DATA NOTE

The genome sequence of the brimstone moth, *Opisthograptis luteolata* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes[1+],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Dominic Phillips[2],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]Natural History Museum, London, UK

[+] Deceased author

v1 **First published:** 12 Sep 2022, **7**:227
https://doi.org/10.12688/wellcomeopenres.18101.1
Latest published: 12 Sep 2022, **7**:227
https://doi.org/10.12688/wellcomeopenres.18101.1

Abstract

We present a genome assembly from an individual male *Opisthograptis luteolata* (the brimstone moth; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 363 megabases in span. The majority of the assembly (99.99%) is scaffolded into 31 chromosomal pseudomolecules with the Z sex chromosome assembled. The complete mitochondrial genome was also assembled and is 16.7 kilobases in length.

Keywords

Opisthograptis luteolata, brimstone moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 12 Sep 2022	✓ view	✓ view

1. **Leidys Murillo-Ramos**, Universidad de Sucre, Sucre, Colombia

2. **Jun Li**, Huaibei Normal University, Huaibei, China

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

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

Author roles: Boyes D: Investigation, Resources; **Phillips D**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Boyes D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the brimstone moth, *Opisthograptis luteolata* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2022, **7**:227 https://doi.org/10.12688/wellcomeopenres.18101.1

First published: 12 Sep 2022, **7**:227 https://doi.org/10.12688/wellcomeopenres.18101.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Geometroidea; Geometridae; Ennominae; *Opisthograptis*; *Opisthograptis luteolata* (Linnaeus, 1758) (NCBI:txid934882).

Background

The brimstone moth, *Opisthograptis luteolata* (Linnaeus, 1758), is a common, brightly coloured, yellow moth with markings along the leading edge of its wings and on each forewing tip; it is sometimes confused with the Brimstone butterfly due to their similar appearance. Very rare white forms of this species have occasionally been reported. *O. luteolata* is a nocturnal species found in Western Asia and across the Palearctic region and overwinters as part-grown larvae or in cocoons as pupae. The larvae mostly feed on plants in the Rosaceae family and emerge in two to three generations each year, with some authors suggesting a three-generation pattern over two years (Waring & Townsend, 2017). In *The colours of animals*, the green form of *O.luteolata* larvae is used as an example to describe counter-shading in insects (Poulton, 1890). This defensive method was more recently confirmed to be an effective form of crypsis in caterpillars (Rowland *et al.*, 2008). Alternatively, the darker larval forms mimic twigs present on host plant species.

The genome of *O.luteolata*, was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all of the named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *O.luteolata*, based on the ilOpiLute1 specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from a single male *O. luteolata* collected from near Chalet, Wytham, Berkshire, UK (Figure 1). A total of 61-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 93-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 2 missing joins, reducing the assembly size by 0.56% and the scaffold number by 23.26%.

The final assembly has a total length of 363 Mb in 33 sequence scaffolds with a scaffold N50 of 13.2 Mb (Table 1). The majority, 99.99%, of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length) and the Z sex chromosome (Figure 2–Figure 5; Table 2).

The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.3% (single 98.0%, duplicated 0.3%) using the lepidoptera_odb10 reference set (n=5,286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition and nucleic acid extraction

A single male *O. luteolata* specimen (ilOpiLute1; genome assembly) was collected using a light trap from near Chalet, Wytham, Berkshire, UK (latitude 51.772, longitude -1.337)

Figure 1. Image of the *Opisthograptis luteolata* specimen taken prior to preservation and processing.

Table 1. Genome data for *Opisthograptis luteolata*, ilOpiLute1.2.

