DATA NOTE

REVISED The genome sequence of the March moth, *Alsophila aescularia* (Denis & Schiffermüller)

[version 2; peer review: 3 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 Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK

[+] Deceased author

v2 **First published:** 19 Feb 2024, **9**:50
https://doi.org/10.12688/wellcomeopenres.20650.1
Latest published: 26 Sep 2025, **9**:50
https://doi.org/10.12688/wellcomeopenres.20650.2

Abstract

We present a genome assembly from an individual male *Alsophila aescularia* (the March moth; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 901.6 megabases in span. Most of the assembly is scaffolded into 14 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.67 kilobases in length. Gene annotation of this assembly on Ensembl identified 13,618 protein coding genes.

Keywords

Alsophila aescularia, March moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

4. **Daniel Doucet**, Natural Resources Canada, Ontario, Canada

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; **Holland PWH**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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 March moth, *Alsophila aescularia* (Denis & Schiffermüller) [version 2; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2025, **9**:50 https://doi.org/10.12688/wellcomeopenres.20650.2

First published: 19 Feb 2024, **9**:50 https://doi.org/10.12688/wellcomeopenres.20650.1

REVISED **Amendments from Version 1**

In Version 2 of this data note we have replaced the unlabelled chromosome map in Figure 5 with a labelled version, also showing a megabase scale. We have added a chromosome plot painted with ancestral Merian elements, illustrating the chromosomal fusions that have taken place in this lineage.
We have also updated the k-mer completeness results, using an updated version of Merqury.FK.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Alsophilinae; *Alsophila*; *Alsophila aescularia* (Denis & Schiffermüller) (NCBI:txid104486).

Background

Female flightlessness, with associated loss or reduction of wings, has evolved several times independently in Lepidoptera (Wahlberg *et al.*, 2010). The March moth *Alsophila aescularia* is a striking example, in which the highly active males have elongate wings and the flightless females completely lack external wings. At rest, the wings of the male are held overlapping and curled around the body; this is a very unusual resting position for moths in the family Geometridae although a similar posture is seen in some other lepidopteran families. The forewings of male *A. aescularia* have a grey-brown ground colour with a distinctive pale scalloped band towards the distal wing margin (Lawrence Sterne Trust, 2014; NBN Atlas Partnership, 2023).

A. aescularia is found across most of Europe, ranging from Scotland and Scandinavia to southern Italy, although there are relatively few records from the Iberian peninsula (GBIF Secretariat, 2023). There are also scattered records from Ukraine and Russia with an eastern limit at the Ural mountains (GBIF Secretariat, 2023). As reflected in the common name, the adult moth is active in spring; the majority of records from Britain and Ireland are from February to April, strongly peaking in March (NBN Atlas Partnership, 2023). Eggs of *A. aescularia* are laid in spring and the larvae feed through late spring and summer before over-wintering at the pupal stage. The larvae are polyphagous, feeding on leaves of deciduous trees including oak (*Quercus* sp.), sycamore (*Acer pseudoplatanus*), mountain ash (*Sorcus aucuparia*) and fruit trees. Occasionally, the species can become a pest in forestry or horticulture, for example in apple orchards in Bulgaria (Velcheva, 2009).

Here we report a complete genome sequence for the March moth *Alsophila aescularia* determined as part of the Darwin Tree of Life project. The genome sequence of *A. aescularia* will facilitate research into the evolution of sexual dimorphism in wing development and into adaptations to polyphagy, and will contribute to the growing set of resources for studying molecular evolution in the Lepidoptera.

Genome sequence report

The genome was sequenced from one male *Alsophila aescularia* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 23-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 12 missing joins or mis-joins and removed two haplotypic duplications, reducing the scaffold number by 12.5%.

The final assembly has a total length of 901.6 Mb in 28 sequence scaffolds with a scaffold N50 of 67.4 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.93%) of the assembly sequence was assigned to 14 chromosomal-level scaffolds, representing 13 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. Chromosome painting with Merian elements illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to the ancestral linkage groups of Lepidoptera (Figure 6).

The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The combined primary and alternate assemblies achieve an estimated QV of 61.5. The *k*-mer completeness is 85.12% for

Figure 1. Photograph of the *Alsophila aescularia* (ilAlsAesc1) specimen used for genome sequencing.

Table 1. Genome data for *Alsophila aescularia*, ilAlsAesc1.1.

