DATA NOTE

The genome sequence of the yellow-tail moth, *Euproctis similis* (Fuessly, 1775) [version 1; peer review: 1 approved, 2 approved with reservations]

Douglas H. Boyes[1], Peter W.H. Holland[2],
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, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK
[2]Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

v1 **First published:** 13 Sep 2021, **6**:227
https://doi.org/10.12688/wellcomeopenres.17188.1

Latest published: 13 Sep 2021, **6**:227
https://doi.org/10.12688/wellcomeopenres.17188.1

Abstract

We present a genome assembly from an individual male *Euproctis similis* (the yellow-tail; Arthropoda; Insecta; Lepidoptera; Lymantriidae). The genome sequence is 508 megabases in span. The majority of the assembly is scaffolded into 22 chromosomal pseudomolecules, with the Z sex chromosome assembled.

Keywords

Euproctis similis, yellow-tail, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Reviewer Status ? ? ✓

	Invited Reviewers		
	1	**2**	**3**
version 1 13 Sep 2021	? report	? report	✓ report

1. **Steven M. Van Belleghem**, University of Puerto Rico-Rio Piedras, San Juan, Puerto Rico
2. **Niclas Backström**, Uppsala University, Uppsala, Sweden

   Daria Shipilina, Uppsala University, Uppsala, Sweden
3. **Thomas Blankers**, University of Amsterdam, Amsterdam, The Netherlands

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 DH: Formal Analysis, Investigation, Resources; **Holland PWH**: Formal Analysis, Investigation, Supervision, 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 DH, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the yellow-tail moth, *Euproctis similis* (Fuessly, 1775) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2021, **6**:227 https://doi.org/10.12688/wellcomeopenres.17188.1

First published: 13 Sep 2021, **6**:227 https://doi.org/10.12688/wellcomeopenres.17188.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Erebidae; Lymantriinae; Euproctis; *Euproctis similis* Fuessly 1775 (NCBI:txid987935).

Introduction

Larvae of *Euproctis similis* (yellow-tail) have long hairs that can cause skin irritation in humans, although the effects are rarely as serious as those caused by larvae of the closely related *Euproctis chrysorrhoea* (brown-tail). The genome of *E. similis* 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 *E. similis*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from a single male *E. similis* collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338). A total of 70-fold coverage in Pacific Biosciences single-molecule long reads (N50 17 kb) and 78-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 40 missing/misjoins and removed 3 haplotypic duplications, reducing the assembly length by 0.10% and the scaffold number by 42.00%, and increasing the scaffold N50 by 14.24%. The final assembly has a total length of 508 Mb in 30 sequence scaffolds with a scaffold N50 of 24 Mb (Table 1). The majority, >99.9%, of assembly sequence was assigned to 22 chromosomal-level scaffolds, representing 21 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 1–Figure 4; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) v5.1.2 completeness of 98.6% using the lepidoptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Table 1. Genome data for *Euproctis similis*, ilEupSimi1.1.

Project accession data	
Assembly identifier	ilEupSimi1.1
Species	*Euproctis similis*
Specimen	ilEupSimi1
NCBI taxonomy ID	NCBI:txid987935
BioProject	PRJEB42127
BioSample ID	SAMEA7519909
Isolate information	Male, head/abdomen/thorax
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6406199
10X Genomics Illumina	ERR6002639-ERR6002642
Hi-C Illumina	ERR6002643, ERR6002644
Genome assembly	
Assembly accession	GCA_905147225.1
Accession of alternate haplotype	GCA_905147215.1
Span (Mb)	508
Number of contigs	55
Contig N50 length (Mb)	21
Number of scaffolds	30
Scaffold N50 length (Mb)	24
Longest scaffold (Mb)	30
BUSCO* genome score	C:98.6%[S:97.7%,D:0.8%],F:0.3%,M:1.1%,n:5286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.1.2. 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/ilEupSimi1.1/dataset/CAJHUZ01/busco.

Methods

A single male *E. similis* was collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338) by Douglas Boyes, University of Oxford, using a light trap. The specimens were snap-frozen in dry ice using a CoolRack before transferring to the Wellcome Sanger Institute (WSI).

