DATA NOTE

The genome sequence of the poplar hawk-moth, *Laothoe populi* (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations]

Douglas 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, OX10 8BB, UK
[2]Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

v1 **First published:** 16 Sep 2021, **6**:237
https://doi.org/10.12688/wellcomeopenres.17191.1
Latest published: 16 Sep 2021, **6**:237
https://doi.org/10.12688/wellcomeopenres.17191.1

Abstract

We present a genome assembly from an individual female *Laothoe populi* (the poplar hawk-moth; Arthropoda; Insecta; Lepidoptera; Sphingidae). The genome sequence is 576 megabases in span. The majority of the assembly is scaffolded into 29 chromosomal pseudomolecules, with the W and Z sex chromosome assembled.

Keywords

Laothoe populi, poplar hawk-moth, genome sequence, chromosomal

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ?

	1	2
version 1 16 Sep 2021	✓ view	? view

1. **Roderic Guigo Guigo**, The Barcelona Institute of Science and Technology, Barcelona, Spain

2. **Petr Nguyen**, University of South Bohemia, Ceske Budejovice, Ceske Budejovice, Czech Republic

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: Formal Analysis, Investigation, Resources; **Holland PWH**: 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 D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the poplar hawk-moth, *Laothoe populi* (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations]** Wellcome Open Research 2021, **6**:237 https://doi.org/10.12688/wellcomeopenres.17191.1

First published: 16 Sep 2021, **6**:237 https://doi.org/10.12688/wellcomeopenres.17191.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Sphingidae; Smerinthinae; Smerinthini; Laothoe; *Laothoe populi* Linnaeus 1758 (NCBI:txid522836).

Introduction

Laothoe populi (Poplar hawk-moth) is one of the largest native Lepidoptera species in the UK; larval colouration varies and relates to differences in sequestration and transport of carotenoids derived from foodplants, poplar (*Populus* sp.) and willow (*Salix* sp.) (Grayson *et al.*, 1991). The genome of *L. populi* 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 *L. populi*, based on one female specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from a single female *L. populi* collected from Wytham Woods, Oxfordshire, UK (latitude 51.768, longitude -1.337). A total of 28-fold coverage in Pacific Biosciences single-molecule long reads and 68-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 103 missing/misjoins and removed 20 haplotypic duplications, reducing the assembly length by 1.19% and the scaffold number by 61.45%, and increasing the scaffold N50 by 12.08%. The final assembly has a total length of 576 Mb in 33 sequence scaffolds with a scaffold N50 of 21 Mb (Table 1). Of the assembly sequence, >99.9% was assigned to 29 chromosomal-level scaffolds, representing 27 autosomes (numbered by sequence length), and the W and Z sex chromosome (Figure 1–Figure 4; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) completeness of 98.8% 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 *Laothoe populi*, ilLaoPopu1.1.

Project accession data	
Assembly identifier	ilLaoPopu1
Species	*Laothoe populi*
Specimen	ilLaoPopu1
NCBI taxonomy ID	NCBI:txid522836
BioProject	PRJEB42952
BioSample ID	SAMEA7520519
Isolate information	Female, head/abdomen/thorax
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6406202, ERR6412028
10X Genomics Illumina	ERR6054412-ERR6054415
Hi-C Illumina	ERR6054411
Genome assembly	
Assembly accession	GCA_905220505.1
Accession of alternate haplotype	GCA_905220495.1
Span (Mb)	576
Number of contigs	135
Contig N50 length (Mb)	7
Number of scaffolds	33
Scaffold N50 length (Mb)	21
Longest scaffold (Mb)	30
BUSCO* genome score	C:98.8%[S:98.5%,D:0.4%],F:0.3%,M:0.8%,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/ilLaoPopu1.1/dataset/CAJNAD01/busco.

Methods

A single female *L. populi* was collected from Wytham Woods, Oxfordshire, UK (latitude 51.768, longitude -1.337) 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 ilLaoPopu1 sample was weighed and dissected on dry ice with head/thorax tissue set aside for Hi-C sequencing. 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.