Project accession data		
Assembly identifier	ilAlsAesc1.1	
Species	*Alsophila aescularia*	
Specimen	ilAlsAesc1	
NCBI taxonomy ID	104486	
BioProject	PRJEB54803	
BioSample ID	SAMEA10107017	
Isolate information	ilAlsAesc1, male: thorax (DNA), abdomen (RNA), head (Hi-C)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	Primary: 61.0; alternate: 61.8; combined: 61.5	*≥ 40*
k-mer completeness	Primary: 85.12%; alternate: 84.21%; combined: 99.14%	*≥ 95%*
BUSCO**	C:98.6%[S:97.8%,D:0.8%], F:0.3%,M:1.0%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.93%	*≥ 90%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR9981094	
Hi-C Illumina	ERR9988136	
PolyA RNA-Seq Illumina	ERR10378023	
Genome assembly		
Assembly accession	GCA_946251855.1	
Accession of alternate haplotype	GCA_946251895.1	
Span (Mb)	901.6	
Number of contigs	140	
Contig N50 length (Mb)	13.3	
Number of scaffolds	28	
Scaffold N50 length (Mb)	67.4	
Longest scaffold (Mb)	107.3	
Genome annotation		
Number of protein-coding genes	13,618	
Number of non-coding genes	2,364	
Number of gene transcripts	26,110	

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

** BUSCO scores based on the lepidoptera_odb10 BUSCO set using 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/Alsophila%20aescularia/dataset/CAMIUD01/busco.

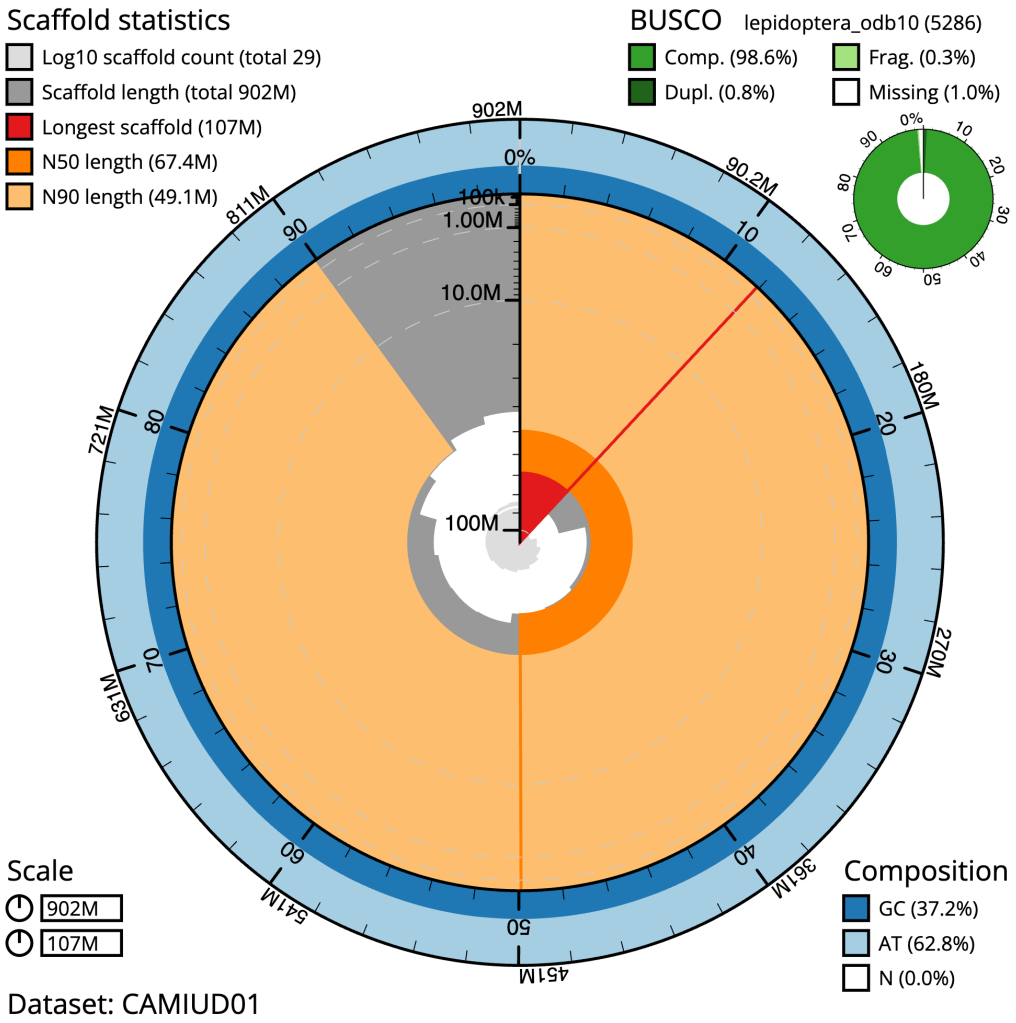

Figure 2. Genome assembly of *Alsophila aescularia*, ilAlsAesc1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 bins around the circumference with each bin representing 0.1% of the 901,612,816 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (107,309,733 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (67,363,832 and 49,117,080 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the lepidoptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Alsophila%20aescularia/dataset/CAMIUD01/snail.

the primary assembly, 84.21% for the alternate haplotype, and 99.14% for the combined assemblies. The primary assembly has a BUSCO v5.3.2 completeness of 98.6% (single = 97.8%, duplicated = 0.8%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/104486.

