



DATA NOTE

The genome sequence of the Figure of Eighty moth *Tethea ocularis* Linnaeus, 1767

[version 1; peer review: 2 approved]

Douglas Boyes¹⁺, Denise C. Wawman^{id}²,
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory
team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹UK Centre for Ecology & Hydrology, Wallingford, England, UK²Edward Grey Institute, Department of Biology, University of Oxford, Oxford, England, UK

+ Deceased author

V1 First published: 28 Jun 2024, 9:348
<https://doi.org/10.12688/wellcomeopenres.21493.1>
Latest published: 28 Jun 2024, 9:348
<https://doi.org/10.12688/wellcomeopenres.21493.1>

Abstract

We present a genome assembly from an individual male *Tethea ocularis* (the Figure of Eighty; Arthropoda; Insecta; Lepidoptera; Drepanidae). The genome sequence is 339.1 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.28 kilobases in length

Keywords

Tethea ocularis, Figure of Eighty moth, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 28 Jun 2024	✓ view	✓ view

1. **Antonio S Ortiz** ^{id}, University of Murcia, Murcia, Spain
2. **Panagiotis Ioannidis** ^{id}, Foundation for Research & Technology - Hellas, Heraklion, Greece

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: **Boyes D:** Investigation, Resources; **Wawman DC:** 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 [206194, <https://doi.org/10.35802/206194>] 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: © 2024 Boyes D *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: Boyes D, Wawman DC, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Figure of Eighty moth *Tethea ocellaris* Linnaeus, 1767 [version 1; peer review: 2 approved]** Wellcome Open Research 2024, 9:348 <https://doi.org/10.12688/wellcomeopenres.21493.1>

First published: 28 Jun 2024, 9:348 <https://doi.org/10.12688/wellcomeopenres.21493.1>

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; Drepanoidea; Drepanidae; Thyatirinae; *Tethea*; *Tethea ocularis* Linnaeus, 1767(NCBI:txid997289).

Background

Tethea ocularis the Figure of Eighty is a moth in the family Drepanidae, in a group known as lutestrings that was formerly the separate family Thyatiridae. The Drepanidae look similar to the noctuid moths, although they are not as thick-bodied. *Tethea ocularis* has a brown forewing, varying in shade from light to dark brown, fine crosslines and a distinctive white “80” in the centre of the forewing; it is superficially similar to the Poplar Lutestring *Tethea or* and the noctuid Figure of Eight *Diloba caeruleocephala* (Waring *et al.*, 2017). The darker melanic form of *Tethea ocularis* *f. fusca* was first recorded in the south and east of England in the 1940’s and has slowly spread across its UK range to become the dominant form (Skinner, 2009).

The current UK range includes most of south-eastern Scotland, and all of England and Wales, with the exception of Cumberland (Waring *et al.*, 2017), although it was first recorded in neighbouring Westmoreland in 1983 (Briggs, 1984). It is not present in Ireland but reached the Isle of Man in 2002 (Skinner, 2009). *Tethea ocularis* is a species of broad-leaved woodland, hedgerows, parks and gardens, where it is on the wing at night from May to July (Waring *et al.*, 2017). The larvae feed on aspen *Populus tremula* and other poplars *Populus* spp., in a cocoon between the leaves, from mid-July to September. *Tethea ocularis* overwinters as a pupa (Waring *et al.*, 2017).

We present a chromosomally complete genome sequence for an adult male *Tethea ocularis*, based on one specimen collected in Wytham Woods, Oxfordshire, as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from a male *Tethea ocularis* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 223-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 10 missing joins or mis-joins, increasing the scaffold number by 1.69%.

The final assembly has a total length of 339.1 Mb in 59 sequence scaffolds with a scaffold N50 of 12.1 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds



Figure 1. Photograph of the *Tethea ocularis* (ilTetOcul2) specimen used for genome sequencing.

assigned to different phyla. Most (99.6%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). The Z sex chromosome was assigned by synteny to *Tetheella fluctuosa* (GCA_951216915.1). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 63.7 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.8% (single = 98.4%, duplicated = 0.3%), using the lepidoptera_odb10 reference set ($n = 5,286$).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at <https://links.tol.sanger.ac.uk/species/997289>.

