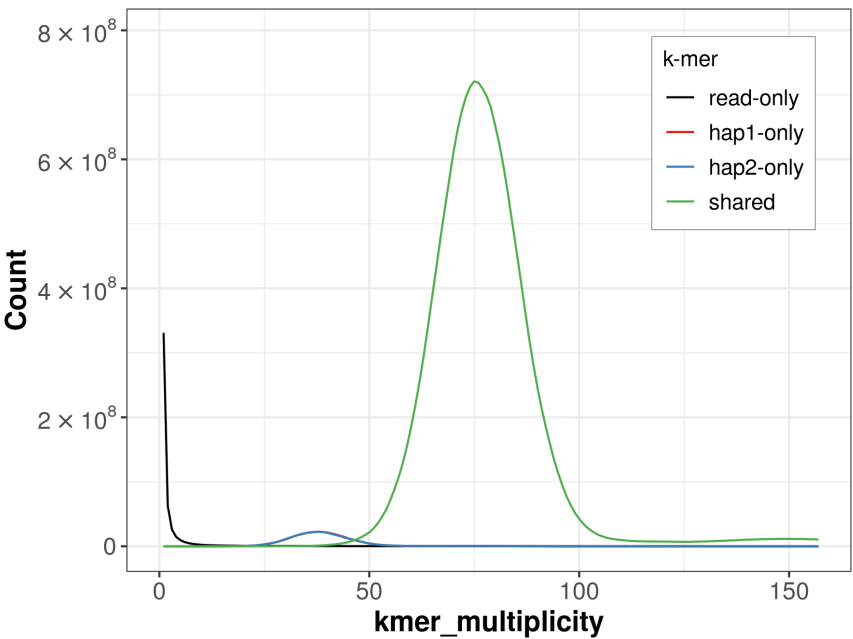


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

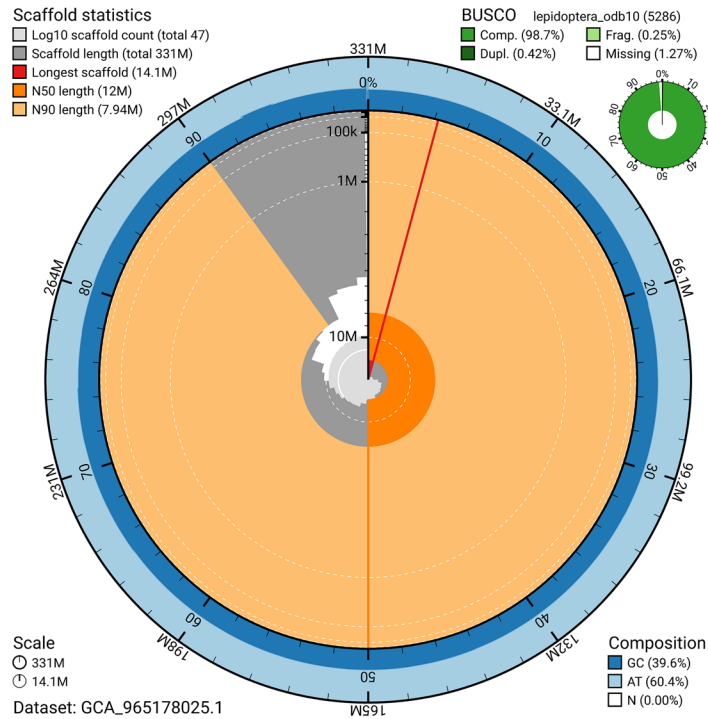


Figure 5. Assembly metrics for *ilDryDodo1.hap1.1*. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