Project accession data	
Assembly identifier	ilOpiLute1.2
Species	*Opisthograptis luteolata*
Specimen	ilOpiLute1 (genome assembly); ilOpiLute2 (Hi-C)
NCBI taxonomy ID	934882
BioProject	PRJEB48397
BioSample ID	SAMEA7519838
Isolate information	Male, whole organism (ilOpiLute1); abdomen tissue (ilOpiLute2)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR7224285
10X Genomics Illumina	ERR7220463-ERR7220466
Hi-C Illumina	ERR7220467
Genome assembly	
Assembly accession	GCA_931315375.2
Accession of alternate haplotype	GCA_931315605.2
Span (Mb)	363
Number of contigs	37
Contig N50 length (Mb)	13.2
Number of scaffolds	33
Scaffold N50 length (Mb)	13.1
Longest scaffold (Mb)	15.55
BUSCO* genome score	C:98.3%[S:98.0%,D:0.3%], F:0.5%,M:1.2%,n:5,286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.3.2. 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/CAKNUS02/dataset/CAKNUS02/busco.

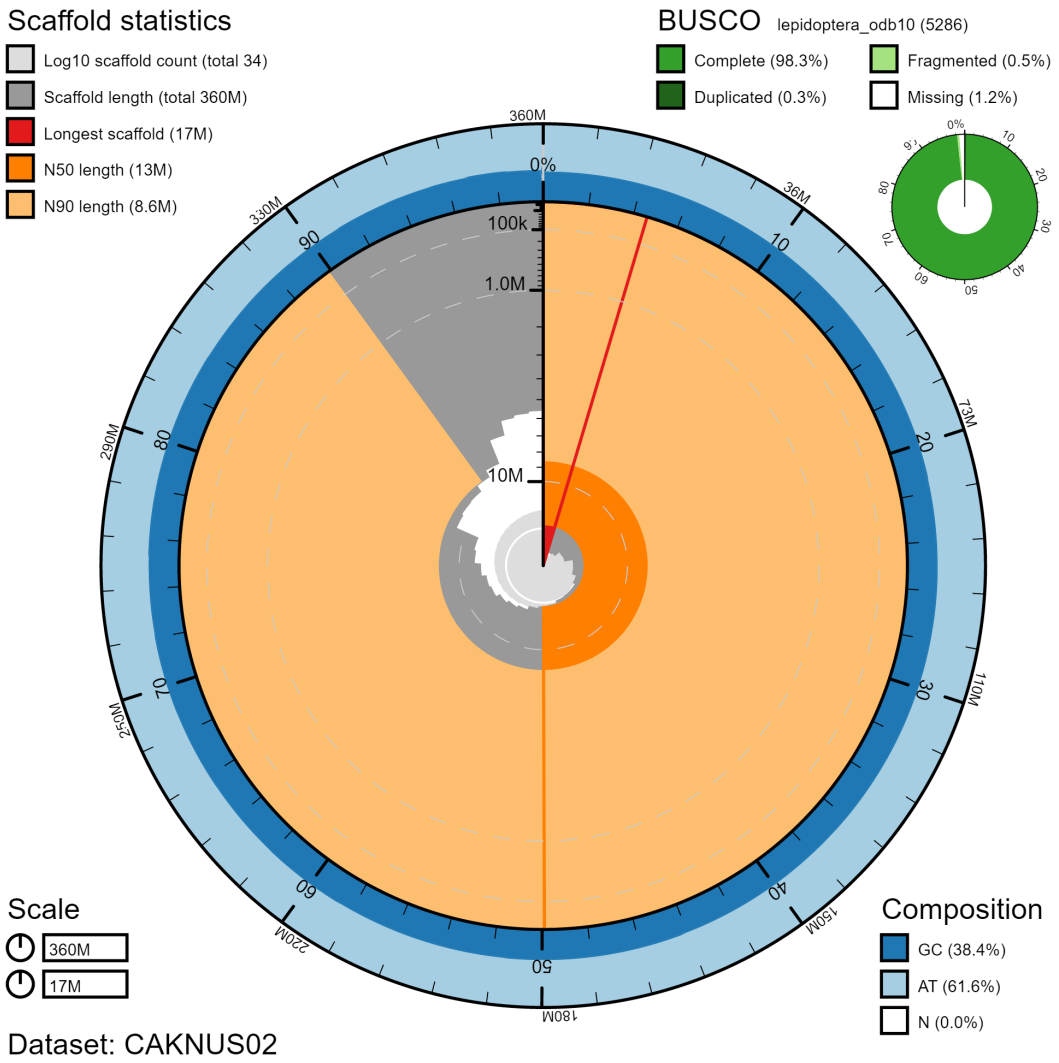

Figure 2. Genome assembly of *Opisthograptis luteolata*, ilOpiLute1.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 363,201,500 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (16,907,887 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (13,236,533 and 8,601,474 bp), respectively. The pale grey spiral shows the cumulative chromosome 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/CAKNUS02/dataset/CAKNUS02/snail.

by Douglas Boyes (University of Oxford). The specimen was identified by Douglas Boyes and snap-frozen on dry ice.

A single *O. luteolata* specimen of unknown sex (ilOpiLute2; Hi-C) was collected using a light trap from Wytham Woods, Berkshire, UK (latitude 51.771, longitude -1.337) by Douglas Boyes (University of Oxford). The specimen was identified by Douglas Boyes and snap-frozen on dry ice.

DNA was extracted at the Scientific Operations Core, Wellcome Sanger Institute. The ilOpiLute1 sample was weighed and dissected on dry ice. Whole organism tissue was disrupted by manual grinding in lysis buffer with a disposable pestle. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA

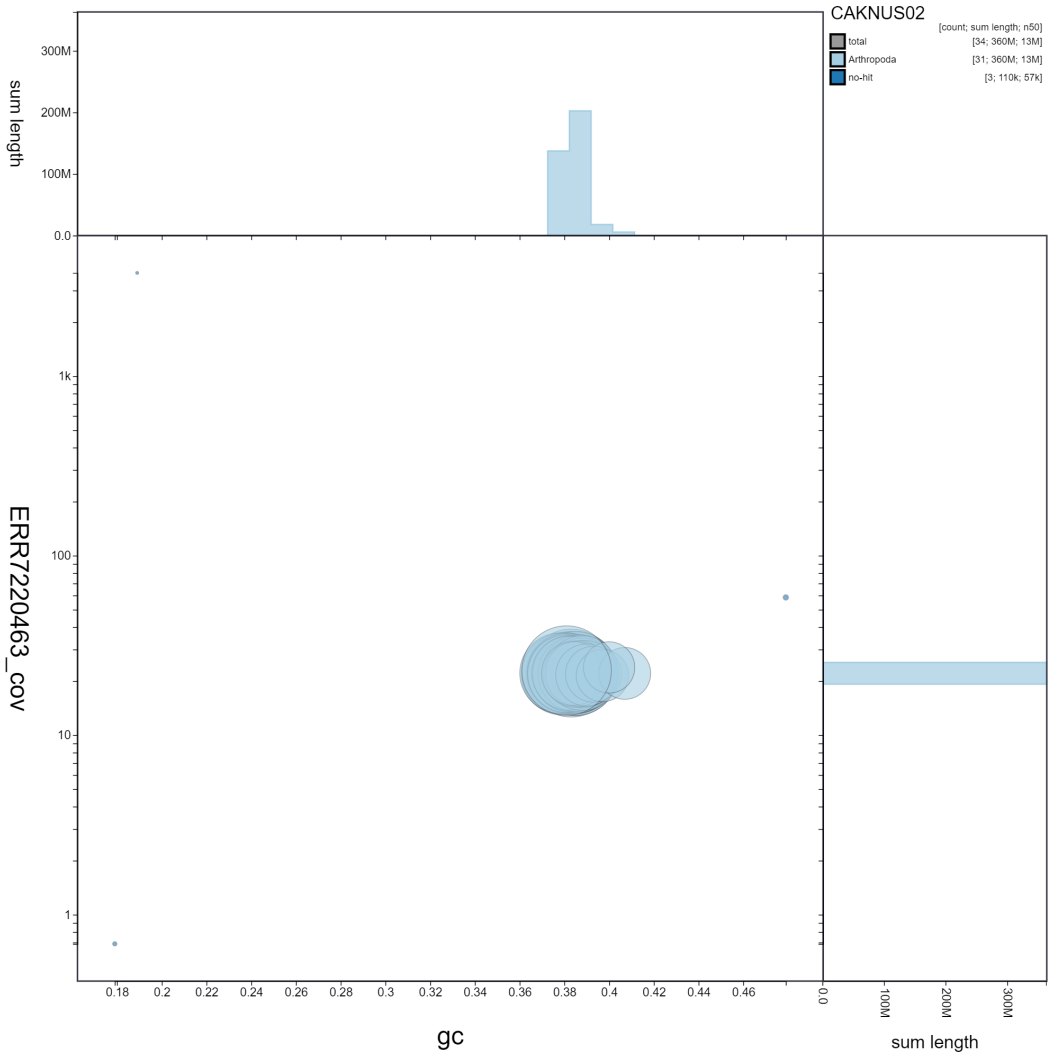

Figure 3. Genome assembly of *Opisthograptis luteolata*, ilOpiLute1.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAKNUS02/dataset/CAKNUS02/blob.

was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. 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.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina HiSeq (10X) instruments. Hi-C data were generated in the Tree of Life laboratory from abdomen tissue of ilOpiLute2 using the Arima v2 kit and sequenced on a NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021); haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2022). The assembly