Methods

Sample acquisition and nucleic acid extraction

A male *Tethea ocularis* (specimen ID Ox001919, ToLID ilTetOcul2) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2021-06-16. The specimen used for Hi-C and RNA sequencing (specimen ID Ox000465, ToLID ilTetOcul1) was collected from the same location on 2020-06-13. Both specimens were collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of

Table 1. Genome data for *Tethea ocularis*, iTetOcul2.1.

Project accession data		
Assembly identifier	iTetOcul2.1	
Species	<i>Tethea ocularis</i>	
Specimen	iTetOcul2	
NCBI taxonomy ID	997289	
BioProject	PRJEB66023	
BioSample ID	SAMEA10979182	
Isolate information	iTetOcul2, male: head and thorax (DNA sequencing) iTetOcul1: head and thorax (Hi-C sequencing), abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	63.7	≥ 50
k-mer completeness	100.0%	≥ 95%
BUSCO**	C:98.8%[S:98.4%,D:0.3%], F:0.3%,M:0.9%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.6%	≥ 95%
Sex chromosomes	ZZ	localised homologous pairs
Organelles	Mitochondrial genome: 15.28 kb	complete single alleles
Raw data accessions		
PacificBiosciences Revio	ERR12055563	
Hi-C Illumina	ERR12071233	
PolyA RNA-Seq Illumina	ERR12071234	
Genome assembly		
Assembly accession	GCA_963555595.1	
Accession of alternate haplotype	GCA_963555585.1	
Span (Mb)	339.1	
Number of contigs	79	
Contig N50 length (Mb)	10.0	
Number of scaffolds	59	
Scaffold N50 length (Mb)	12.1	
Longest scaffold (Mb)	15.95	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from [Rhie et al. \(2021\)](#).

** BUSCO scores based on the lepidoptera_odb10 BUSCO set using version 5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at <https://blobtoolkit.genomehubs.org/view/CAUVWG01/dataset/CAUVWG01/busco>.

core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. The sample was prepared for extraction in the WSI Tree of Life Core Laboratory: the iTetOcul2 sample was weighed and dissected on dry ice ([Jay et al., 2023](#)), and tissue from the head and thorax

was homogenised using a PowerMasher II tissue disruptor ([Denton et al., 2023a](#)).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol