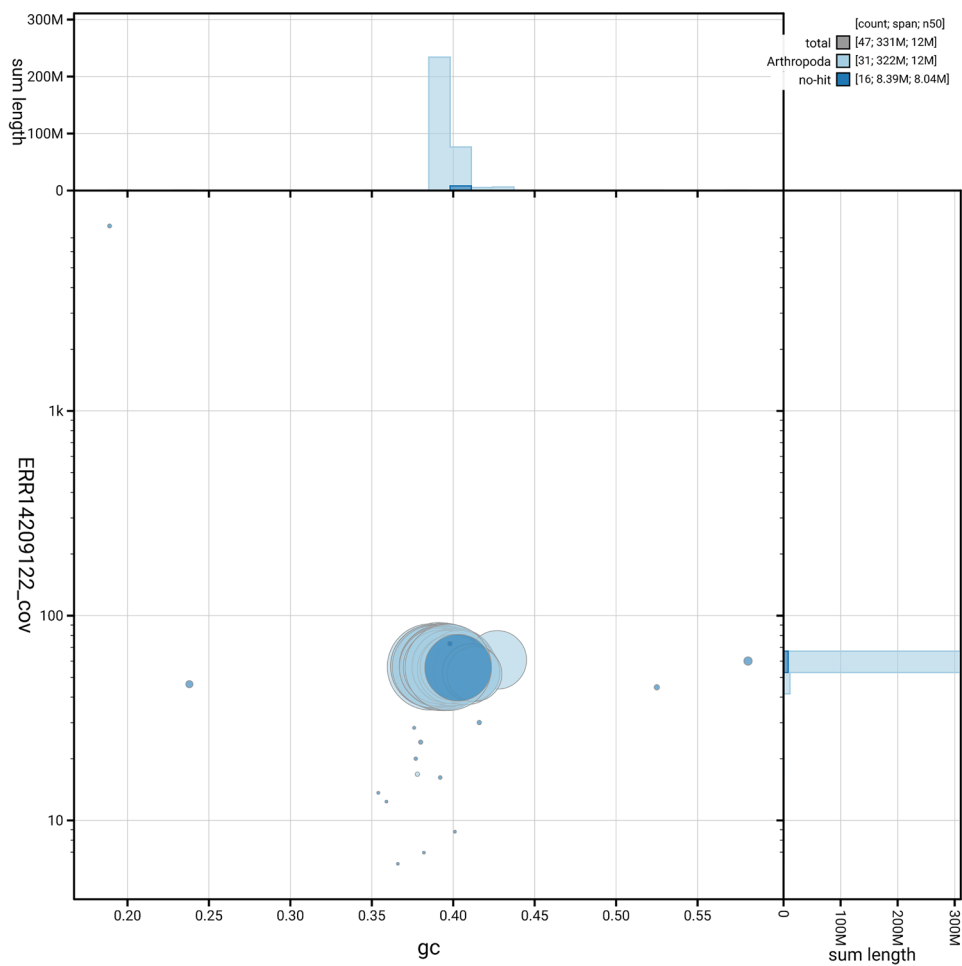


Figure 6. BlobToolKit GC-coverage plot for iLDryDodo1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Drymonia dodonaea* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q62	6.C.Q40
Contig N50 length	6.16 Mb	≥ 1 Mb
Scaffold N50 length	12 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 62.2; haplotype 2: 62.1; combined: 62.1	≥ 40
<i>k</i> -mer completeness	Haplotype 1: 95.91%; Haplotype 2: 95.91%; combined: 99.73%	≥ 95%
BUSCO	C:98.7% [S:98.3%; D:0.4%]; F:0.2%; M:1.0%; n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.89%	≥ 90%

out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

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

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

Data availability

European Nucleotide Archive: *Drymonia dodonaea* (marbled brown). Accession number [PRJEB85001](#). The genome sequence is released openly for reuse. The *Drymonia dodonaea* 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 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 are 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 [University of Oxford and Wytham Woods Genome Acquisition Lab](#)
- Members of the [Darwin Tree of Life Barcoding collective](#)
- Members of the [Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team](#)
- Members of [Wellcome Sanger Institute Scientific Operations – Sequencing Operations](#)
- 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	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	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
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquyFK	1.1.2	https://github.com/thegenemyers/MERQUERY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2

Software	Version	Source
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	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	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	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

References

- Alлио R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Conrad KF, Warren MS, Fox R, *et al.*: **Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis.** *Biol Conserv.* 2006; **132**(3): 279–291.
[Publisher 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](#)
- 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: ***Drymonia dodonaea* (denis & Schiffermüller) 1775.** 2023.
[Reference Source](#)
- 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](#)
- Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Henwood B, Sterling P, Lewington R: **Field guide to the caterpillars of Great Britain and Ireland.** Bloomsbury Wildlife, 2020.
[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](#)
- 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, 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](#)
- 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](#)

Skinner B, Wilson D: **Colour identification guide to moths of the British Isles.** Apollo Books, 2009.