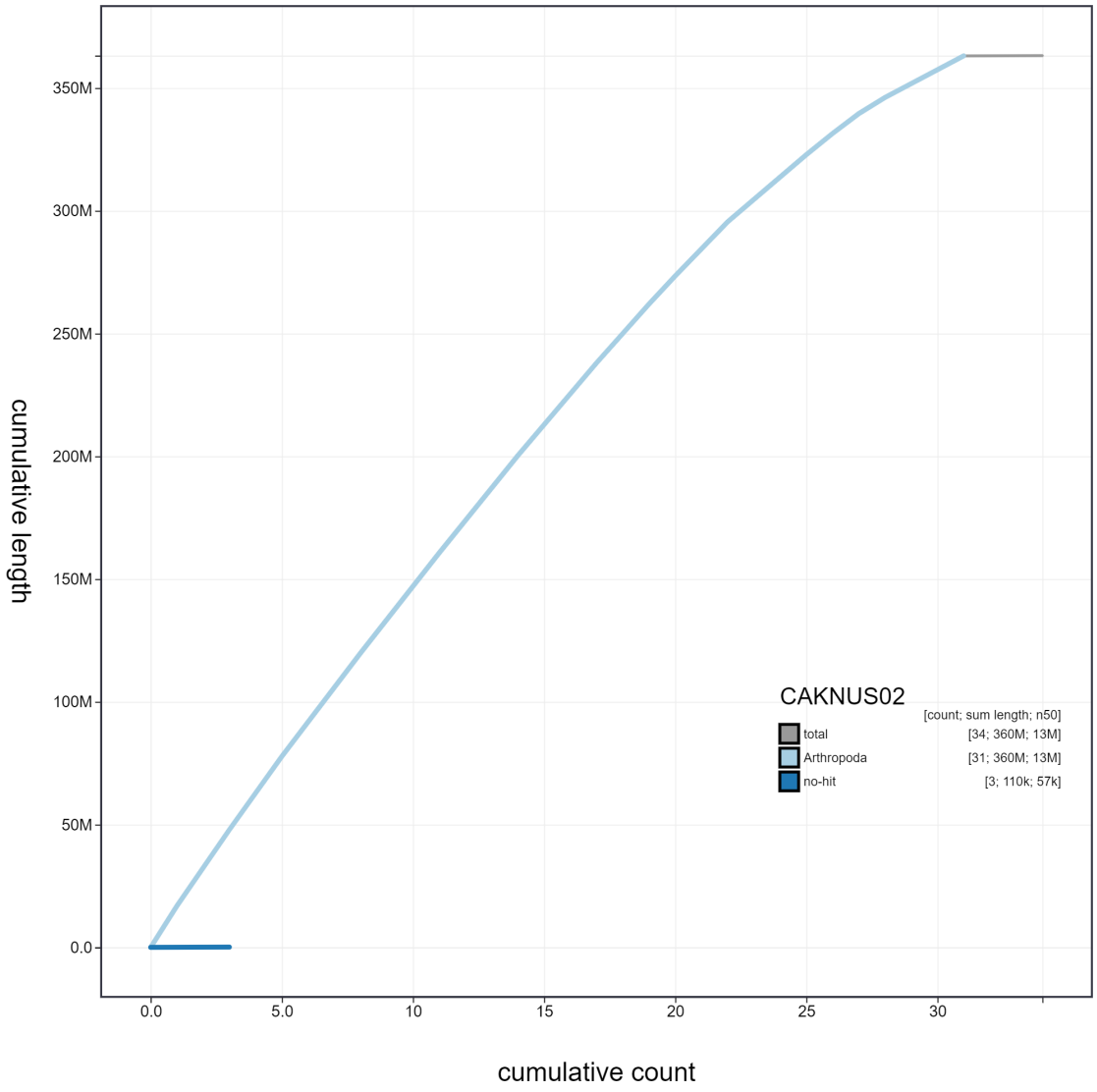

Figure 4. Genome assembly of *Opisthograptis luteolata*, ilOpiLute1.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAKNUS02/dataset/CAKNUS02/cumulative.

was checked for contamination as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performs annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Ethics/compliance issues

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. 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

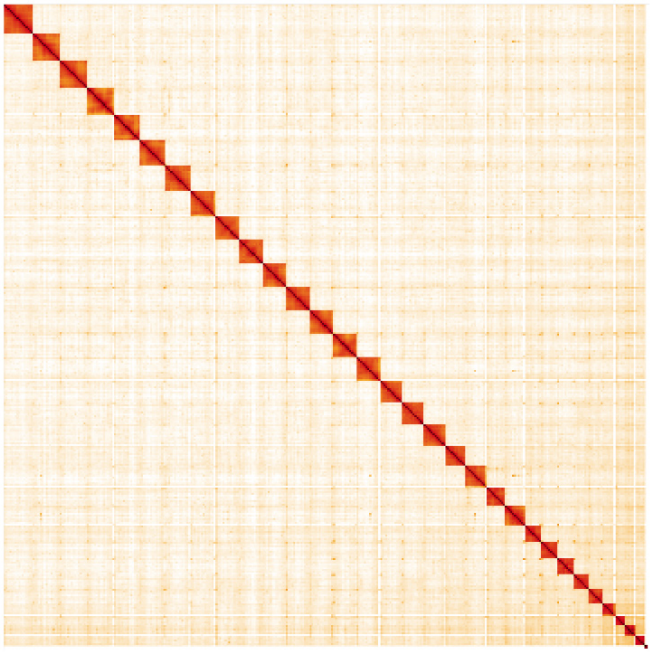

Figure 5. Genome assembly of *Opisthograptis luteolata*, ilOpiLute1.2: Hi-C contact map. Hi-C contact map of the ilOpiLute1.2 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=caRV68qHQESsjAlWXei8Fg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Opisthograptis luteolata*, ilOpiLute1.2.