DNA was extracted at the Tree of Life laboratory, WSI. The ilEupSimi1 sample was weighed and dissected on dry ice with tissue set aside for RNA extraction and Hi-C sequencing. Thorax/abdomen tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. 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 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.

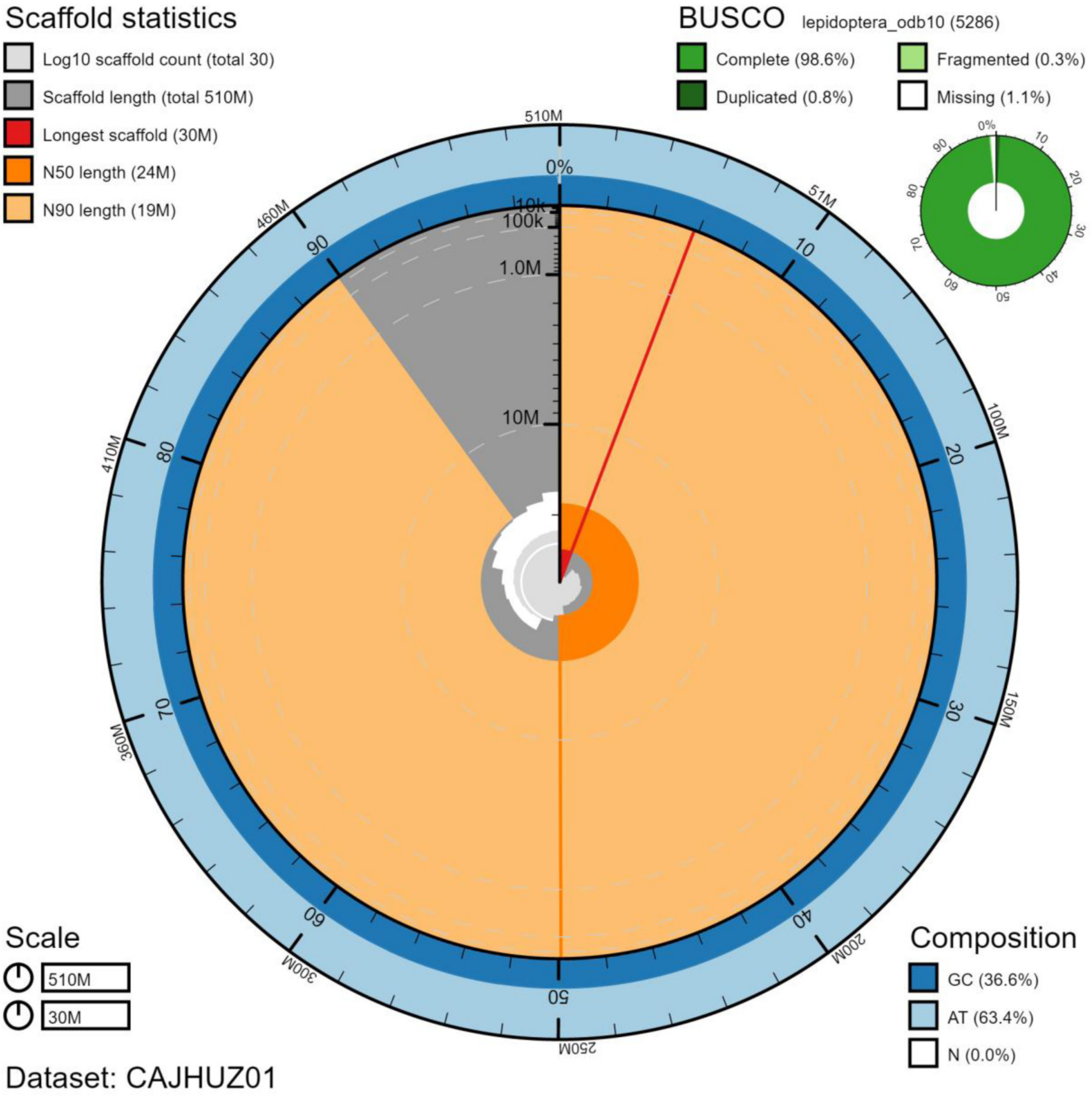

Figure 1. Genome assembly of *Euproctis similis*, ilEupSimi1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupSimi1.1/dataset/CAJHUZ01/snail.