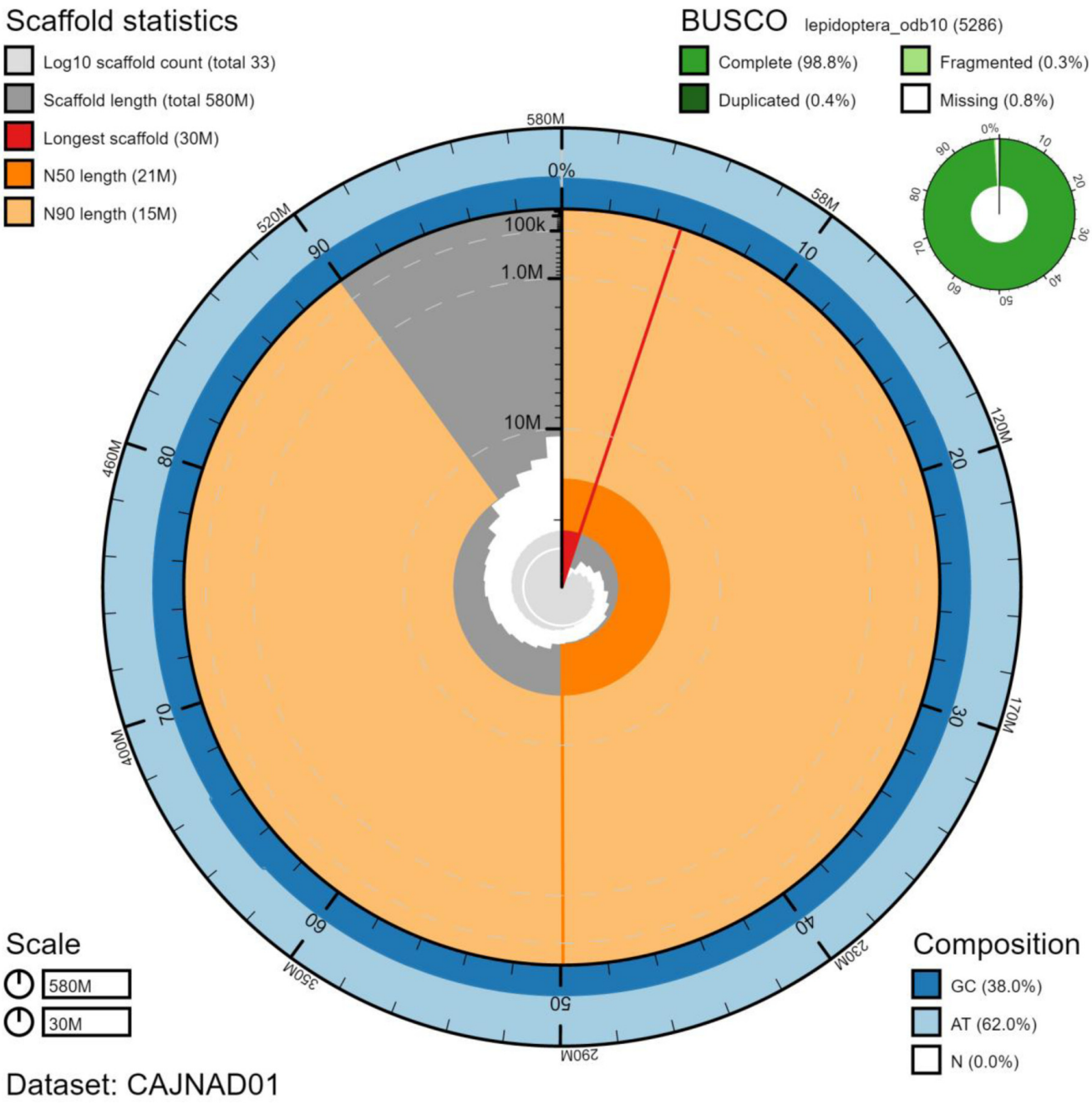

Figure 1. Genome assembly of *Laothoe populi*, ilLaoPopu1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilLaoPopu1.1/dataset/CAJNAD01/snail.