INSDC accession	Chromosome	Size (Mb)	GC%
OV928034.1	1	15.55	38.3
OV928035.1	2	15.45	38.3
OV928036.1	3	14.94	38.5
OV928037.1	4	14.88	38.5
OV928038.1	5	14.19	37.9
OV928039.1	6	14.15	38
OV928040.1	7	13.82	38.1
OV928041.1	8	13.61	38.1
OV928042.1	9	13.6	37.9
OV928043.1	10	13.56	37.8
OV928044.1	11	13.33	38.5
OV928045.1	12	13.24	38.5
OV928046.1	13	13.1	38.1
OV928047.1	14	12.78	38.1
OV928048.1	15	12.59	38.6
OV928049.1	16	12.36	38.3
OV928050.1	17	12.03	38.5
OV928051.1	18	11.9	38.7
OV928052.1	19	11.54	38.2
OV928053.1	20	11.06	38.8
OV928054.1	21	10.92	38.8
OV928055.1	22	9.21	38.9
OV928056.1	23	9.1	38.5
OV928057.1	24	9.06	38.6
OV928058.1	25	8.6	38.4
OV928059.1	26	8.1	39
OV928060.1	27	6.58	39.3
OV928061.1	28	5.76	39.7
OV928062.1	29	5.64	40.7
OV928063.1	30	5.53	40
OV928033.1	Z	16.91	38.1
OV928065.1	MT	0.02	19
-	Unplaced	0.09	36.5