RNA was extracted from thorax/abdomen tissue in the Tree of Life Laboratory at the WSI using TRIzol (Invitrogen), according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration RNA assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated from head tissue using the Qiagen EpiTect Hi-C kit and sequenced on HiSeq X.

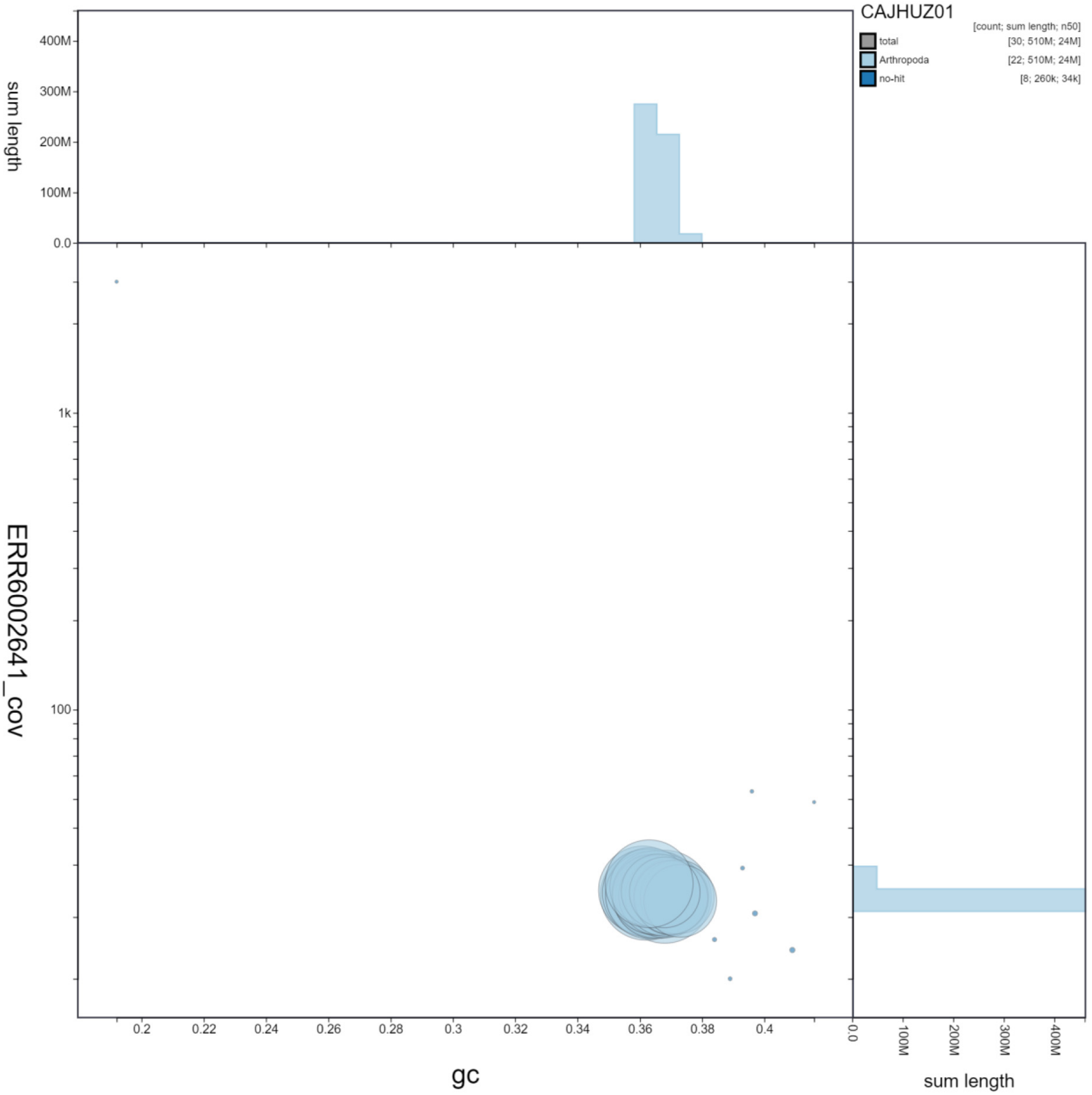

Figure 2. Genome assembly of *Euproctis similis*, ilEupSimi1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupSimi1.1/dataset/CAJHUZ01/blob.

Assembly was carried out with HiCanu (Nurk *et al.*, 2020); haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was polished with the 10X Genomics Illumina data by aligning to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). One round of the Illumina polishing was applied. Scaffolding with Hi-C data (Rao *et al.*, 2014) was carried out with SALSA2 (Ghurye *et al.*, 2019). 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 Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021). The genome was analysed and BUSCO scores generated