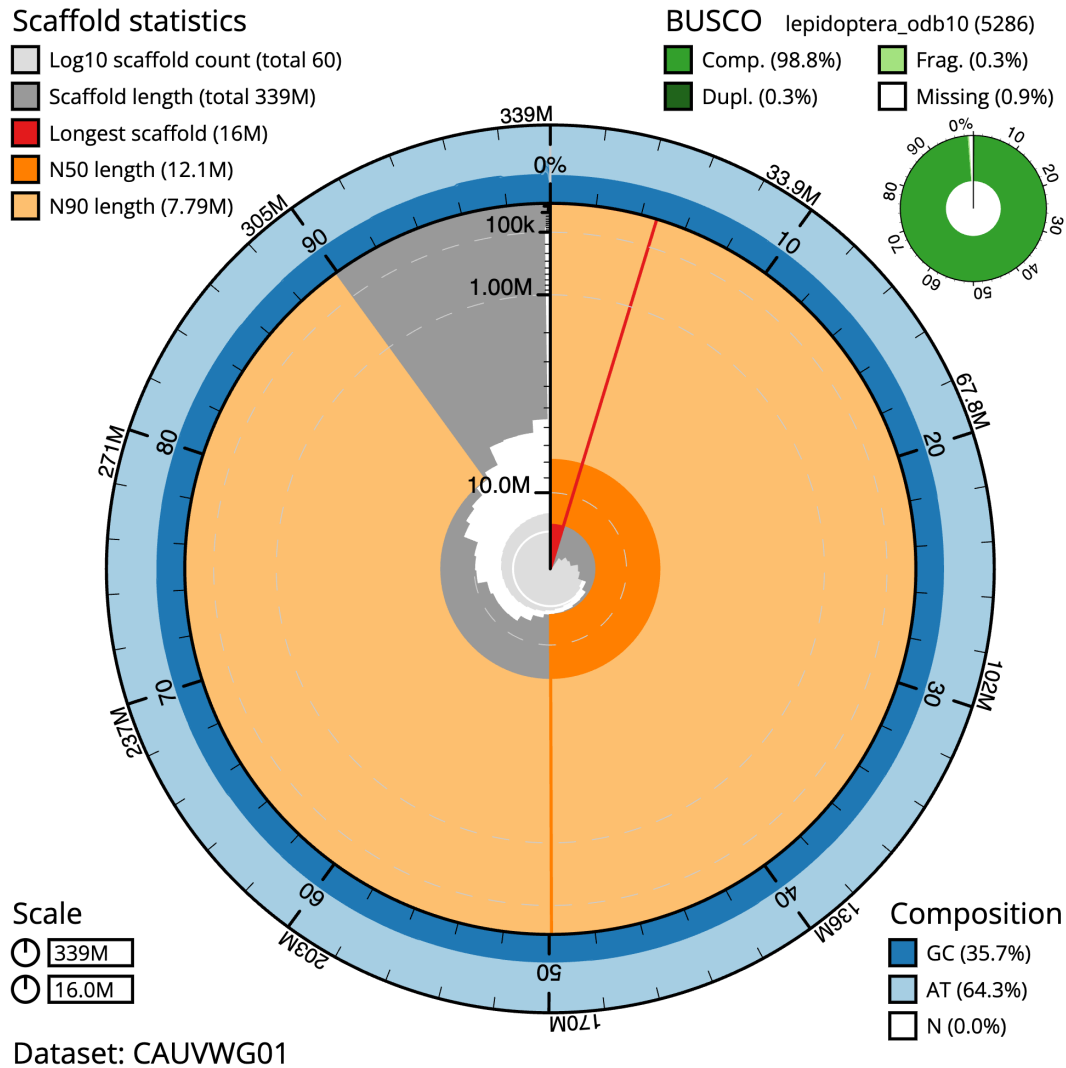


Figure 2. Genome assembly of *Tethea ocularis*, ilTetOcul2.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 339,142,399 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (15,953,683 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,138,052 and 7,792,192 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAUVWG01/dataset/CAUVWG01/snail>.

(Oatley *et al.*, 2023a). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 31 (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023a): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from abdomen tissue of ilTetOcul1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*-Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

