




## DATA NOTE

# The genome sequence of the angle shades moth, *Phlogophora meticulosa* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes<sup>1†</sup>, Peter W.H. Holland <sup>2</sup>,  
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

<sup>1</sup>UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK

<sup>2</sup>Department of Zoology, University of Oxford, Oxford, UK

<sup>†</sup> Deceased author

**V1** First published: 14 Mar 2022, 7:89  
<https://doi.org/10.12688/wellcomeopenres.17757.1>

Latest published: 14 Mar 2022, 7:89  
<https://doi.org/10.12688/wellcomeopenres.17757.1>

## Abstract



We present a genome assembly from an individual female *Phlogophora meticulosa* (the angle shades; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 539 megabases in span. The majority of the assembly, 95.17%, is scaffolded into 31 chromosomal pseudomolecules, with the Z sex chromosome assembled. Some unassigned scaffolds are identified as belonging to the W chromosome based on half-depth coverage and comparison to other Noctuidae W chromosomes. The mitochondrial genome was also assembled and is 15.4 kilobases in length.



## Keywords

*Phlogophora meticulosa*, angle shades, genome sequence, chromosomal, Lepidoptera

## Open Peer Review

Approval Status  

	1	2
version 1		
14 Mar 2022	<a href="#">view</a>	<a href="#">view</a>

1. Sara J. Oppenheim , American Museum of Natural History, New York, USA
2. Jing Zhang , University of Texas Southwestern Medical Center, Dallas, USA

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



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

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

**Author roles:** Boyes D: Investigation, Resources; Holland PWH: 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, <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:** © 2022 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, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the angle shades moth, *Phlogophora meticulosa* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2022, 7:89 <https://doi.org/10.12688/wellcomeopenres.17757.1>

**First published:** 14 Mar 2022, 7:89 <https://doi.org/10.12688/wellcomeopenres.17757.1>

## Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Hadeninae; Phlogophora; *Phlogophora meticulosa* (Linnaeus, 1758) (NCBI: txid875884).

## Background

*Phlogophora meticulosa* (angle shades) is a delicately-patterned moth with characteristically ‘wrinkled’ wings that do not lie in a flat plane; the unusual wing form contributes to camouflage amongst dead leaves and may also influence flight through permitting deformation (Wootton, 2020). The species can be found across most of Europe, and throughout the UK and Ireland where it is more frequent in the south. The species is migratory and occurs commonly in parks, gardens, woodland edges and coastal areas. In the UK, adults are generally observed from March to November in two flight periods, with the peak of each flight period having shifted earlier by about 3 weeks over the past 30 years (Randle *et al.*, 2019). The larvae of *P. meticulosa* are polyphagous and have variable colouration ranging from bright green to dark brown, with colour being controlled by food plant in early larval instars and by multiple, as yet unidentified, genetic loci at later larval instars (Majerus, 1978). The genome of *P. meticulosa* 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 *P. meticulosa*, based on one female specimen from Wytham Woods, Oxfordshire, UK.

## Genome sequence report

The genome was sequenced from a single female *P. meticulosa* (ilPhlMeti2; Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338). A total of 45-fold

coverage in Pacific Biosciences single-molecule long reads (N50 12 kb) and 77-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 10 missing/misjoins, reducing the scaffold number by 14.81%, and increasing the scaffold N50 by 2.73%.