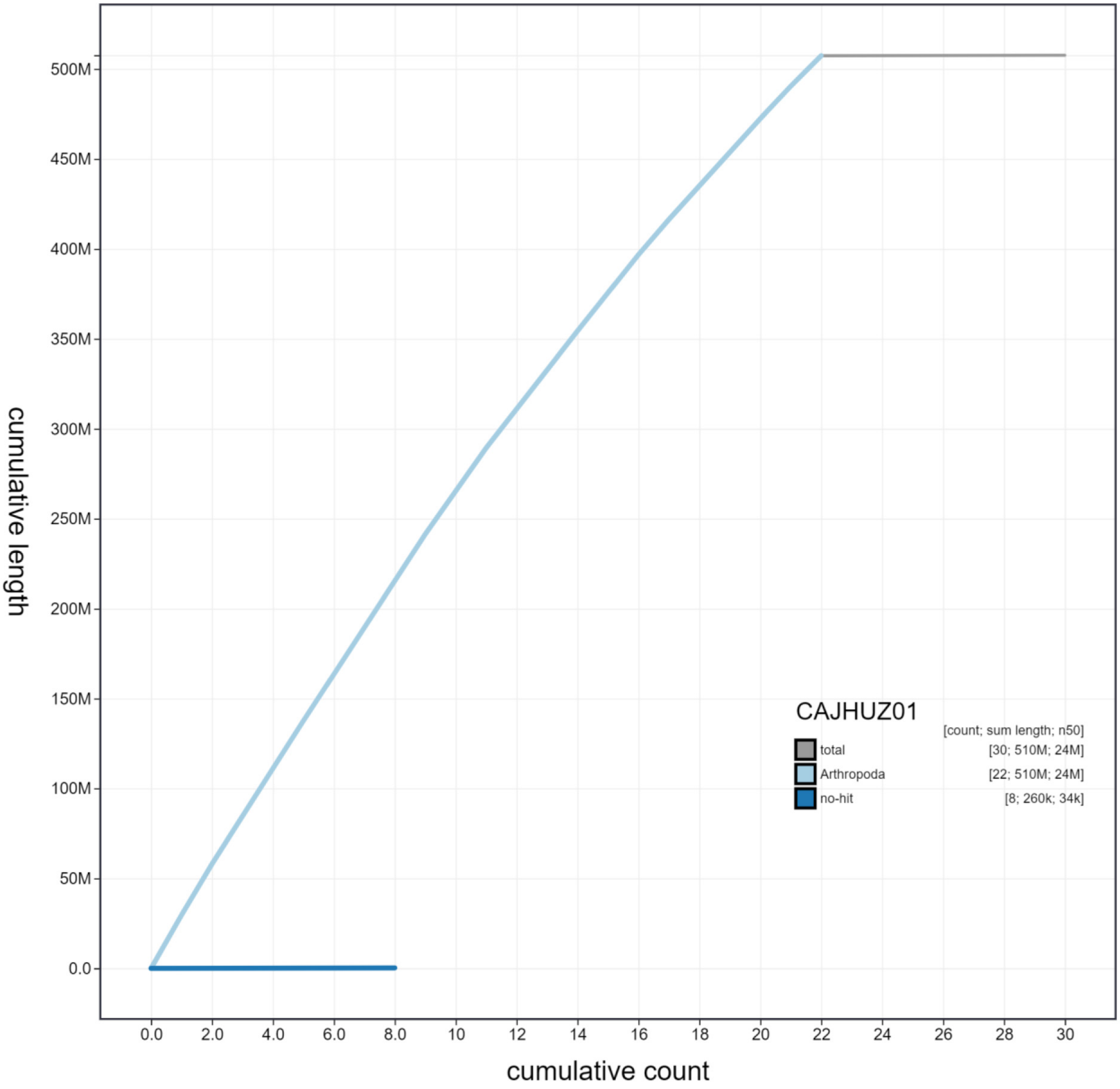

Figure 3. Genome assembly of *Euproctis similis*, ilEupSimi1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupSimi1.1/dataset/CAJHUZ01/cumulative.

within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

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 to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration