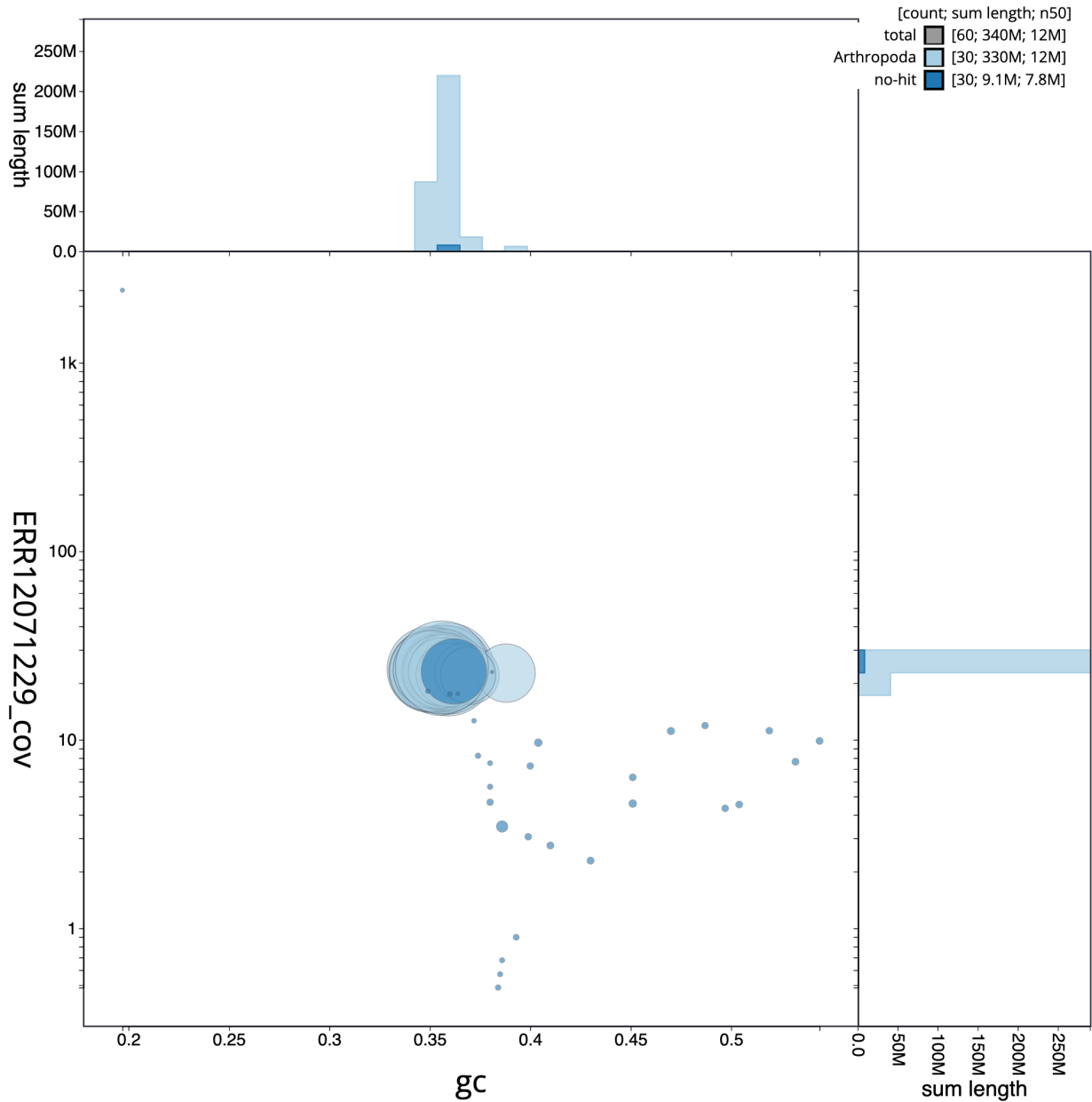


Figure 3. Genome assembly of *Tethea ocularis*, iTetOcul2.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAUVWG01/dataset/CAUVWG01/blob>.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences Revio (HiFi) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were also generated from head and thorax tissue of iTetOcul1 using the Arima2 kit and sequenced on the HiSeq X Ten instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was