Genome annotation report

The *Alsophila aescularia* genome assembly (GCA_946251855.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Alsophila_aescularia_GCA_946251855.1/Info/Index). The resulting annotation includes 26,110 transcribed mRNAs from 13,618 protein-coding and 2,364 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Alsophila aescularia* (specimen ID Ox001094, ToLID ilAlsAesc1) was collected in a light trap from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2021-03-31. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

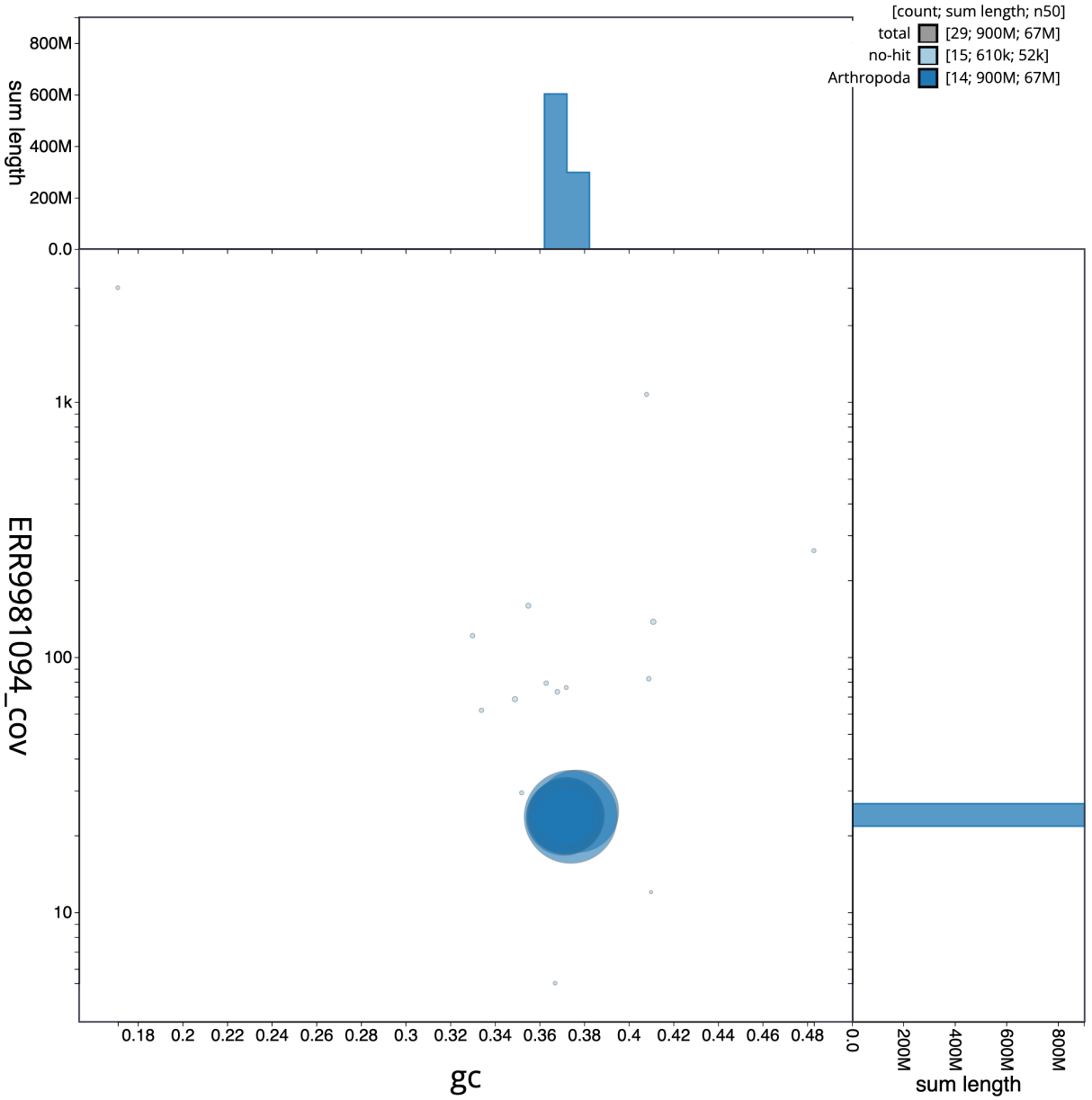

Figure 3. Genome assembly of *Alsophila aescularia*, ilAlsAesc1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Alsophila%20aescularia/dataset/CAMIUD01/blob.