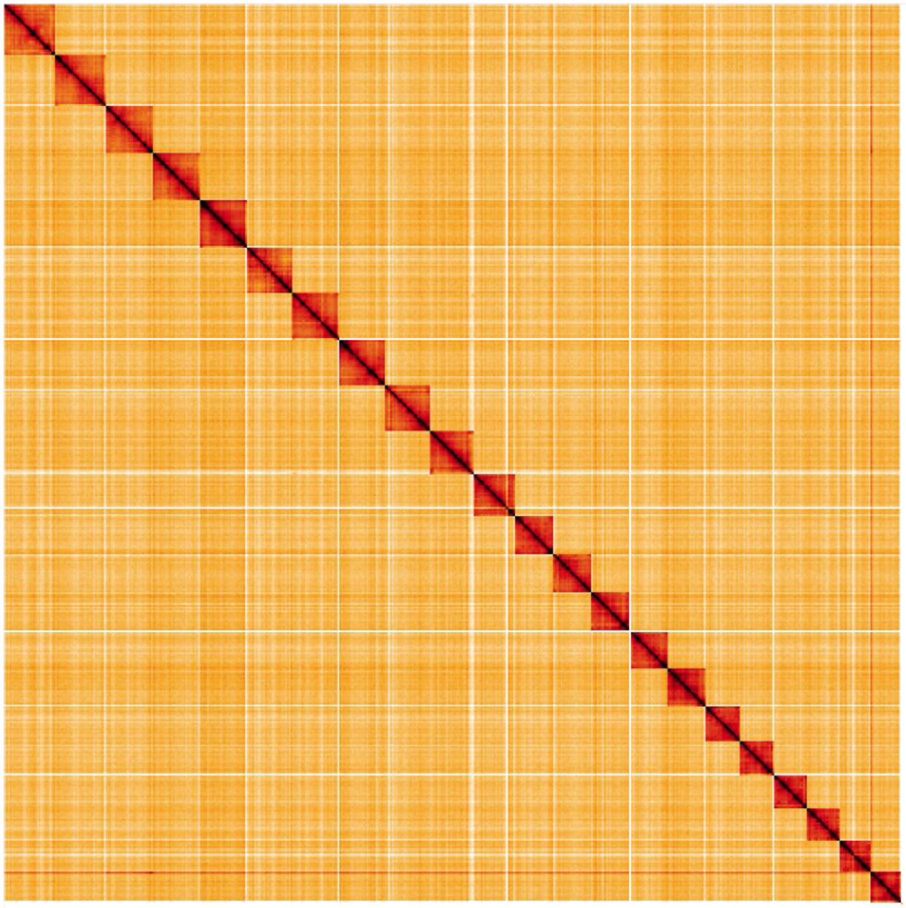

Figure 4. Genome assembly of *Euproctis similis*, ilEupSimi1.1: Hi-C contact map. Hi-C contact map of the ilEupSimi1.1 assembly, visualised in HiGlass.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Euproctis similis*, ilEupSimi1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
LR990103.1	1	29.63	36.8
LR990104.1	2	28.43	36.2
LR990105.1	3	26.72	36.1
LR990106.1	4	26.40	36.3
LR990108.1	5	26.05	36.2
LR990109.1	6	26.03	36.3
LR990110.1	7	25.95	36.6
LR990111.1	8	25.82	36.5
LR990112.1	9	24.45	36.5
LR990113.1	10	23.51	36.8
LR990114.1	11	21.70	36.5
LR990115.1	12	21.66	36.4
LR990116.1	13	21.63	36.6
LR990117.1	14	21.40	36.8
LR990118.1	15	21.08	36.3
LR990119.1	16	19.59	37
LR990120.1	17	18.87	36.6
LR990121.1	18	18.67	37.2
LR990122.1	19	18.51	37
LR990123.1	20	17.96	37.3
LR990124.1	21	16.98	36.8
LR990107.1	Z	26.35	36.3
LR990125.1	MT	0.02	19.4
-	Unplaced	0.24	39.9

Table 3. Software tools used.

Software tool	Version	Source
HiCanu	2.1	Nurk *et al.*, 2020
purge_dups	1.2.3	Guan *et al.*, 2020
SALSA2	2.2	Ghurye *et al.*, 2019
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	1	Uliano-Silva *et al.*, 2021
gEVAL	N/A	Chow *et al.*, 2016
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.1.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	2.6.2	Challis *et al.*, 2020

Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the WSI), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Euproctis similis (yellow-tail). Accession number PRJEB42127: https://identifiers.org/ena.embl:PRJEB42127

Acknowledgements

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789929.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893704.

Members of the Wellcome Sanger Institute Tree of Life programme collective are listed here: https://doi.org/10.5281/zenodo.5377053.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790456.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013542.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783559.

References

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

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a Web-Based Browser for Evaluating Genome Assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–10.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
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): 2896–98.
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

Nurk S, Walenz BP, Rhie A, *et al.*: **HiCanu: Accurate Assembly of Segmental Duplications, Satellites, and Allelic Variants from High-Fidelity Long Reads.** *Genome Res.* 2020; **30**(9): 1291–1305.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ? ? ✓