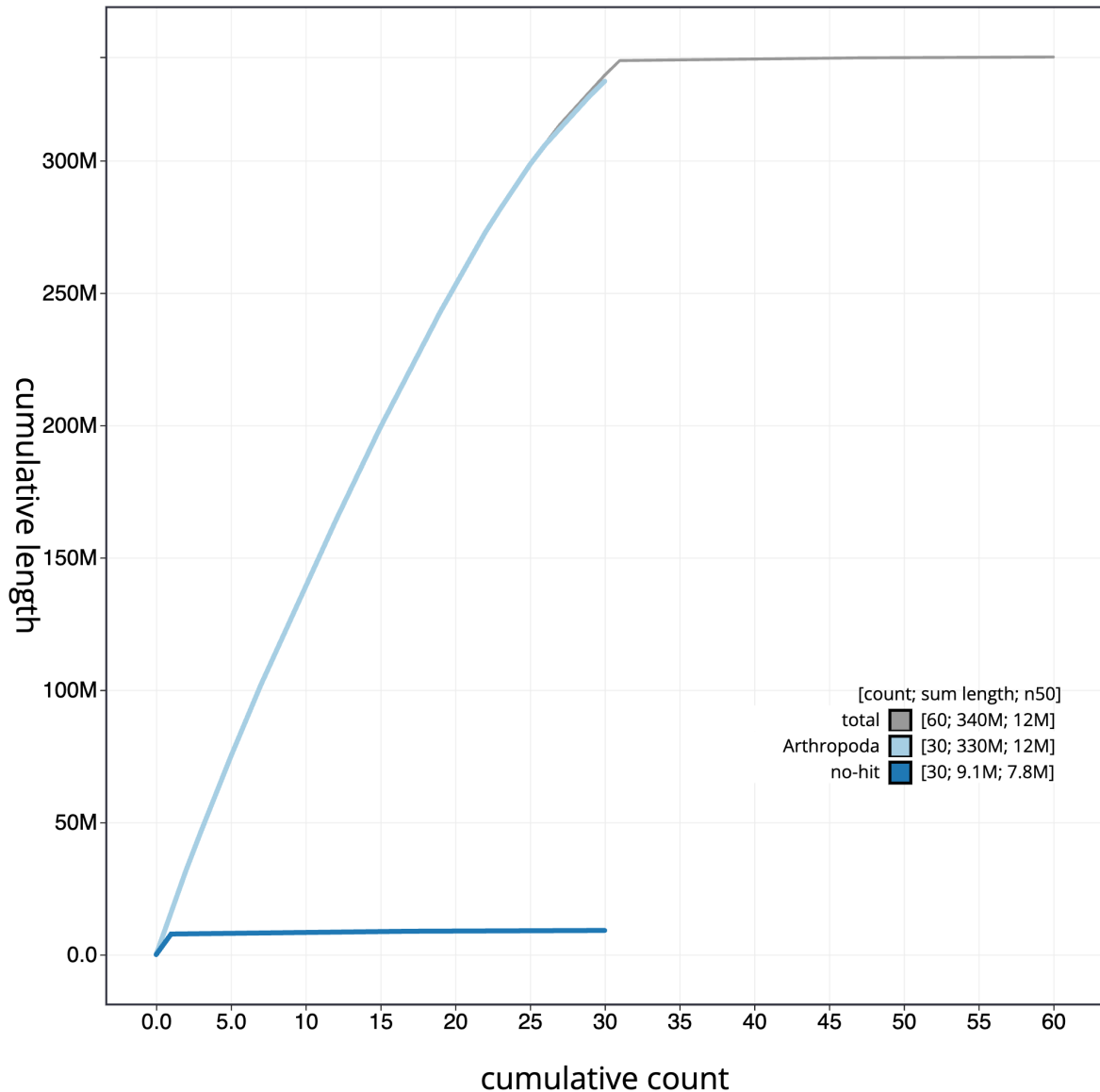


Figure 4. Genome assembly of *Tethea ocularis*, iTetOcul2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAUVWG01/dataset/CAUVWG01/cumulative>.

assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwamem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a) and

“sanger-tol/genomenote” (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

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

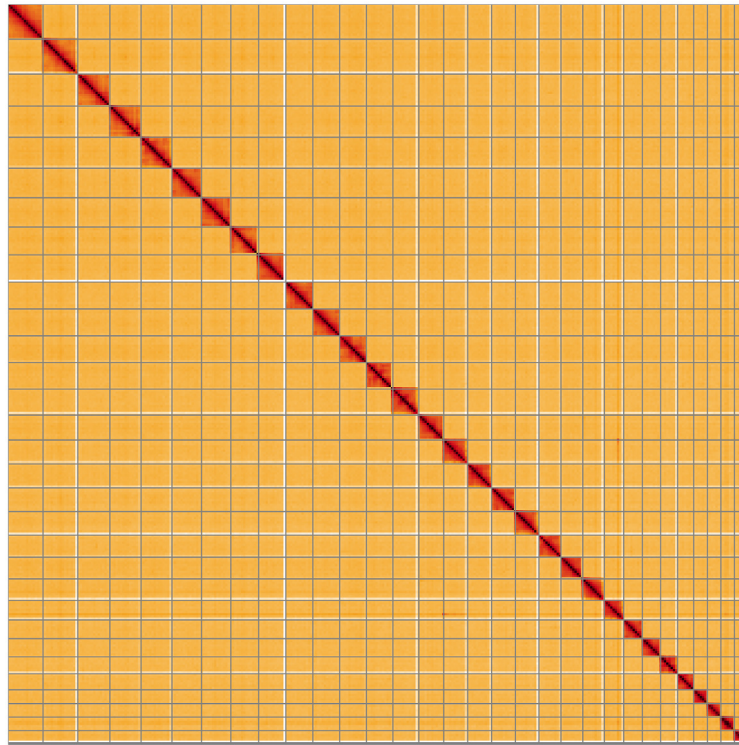


Figure 5. Genome assembly of *Tethea ocularis*, iTetOcul2.1: Hi-C contact map of the iTetOcul2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=BwYrBgisQ3-PD8AWa8mIQa>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Tethea ocularis*, iTetOcul2.