Pacific Biosciences HiFi circular consensus and 10X Genomics 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 and Illumina HiSeq X instruments. HiC data were generated from head/thorax tissue using the Arima v2.0 kit and sequenced on HiSeq X.

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021); haplotypic duplication was identified and removed with purge_dups

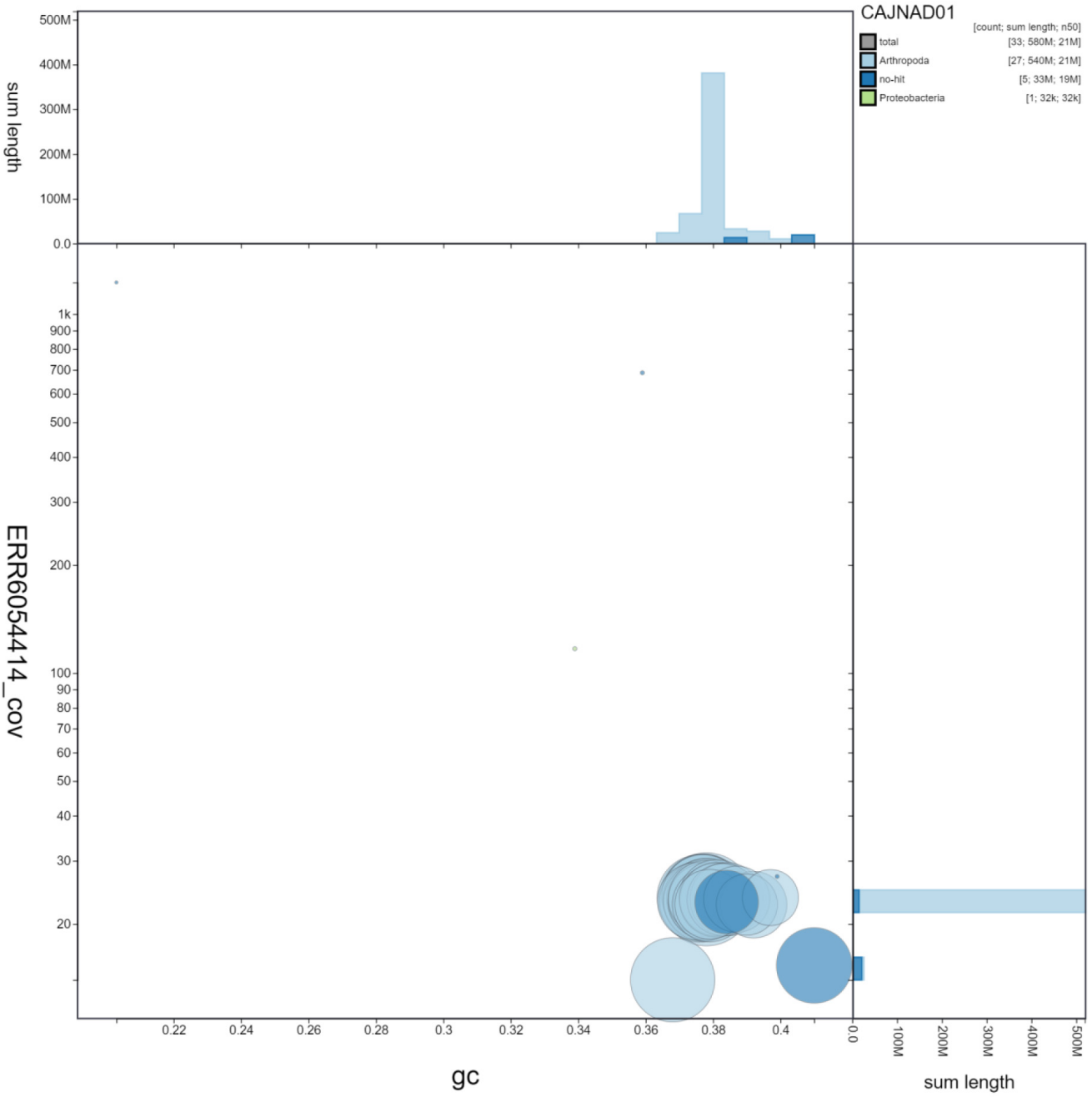

Figure 2. Genome assembly of *Laothoe populi*, ilLaoPopu1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilLaoPopu1.1/dataset/CAJNAD01/blob.