Version 1

Reviewer Report 24 September 2021

https://doi.org/10.21956/wellcomeopenres.18991.r45902

✓ **Thomas Blankers**

Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

This data note presents the genome assembly of *Euproctis similis* as part of the Darwin Tree of Life Project. The information about the molecular lab methods is sufficient and clear. The appropriate, protocols have been used, following recommendations of the manufacturers and sequencing platforms. The quality of the genome is high, owing to the use of complementary sequencing technologies that allow for contiguous assemblies.

I only have three small suggestions for additional information, although I also see that none of these are commonly supplied in the notes coming from the Darwin Tree of Life project. First, as a biologist, I would be interested in knowing a little bit more about the organism. For example that it is a night-active moth, wide-spread across the Eurasian continent, that they're active from August to June and that they are associated with both urban and non-urban habitats and with several host plants. Second, in the presentation of the methods there are no details about the bioinformatic analyses beyond the programs that were used. It would be good to specify any deviation from default settings or even to have a brief summary of the commands used to perform the analyses. This could be done in a separate file archived along with the note or in a Table, possibly integrated in Table 3. Third, the data presentation can benefit from brief expansion of the results. The interactive figures are nice, because some explanation of what is shown can also be found at the corresponding blobtoolkit repository. However, there is no text accompanying these figures beyond a single sentence referencing the number of scaffolds and citing figures 1 through 4. Some expansion of the genome assembly statistics seems desirable. And figure 4 could use a legend as well as axis labels with the chromosome numbers. Again, I do see that other examples of notes on genomes coming from this project also do not necessarily contain these additional pieces of information, so I guess it is up to the authors to decide whether that continuity matters more or whether the details are simply not necessary.

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: Evolutionary genomics, currently working on crickets, moths, and C. elegans

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 20 September 2021

https://doi.org/10.21956/wellcomeopenres.18991.r45899

?