INSDC accession	Chromosome	Length (Mb)	GC%
OY741957.1	1	15.91	36.0
OY741958.1	2	14.69	35.5
OY741959.1	3	14.26	36.0
OY741960.1	4	14.13	35.5
OY741961.1	5	13.54	35.0
OY741962.1	6	13.4	35.0
OY741963.1	7	12.75	35.5
OY741964.1	8	12.41	35.5
OY741965.1	9	12.37	35.0
OY741966.1	10	12.27	35.5
OY741967.1	11	12.22	35.5
OY741968.1	12	12.14	35.0
OY741969.1	13	11.87	35.0
OY741970.1	14	11.4	35.5
OY741971.1	15	10.96	35.5

INSDC accession	Chromosome	Length (Mb)	GC%
OY741972.1	16	10.93	35.0
OY741973.1	17	10.91	36.0
OY741974.1	18	10.75	35.5
OY741975.1	19	10.08	36.0
OY741976.1	20	9.99	35.5
OY741977.1	21	9.86	36.0
OY741978.1	22	8.86	36.0
OY741979.1	23	8.55	36.0
OY741980.1	24	8.34	35.5
OY741981.1	25	7.79	36.0
OY741982.1	26	7.35	36.0
OY741983.1	27	6.25	36.5
OY741984.1	28	6.19	37.0
OY741985.1	29	6.17	39.0
OY741986.1	30	5.51	37.0
OY741956.1	Z	15.95	35.5
OY741987.1	MT	0.02	19.5

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	1.2a.2	https://github.com/c-zhou/yahs

the ‘**Darwin Tree of Life Project Sampling Code of Practice**’, which can be found in full on the Darwin Tree of Life website [here](#). By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

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

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

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

Data availability

European Nucleotide Archive: *Tethea ocularis* (figure of eighty). Accession number PRJEB66023; <https://identifiers.org/ena.embl/PRJEB66023> (Wellcome Sanger Institute, 2023). The

genome sequence is released openly for reuse. The *Tethea ocularis* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project. 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](#).

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.7125292>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.4893703>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.10066175>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.10043364>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.10066637>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.5013541>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.4783558>.

References

- Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Allio R, Schomaker-Bastos A, Romiguié 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](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor[®]3 for LI PacBio.** *protocols.io.* 2023.
[Publisher Full Text](#)
- Bernt M, Donath A, Jühling F, et al.: **MITOS: improved de novo metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Briggs J: **Northward spread of *Tethea ocellaris* L. ssp. octogenesimea Hbn.: figure of eighty.** *The Entomologist's Record and Journal of Variation.* 1984; **96**: 81–82.
[Reference Source](#)
- 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](#)
- Chow W, Brugger K, Caccamo M, et al.: **gEVAL – a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Denton A, Oatley G, Cornwell C, et al.: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
[Publisher Full Text](#)
- Denton A, Yatsenko H, Jay J, et al.: **Sanger Tree of Life wet laboratory protocol collection v.1.** *protocols.io.* 2023b.
[Publisher Full Text](#)
- Di Tommaso P, Chatzou M, Floden EW, et al.: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
- do Amaral RJV, Bates A, Denton A, et al.: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
[Publisher 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](#)
- Harry E: **PretextView (Paired Read Texture Viewer): a desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
[Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): gaa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jay J, Yatsenko H, Narváez-Gómez JP, et al.: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
[Publisher 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](#)
- Manni M, Berkeley MR, Seppy 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](#)
- Oatley G, Denton A, Howard C, et al.: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023a.
[Publisher 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](#)
- Simão FA, Waterhouse RM, Ioannidis P, et al.: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Skinner B: **Colour identification guide to moths of the British Isles.** First edit. Stenstrup, Denmark: Apollo Books, 2009.
- Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023a.
[Publisher Full Text](#)
- Strickland M, Moll R, Cornwell C, et al.: **Sanger Tree of Life HMW DNA extraction: manual MagAttract.** *protocols.io.* 2023b; [Accessed 24 November 2023].
[Publisher Full Text](#)
- Surana P, Muffato M, Qi G: **Sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - hebridean black (1.1.0).** *Zenodo.* 2023a.
[Publisher Full Text](#)
- Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
[Publisher Full Text](#)
- Todorovic M, Sampaio F, Howard C, et al.: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor[®]3 for PacBio HiFi.** *protocols.io.* 2023.
[Publisher 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](#)
- Waring P, Townsend M, Lewington R: **Field guide to the moths of Great Britain and Ireland: third edition.** Bloomsbury Wildlife Guides, 2017.
- Wellcome Sanger Institute: **The genome sequence of the figure of eighty moth *Tethea ocellaris* Linnaeus, 1767.** European Nucleotide Archive. [dataset], accession number PRJEB66023, 2023.
- 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 21 September 2024