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.3	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
YaHS	1.0	Zhou *et al.*, 2022
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.0	Uliano-Silva *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.2.6	Challis *et al.*, 2020

to, the Darwin Tree of Life Project. 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: Opisthograptis luteolata (brimstone moth). Accession number PRJEB48397; https://identifiers.org/ena.embl/PRJEB48397.

The genome sequence is released openly for reuse. The *O. luteolata* 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 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.6418202.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.6418156.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.6866293.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.5746904.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.6125046.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.6418363.

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

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit-Interactive Quality Assessment of Genome Assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–74.
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): 170–75.
PubMed Abstract | Publisher Full Text | Free Full Text

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** 2012; arXiv:1207.3907.
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

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Poulton EB: **The Colours of Animals: Their Meaning and Use, Especially Considered in the Case of Insects**. D. Appleton and Company. 1890.
Reference Source

Rao SS, 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): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Rowland HM, Cuthill IC, Harvey IF, *et al.*: **Can't Tell the Caterpillars from the Trees: Countershading Enhances Survival in a Woodland.** *Proc Biol Sci.* 2008; **275**(1651): 2539–45.
PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0**. 2021.
Publisher Full Text

Waring P, Townsend M: **Field Guide to the Moths of Great Britain and Ireland**. 3rd ed. Bloomsbury Publishing. 2017.
Reference Source

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet Another Hi-C Scaffolding Tool.** *bioRxiv.* 2022.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 29 September 2022

https://doi.org/10.21956/wellcomeopenres.20072.r52336

✔ **Jun Li**
College of Life Sciences, Huaibei Normal University, Huaibei, China

In this article, the genomes of *Opisthograptis luteolata* (Linnaeus, 1758) were sequenced. The nuclear genome is 363 megabases and mitochondrial genome is 16.7 kilobases in length. Most of the assembly sequence were assigned to 31 chromosomal-level scaffolds which were corresponding to 30 autosomes and one sex chromosome. These are helpful for us to understand the moth.

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: Animal 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 27 September 2022

https://doi.org/10.21956/wellcomeopenres.20072.r52338

Leidys Murillo-Ramos
Department of Biology, Universidad de Sucre, Sucre, Colombia

The genome sequence of the brimstone moth, *Opisthograptis luteolata* (Linnaeus, 1758)

The authors present a genome assembly from Ophisthograptis luteolata (Geometridae), commonly known as the brimstone moth. The genome was sequenced from a fresh male sample collected near Chalet, Wytham, Berkshire, UK. The DNA extraction, library preparation and sequencing procedures followed the latest protocols in genome processing. The genome assembly was carried out with Hifiasm and following well known protocols. The genome sequencing resulted in 363 MB in 33 sequence scaffolds. As stated by the authors, the majority of the assembly (99.99%) is scaffolded into 31 chromosomal pseudomolecules with the Z sex chromosome assembled. The complete mitochondrial genome was also assembled and is 16.7 kilobases in length. Moreover, the data was compared to the lepidoptera_obd10 reference set from BUSCO dataset and they retrieved 98,3% of completeness. In figure 2, the authors present a great summary of the results. The plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 363,201,500 bp assembly. I do not have any further comments, and I encourage the publication of this research. I personally congratulate the authors for the initiative and collaborative effort that will be available to researchers and public interested in moths.

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 systematics

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.