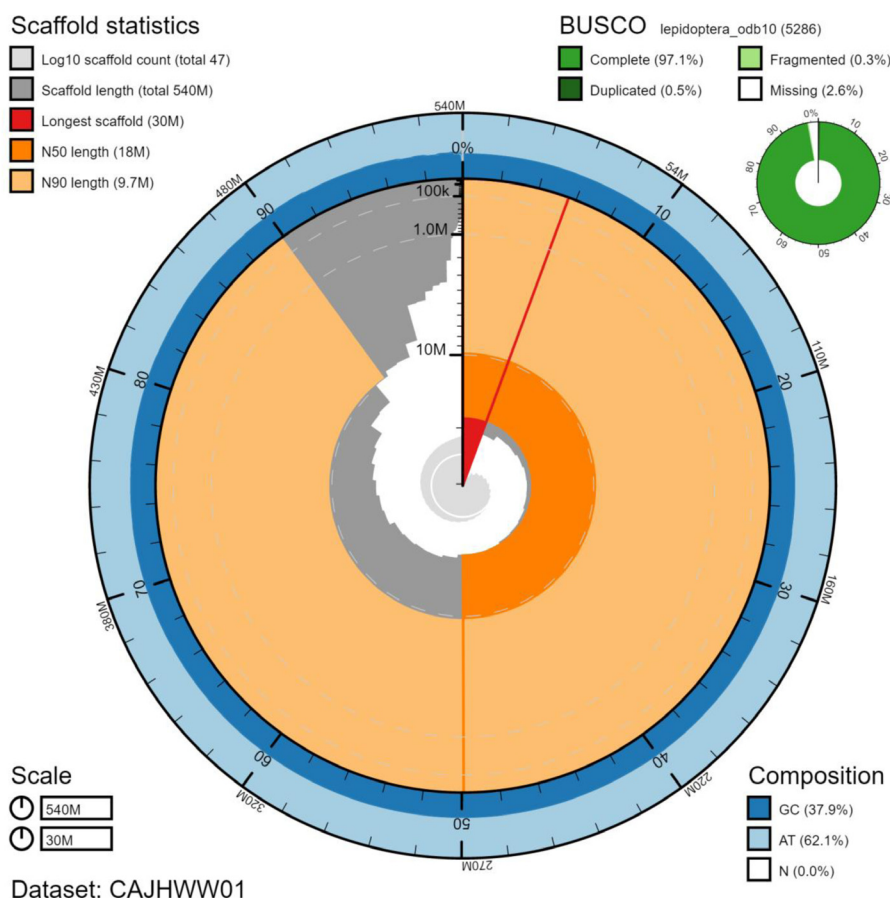
The final assembly has a total length of 539 Mb in 47 sequence scaffolds with a scaffold N50 of 18 Mb (Table 1). Of the assembly sequence, 95.2% was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 5; Table 2). Some unassigned scaffolds are compatible with belonging to the W chromosome based on half-depth read coverage, high repeat content and comparison with other Noctuidae W chromosomes. The assembly has a BUSCO (Manni *et al.*, 2021) completeness of 97.1% (single 96.7%, duplicated 0.5%) using the

**Table 1. Genome data for *Phlogophora meticulosa*, ilPhlMeti2.1.**

Project accession data	
Assembly identifier	ilPhlMeti2
Species	<i>Phlogophora meticulosa</i>
Specimen	ilPhlMeti2
NCBI taxonomy ID	NCBI:txid875884
BioProject	PRJEB41949
BioSample ID	SAMEA7520192
Isolate information	Female, head/abdomen/thorax
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6594496
10X Genomics Illumina	ERR6002732-ERR6002735
Hi-C Illumina	ERR6002736
PolyA RNA-Seq Illumina	ERR6002737, ERR6787418
Genome assembly	
Assembly accession	GCA_905147745.1
Accession of alternate haplotype	GCA_905147725.1
Span (Mb)	539
Number of contigs	61
Contig N50 length (Mb)	17
Number of scaffolds	47
Scaffold N50 length (Mb)	18
Longest scaffold (Mb)	20
BUSCO* genome score	C:97.5%,S:96.5%,D:1.0%,F:0.4%,M:2.1%,n:1658



**Figure 1. Image of the ilPhlMeti2 specimen taken prior to preservation and processing.** Specimen shown next to FluidX storage tube, 43.9 mm in length.



**Figure 2. Genome assembly of *Phlogophora meticulosa*, ilPhlMeti2.1: 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 538,659,867 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 (30,333,309 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (18,098,236 and 9,704,819 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/ilPhlMeti2.1/dataset/CAJHWW01/snail>.

lepidoptera\_odb10 reference set (n=5286). 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