(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 within the

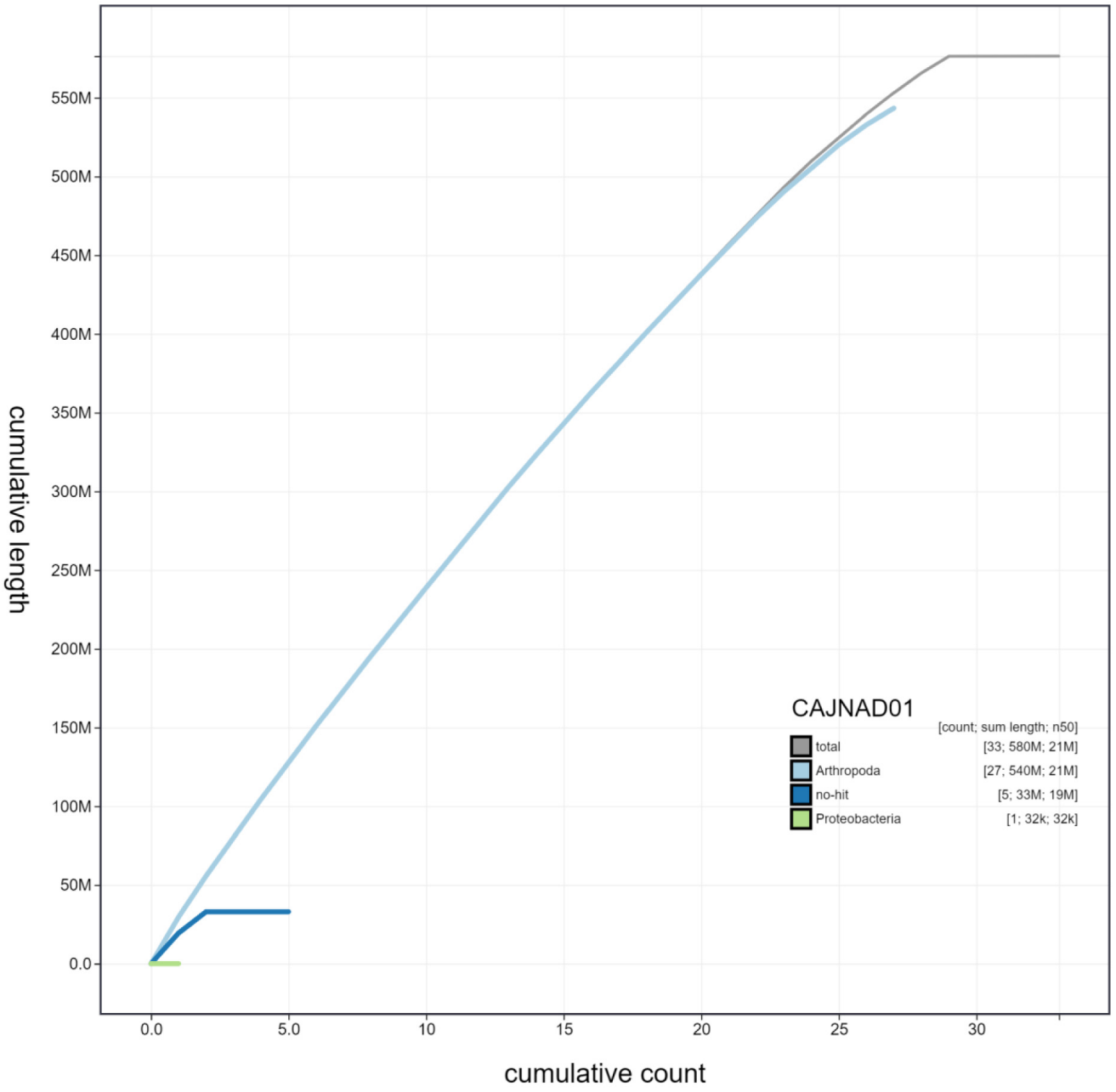

Figure 3. Genome assembly of *Laothoe populi*, ilLaoPopu1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilLaoPopu1.1/dataset/CAJNAD01/cumulative.

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

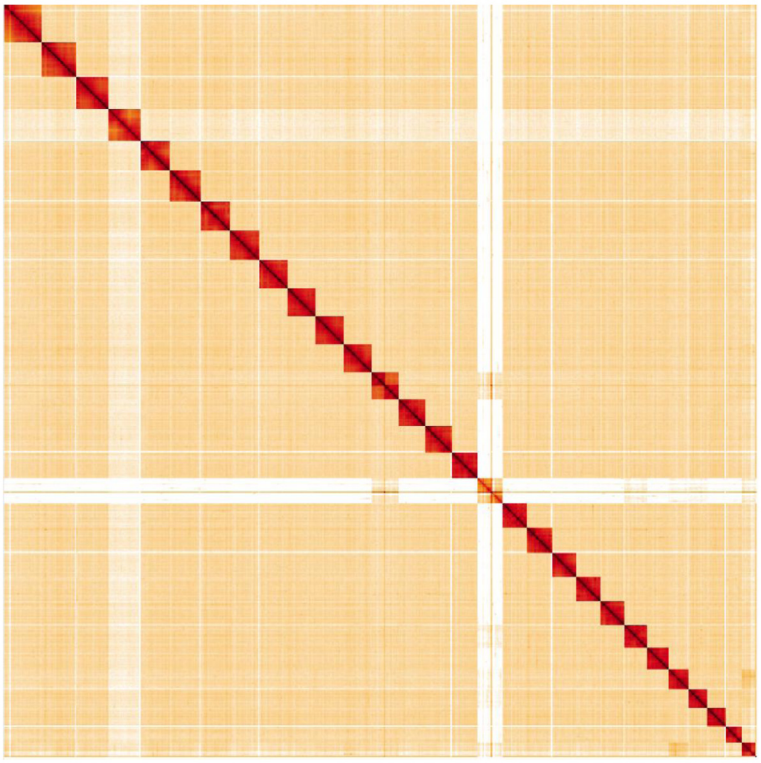

Figure 4. Genome assembly of *Laothoe populi*, ilLaoPopu1.1: Hi-C contact map. Hi-C contact map of the ilLaoPopu1.1 assembly, visualised in HiGlass.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Laothoe populi*, ilLaoPopu1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
HG992146.1	1	29.55	37.8
HG992147.1	2	26.14	37.7
HG992148.1	3	24.61	37.6
HG992150.1	4	23.33	37.7
HG992151.1	5	23.18	37.7
HG992152.1	6	22.57	37.7
HG992153.1	7	22.07	37.6
HG992154.1	8	21.63	37.7
HG992155.1	9	21.58	37.8
HG992156.1	10	21.43	37.8
HG992157.1	11	21.37	37.8
HG992158.1	12	21.13	37.7
HG992159.1	13	20.40	37.5
HG992160.1	14	20.01	37.8
HG992161.1	15	19.67	37.8
HG992163.1	16	19.03	37.9
HG992164.1	17	18.92	38.1
HG992165.1	18	18.59	38.1
HG992166.1	19	18.42	38.1
HG992167.1	20	18.17	38.4
HG992168.1	21	17.79	38.3
HG992169.1	22	16.54	37.9
HG992170.1	23	15.01	39.2
HG992171.1	24	14.83	38.7
HG992172.1	25	13.54	38.4
HG992173.1	26	12.63	39
HG992174.1	27	10.59	39.7
HG992162.1	W	19.39	41
HG992149.1	Z	24.18	36.8
HG992175.1	MT	0.02	20.5
-	Unplaced	0.09	36.4