<https://doi.org/10.21956/wellcomeopenres.23763.r93819>

© 2024 Ioannidis P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Panagiotis Ioannidis 

Foundation for Research & Technology - Hellas, Heraklion, Crete, GR 70013, Greece

The herein reported genome sequence of the *Tethea ocularis* moth is a high-quality assembly which is based on deep, PacBio sequencing (223-fold coverage). Additionally, BUSCO scores are also very good at almost 99%.

I expected that this genome is ready to be used for downstream analyses and as such I fully endorse this manuscript in its current form.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Insect genomics, bioinformatics

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 18 September 2024

<https://doi.org/10.21956/wellcomeopenres.23763.r94816>

© 2024 Ortiz 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.



Antonio S Ortiz 

University of Murcia, Murcia, Spain

Boyes & Wawman describes the genome assembly of *Tethea ocularis* and provides important genomic information about its size (near 340 Mb) in 31 chromosomes.

A new genome sequenced from the moth family Drepanidae is added to others as

- *Tethea albicostata* (Bremer 1861) by Liang *et al.* (2023) published in Mitochondrial DNA part B,
- *Drepana arcuata* Walker, 1855 by Yadav *et al.* (2020) published in Data in Brief and
- *Drepana falcataria* (Linnaeus, 1758) by Boyes & Lewis (2023),
- *Watsonalla binaria* (Hufnagel, 1767) by Boyes & Chua (2023) and
- *Achlya flavicornis* (Linnaeus, 1758) by Crowley & Phillips (2023), the last three published in Wellcome Open Research (WOR).

This assembly provides valuable information for the genomics of Lepidoptera and for further phylogenetic research in the evolutionary pattern in Lepidoptera.

The methods for the sequencing and assembly used are appropriate and technically supported but it would be interesting to offer a calculation of the number of genes that encode proteins.

I suggest the addition of author names and year in the first mention of the species in the text (*Tethea ocularis*, *Tethea or*, *Diloba caeruleocephala*, *T. ocularis f. fusca* and *Tetheella fluctuosa*).

References

1. Jiang N, Yang C, Xue D, Han H: An updated checklist of Thyatirinae (Lepidoptera, Drepanidae) from China, with descriptions of one new species. *Zootaxa*. 2015; **3941** (1): 1-48 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Yadav C, Smith ML, Yack JE: Transcriptome analysis of a social caterpillar, *Drepana arcuata*: De novo assembly, functional annotation and developmental analysis. *PLoS One*. 2020; **15** (6): e0234903 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Boyes D, Lewis OT, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: The genome sequence of the Pebble Hook-tip, *Drepana falcataria* (Linnaeus, 1758). *Wellcome Open Res.* 2023; **8**: 207 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Boyes D, Chua P, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: The genome sequence of the Oak Hook-tip, *Watsonalla binaria* (Hufnagel, 1767). *Wellcome Open Res.* 2023; **8**: 324 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Crowley L, Phillips D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: The genome sequence of the Yellow Horned, *Achlya flavicornis* (Linnaeus, 1758). *Wellcome Open Research*. 2023; **8**. [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?

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetics on Lepidoptera with special interest in evolutionary processes.

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.