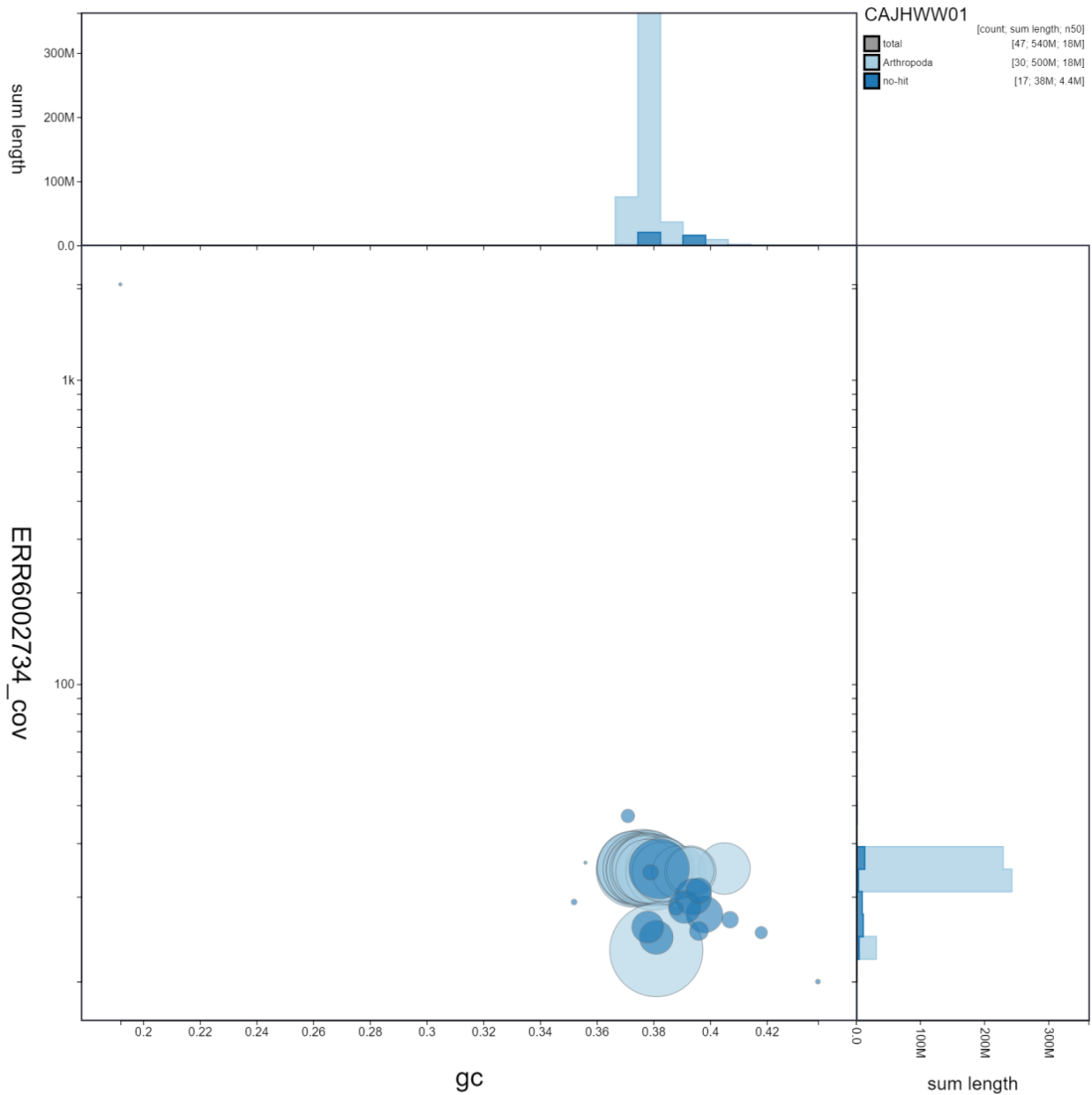
One female *P. meticulosa* (ilPhlMeti2) and one *P. meticulosa* of unknown sex (ilPhlMeti1) were collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338) by Douglas Boyes, UKCEH, using a light trap. The specimen was identified by the same individual and preserved on dry ice.

DNA was extracted from head/thorax tissue at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. RNA was extracted

from head/thorax and abdomen tissue of ilPhlMeti1 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.

### Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud 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. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on