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.12	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
longranger	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	1.0	Uliano-Silva *et al.*, 2021
SALSA2	2.2	Ghurye *et al.*, 2019
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

for, and supplied 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: Laothoe populi (poplar hawk-moth). Accession number PRJEB42952: https://identifiers.org/ena.embl:PRJEB42952

The genome sequence is released openly for reuse. The *L. populi* 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. Raw data and assembly accession identifiers are reported in Table 1.

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

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

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

Grayson J, Edmunds M, Evans EH, *et al.*: **Carotenoids and Colouration of Poplar Hawkmoth Caterpillars (*Laothoe Populi*).** *Biological Journal of the Linnean Society.* Linnean Society of London, 1991; **42**(4): 457–65.
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

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 22 November 2021

https://doi.org/10.21956/wellcomeopenres.18994.r46904

?

Petr Nguyen
University of South Bohemia, Ceske Budejovice, Ceske Budejovice, Czech Republic

The authors report a chromosome-level assembly of the female poplar hawk-moth, *Laothoe populi*, sequenced by PacBio HiFi and 10x Genomics technologies. There is no clear reason for sequencing the genome of *L. populi* as the species has not been of any particular interest to researchers.

The *L. populi* assembly is 576 Mb long. The *L. populi* genome size is not known but the assembly length falls well within the genome size range of other representatives of the family Sphingidae (www.genomesize.com). Some comparison with a k-mer based estimate would be nice. More than 99.9% of the sequence was assigned to 29 chromosome-level scaffolds, which correspond to 27 autosomes and sex chromosomes W and Z. This is in agreement with a haploid chromosome number n=28 reported for the species (Robinson 1971, Lepidoptera genetics). The sex chromosomes were probably identified based on their coverage (Figure 2) and Hi-C contact map (Figure 4), but this is not clear from the text and could be explained in more detail in figure legends. Based on Figure 4 it seems that the W chromosome interacts with some of autosomes. I wonder whether it could point to some misassemblies or there is some interesting biology involved. The autosomes are numbered by their size which will be problematic and confusing in future comparative studies. Numbering based on gene synteny with species with ancestral chromosome number n=31 would be preferable.

The methods section describes the pipeline but not the parameters used. For better reproducibility, parameters used for individual processing and assembly steps should be specified. I believe a tool used for polishing is missing.

Minor comments:
In methods, “The specimens were snap-frozen” should read “The specimen was snap-frozen” as only one female was collected.
I believe the sentence “Fragment size analysis of 0.01-0.5 ng of DNA was then performed using an Agilent FemtoPulse” should follow DNA extraction. Or was the fragment size analysis performed with crude homogenate?

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: Genetics of Lepidoptera

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.

Reviewer Report 18 October 2021

https://doi.org/10.21956/wellcomeopenres.18994.r46101

Roderic Guigo Guigo
Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain

This is a concise, clear, precise note that describes the assembly of a moth species within the Darwin Tree of Life project. I am informed, but not an expert in genome assembly. I believe, however, that the metrics provided show that this is a high quality assembly, well exceeding the VGP/EBP standards.

As a non-expert in genome assembly, it took me a while to fully understand the information displayed in the Figures - in particular, those generated by the Blob Toolkit. Accessing the interactive version of the figures helped some, but the meaning of the different figure components is not immediately obvious either through the interactive interface. If these plots will become the standard for the DToL genome data notes, it maybe helpful to have a direct link to a description of what is shown in the figures.

Regarding the static figures, maybe there should be an easy way to show the chromosomal scaffolds, I did not find a way to easily visualize the chromosomes in the snail plot.

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: Transcriptomics, Gene finding, 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.