[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, Krashennikova 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](#)

Waring P, Townsend M, Lewington R: **Field guide to the moths of Great Britain and Ireland.** Bloomsbury Publishing Plc, 2017.

[Reference Source](#)

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

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

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 16 October 2025

<https://doi.org/10.21956/wellcomeopenres.27320.r132871>

© 2025 Inwood S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sarah Inwood 

University of Otago, Dunedin, New Zealand

The authors have generated a high quality genome assembly for the Marbled Brown moth *Drymonia dodonaea*. This involved using state of the art PacBio HiFi sequencing alongside Hi-C sequencing to generate a chromosome-level scaffolded assembly. Quality control metrics for this assembly meet literature standards, with an N50 of 12 Mb, 99.89% of the assembled sequence being assigned to chromosomal-level scaffolds, above 95% k-mer completeness, and a high 98.7% of lepidopteran BUSCO orthologs being found within the assembly. Alongside this a mitochondrial genome was also assembled, with a length of 16.37 kb.

Minor comments:

- Length of the assembled mitochondrial genome could be stated in the results of the paper where it is reported that the mitochondrial genome was also assembled
- Hifiasm, purge_dups, YaHS, and MitoHiFi missing from software table, therefore version number used is not stated
- Conversely the table includes BEDTools, Cooler, fasta_windows and GoaT CLI which are never mentioned in text - were these tools used for this assembly and if so, what were they used for?

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?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, host-parasite interactions, parasitoid wasps, insects, viruses, biocontrol

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 16 October 2025

<https://doi.org/10.21956/wellcomeopenres.27320.r132873>

© 2025 Zahoor K. 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.



Kashif Zahoor 

Government College University Faisalabad, Faisalabad, Punjab, Pakistan

The article titled "The genome sequence of the Marbled Brown, *Drymonia dodonaea* (Denis & Schiffermüller), 1775 (Lepidoptera: Notodontidae)" describes genome of Marbled Brown Moth, *Drymonia dodonaea*. Its appreciable that mitochondrial genome has also been assembled. Its very well performed scientific work, well written, very well explained and beautifully presented with Figures and Tables. The work shows huge work rather I should say hard work. Authors covered all the aspects that there's no such room to criticize or even to make critical comments. The article fully merits to be indexed, and I strongly recommend to get this indexed in your good journal.

There's only 2-3 points I want to raise:

- Is there any possibility to get more than these data; 25.37Gb from 2.61 million reads PacBio??? and 151.63 Gb from 1 004.20 million reads through Hi-C??? It seems likely that there could be more data if one may use any other technique/tool?
- Hi-C shows more reads but less data???
- PacBio has less reads but more gigabases???
- Is there any kind of optimum value for QV?

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, Molecular Biology, Entomology, Zoology

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

Reviewer Report 27 September 2025

<https://doi.org/10.21956/wellcomeopenres.27320.r132874>

© 2025 De Araujo 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.



Alexandra Cerqueira De Araujo 

University College London, London, UK

The authors present here a chromosome-level *de novo* assembly of the genome of *Drymonia dodonaea* (Notodontidae, Lepidoptera). This is the first assembly available for this species. This assembly is composed of 31 chromosome-level pseudomolecules (including a sex chromosome Z) and one mitochondrial chromosome. The assembly has a total length of 330.51 Mb and a N50 of 12.0 Mb. This assembly achieves a EBP reference standard of 6.C.Q62 and contains most of the orthologous genes found in lepidopterans (i.e. 98.7% of the complete BUSCO sequences). The quality of this assembly meets the recommended reference standard for genome assemblies, with no identifiable major issues. All methods used to produce this data is reported in this manuscript and the code is available here <https://github.com/sanger-tol>.

Minor comments:

1- in "Genome assembly"

"Contigs were further scaffolded with Hi-C data in YaHS" is a bit ambiguous, may be consider changing "in YaHS" to "using YaHS" as it could sound like the Hi-C data used was stored in YaHS.

2- Regarding the number of chromosomal pseudomolecules, although other chromosome-level assemblies of species of the same genus have between 29 to 31 chromosomes, is there any karyotype data supporting it? (this question does not invalidate the results found by the authors, this is pure curiosity)

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

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.