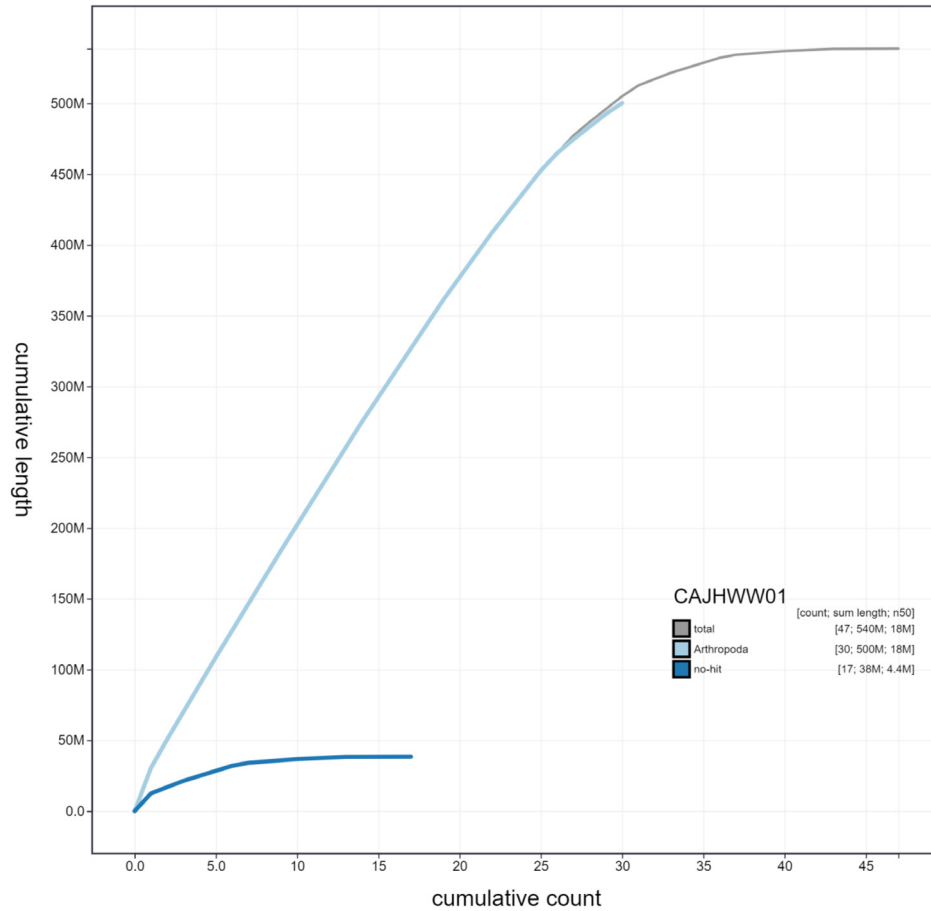
**Figure 3. Genome assembly of *Phlogophora meticulosa*, ilPhlMeti2.1: GC coverage.** BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilPhlMeti2.1/dataset/CAJHWW01/blob>.

Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated from abdomen tissue of ilPhlMeti2 using the Arima v1 Hi-C kit and sequenced on HiSeq X.

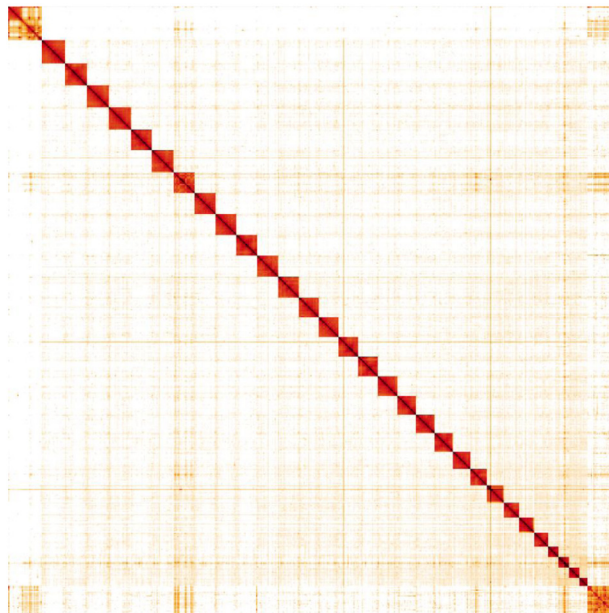
#### Genome assembly

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 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 (Howe *et al.*, 2021) was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome



**Figure 4. Genome assembly of *Phlogophora meticulosa*, ilPhlMeti2.1: 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/ilPhlMeti2.1/dataset/CAJHWW01/cumulative>.



**Figure 5. Genome assembly of *Phlogophora meticulosa*, ilPhlMeti2.1: Hi-C contact map.** Hi-C contact map of the ilPhlMeti2.1 assembly, visualised in HiGlass.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Phlogophora meticulosa*, ilPhlMeti2.1.**