Niclas Backström
Evolutionary Biology Program, Department of Ecology and Genetics (IEG), Uppsala University, Uppsala, Sweden

Daria Shipilina
Evolutionary Biology Program, Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

The manuscript by Boyes and Holland describes an effort to assemble both the nuclear and mitochondrial genome of a male yellow-tailed moth. The procedure includes multiple sequencing techniques allowing for both primary assembly of contigs from long-read libraries (PacBio), polishing with linked reads (10X) and scaffolding with chromosome interaction information (HiC). The highly contiguous (seven very short scaffolds, that only sum up to 0.24 Mb in total, that are not linked to chromosomes) final assembly contains 21 autosomes, the Z-chromosome and the mitochondrial genome. This will be a valuable resource for both population genetic analyses within the species/genus and comparative genomics studies in insects in general and in Lepidoptera in particular. The study sets high standards for the methodology of genome assembly by using and efficiently combining state-of-the-art sequencing technologies and novel bioinformatic pipelines.

Minor comments:

Abstract:

1. Maybe omit 'majority of the'.

2. Information about that the mitochondrial genome is also assembled should be included.

3. Perhaps mention briefly that the number of assembled chromosomes is lower than many previously sequenced lepidopterans – perhaps indicating fusions in the *Euproctis* lineage?

Introduction:

1. A brief description of the abundance and distribution range of this species in the AABI area and globally would be very informative.

2. Perhaps it could be of interest to characterize species specific and general biological questions that can be addressed using this genomic resource.

Genome sequence report + Methods:

1. The first sentence in these two sections is redundant. Perhaps it is sufficient to describe the sampling location, collector, sample ID and date of sampling (missing) in the methods section and omit the redundant details in the GSR section.

Genome sequence report:

1. Omit 'The majority,"?

2. Add information about that the mitochondrial genome was assembled (including assessment of completeness, e. g. exact length of the scaffold)?

3. The finding that scaffolds are anchored to only 21 autosomes is a bit surprising given that most lepidopterans sequenced so far have 31 chromosome pairs. This could be another example of chromosome rearrangements (here potentially fusions, similar to what has been observed in e.g. *Heliconius melpomene*)? This might be worth mentioning briefly.

4. Were RNAseq sequencing reads used in the downstream analysis?

5. Table 1. Unclear if the "Number of contigs" corresponds to the number of contigs before or after the manual correction of the Hi-C map?

Methods:

1. 'specimen was' instead of 'specimens were'.

2. Second paragraph. The sample ID can be given in the first part of the section where sample location, collector, date of sampling etc. is stated.

3. 4th line, left column under Figure 1. 'its concentration RNA assessed' should perhaps be 'the RNA concentration assessed'.

4. Information about RNAseq library preparation method, sequencing technique applied and data availability of RNAseq reads is missing.

Reflection:

1. An annotation with the RNAseq data generated here in combination with previously available lepidopteran gene models would add useful information for downstream analyses using this resource.

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?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative genomics, population genetics, molecular evolution.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 13 September 2021

https://doi.org/10.21956/wellcomeopenres.18991.r45833

?

Steven M. Van Belleghem

Department of Biology, University of Puerto Rico-Rio Piedras, San Juan, Puerto Rico

Douglas Boyes and Peter Holland report a high-quality genome assembly for the yellow-tail moth from the UK. They do this using PacBio and 10X sequencing and scaffolded the initially obtained contigs using Hi-C data. The data and assembly seem of high quality and the methods used are rigorous. The assembly contains 30 scaffolds assigned to 22 chromosomes and has a 98.6 BUSCO completeness. I believe this chromosome level assembly will be a valuable tool for future studies.

Abstract:

- I suggest using a more precise alternative to the word ‘majority’.

Introduction:

- Is it possible to give more precise information on the geographic distribution of this species?

Results:

- I do not see any results on the RNA-seq data. Was a transcriptome assembly performed? I also do not see accession numbers to these data in Table 1.

- An assessment of the repetitiveness (TE content) of the genome could be useful.

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?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, functional genomics, insects

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.