INSDC accession	Chromosome	Size (Mb)	GC%
LR990518.1	1	20.34	37.6
LR990519.1	2	19.51	37.8
LR990520.1	3	19.41	37.3
LR990521.1	4	19.14	37.4
LR990522.1	5	18.92	37.8
LR990523.1	6	18.86	37.3
LR990524.1	7	18.80	38
LR990525.1	8	18.58	37.5
LR990526.1	9	18.45	37.7
LR990527.1	10	18.40	37.8
LR990528.1	11	18.17	37.3
LR990529.1	12	18.16	37.5
LR990530.1	13	18.10	37.6
LR990531.1	14	17.62	37.6
LR990532.1	15	17.29	37.7
LR990533.1	16	17.26	37.8
LR990534.1	17	17.07	38
LR990535.1	18	16.88	37.8
LR990536.1	19	16.09	38.2
LR990537.1	20	15.90	38.2
LR990538.1	21	15.50	37.7
LR990539.1	22	15.02	38.2
LR990540.1	23	14.52	38.3
LR990541.1	24	14.19	37.8
LR990542.1	25	12.45	38.2
LR990543.1	26	12.28	38.3
LR990544.1	27	9.70	38.9
LR990545.1	28	9.12	40.5
LR990546.1	29	8.99	39.3
LR990547.1	30	7.61	39.3
LR990517.1	Z	30.33	38.1
LR990548.1	MT	0.02	19.5
-	Unplaced	25.97	39

was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performed 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



**Table 3. Software tools used.**

Software tool	Version	Source
Hifiasm	0.12	<a href="#">Cheng et al., 2021</a>
purge_dups	1.2.3	<a href="#">Guan et al., 2020</a>
SALSA2	2.2	<a href="#">Ghurye et al., 2019</a>
longranger align	2.2.2	<a href="https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines">https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines</a>
freebayes	1.3.1-17-gaa2ace8	<a href="#">Garrison &amp; Marth, 2012</a>
MitoHiFi	1.0	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
gEVAL	0.1.x	<a href="#">Chow et al., 2016</a>
HiGlass	1.11.6	<a href="#">Kerpedjiev et al., 2018</a>
BlobToolKit	2.6.2	<a href="#">Challis et al., 2020</a>

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 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: *Phlogophora meticulosa* (angle shades). Accession number [PRJEB42141](#); <https://identifiers.org/ena.embl/PRJEB42141>.

The genome sequence is released openly for reuse. The *P. meticulosa* 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 the 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.5746938>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.5744972>.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: <https://doi.org/10.5281/zenodo.6125027>.

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.5638618>.

### 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](#)
- 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**, July, 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): 2896–2898.

[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](#)

Majerus MEN: **The Control of Larval Colour in Phlogophora Meticulosa L. (Lepidoptera: Noctuidae) and Some of Its Consequences**. University of

London. 1978.

[Reference Source](#)

Manni M, Berkeley MR, Seppely 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): 4647–54.

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

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain and Ireland's Larger Moths**. Pisces Publications, Newbury. 2019.

[Reference Source](#)

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): 1665–80.

[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](#)

Wootton R: **The Geometry and Mechanics of Insect Wing Deformations in Flight: A Modelling Approach**. *Insects.* 2020; **11**(7): 446.

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

# Open Peer Review

Current Peer Review Status:  

---

Version 1

Reviewer Report 22 June 2022

<https://doi.org/10.21956/wellcomeopenres.19653.r50941>

© 2022 Zhang J. 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.



Jing Zhang 

Eugene McDermott Center for Human Growth and Development, University of Texas  
Southwestern Medical Center, Dallas, TX, USA

The authors assembled a high-quality chromosome-level genome of angle shades moth with various sequencing techniques, including PacBio, 10X Genomics and Hi-C. The manuscript is well written. The only question I have is if authors also used 10X Genomics read data for scaffolding or generating contigs. From the manuscript, I noticed that 10X Genomics read data is only used for polishing the assembly and correcting small polymorphisms. But the advantage of 10X Genomics read data over common Illumina pair-end short-read data is to provide long-range genomic information and to help scaffold contigs. Conventional Illumina short read sequencing data is enough for fixing base errors (SNV/Indel)

**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:** Genomics, Computational biology

**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 01 April 2022

<https://doi.org/10.21956/wellcomeopenres.19653.r49197>

© 2022 Oppenheim S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Sara J. Oppenheim** 

Sackler Institute for Comparative Genomics,, American Museum of Natural History, New York, NY, USA

This is a very nicely done sequencing project and the assembly stats are excellent. The MS will be improved by adding a more detailed justification of why the sequencing was carried out and what sort of interesting biological phenomena might be explored. Such information would enhance the intro and allow for some discussion/conclusions.

The figures are largely useless and all of them are hard to read because the legends are super tiny.

Some further details are needed regarding the treatment of samples and which samples were used for which approach.

These issues are detailed in the annotated version of the MS that I have provided [here](#).

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

Partly

**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 biology; quantitative genetics; genome sequencing, assembly, and annotation; ecological adaptation in the Lepidoptera; 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.**

---