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up, and the full protocols have been deposited in the protocols.io repository (Denton *et al.*, 2023b). The sample was prepared for DNA extraction at the WSI Tree of Life laboratory: the ilAlsAesc1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023), with tissue set aside for Hi-C sequencing. Tissue from the thorax was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). DNA was extracted at the WSI Scientific Operations core using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): In brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA

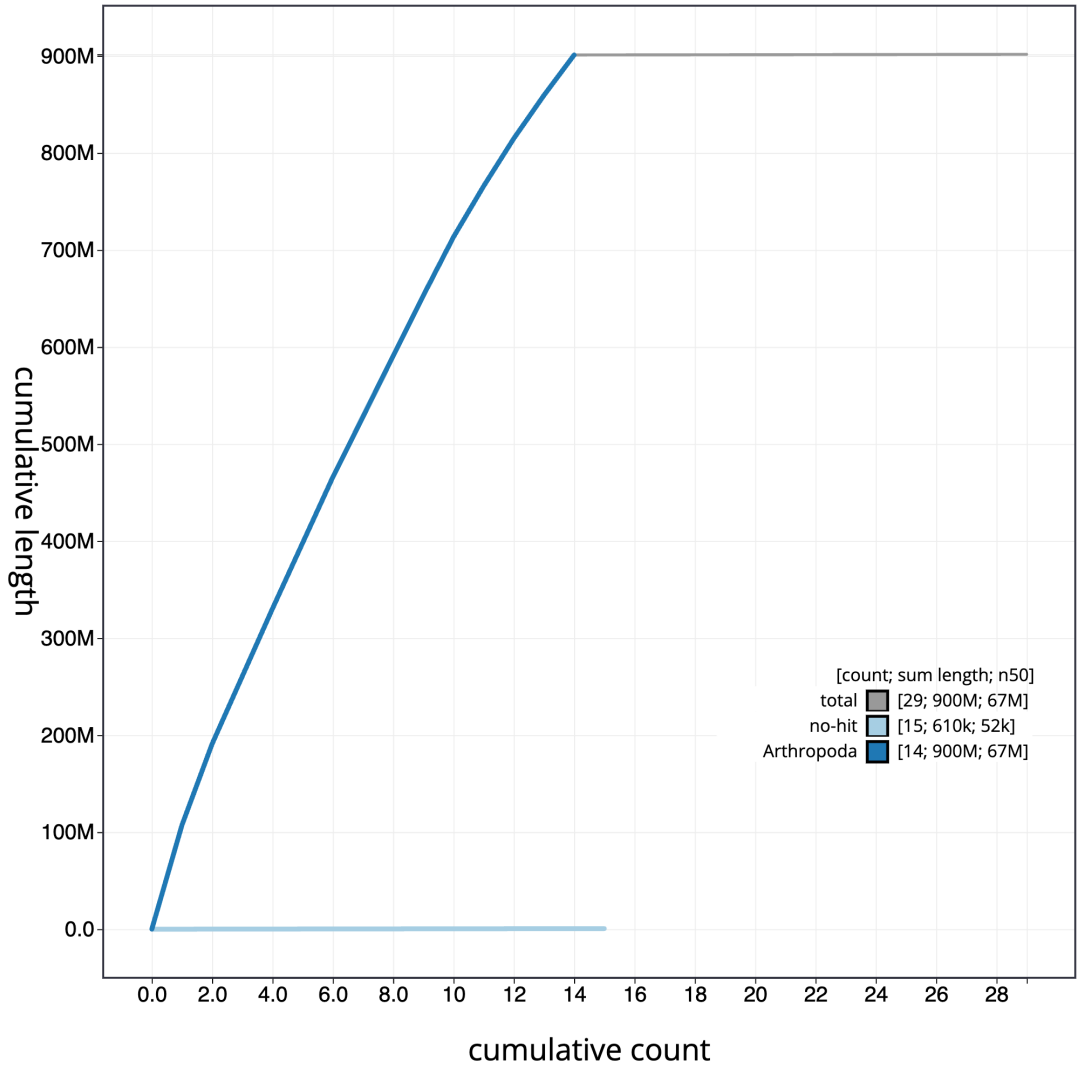

Figure 4. Genome assembly of *Alsophila aescularia*, ilAlsAesc1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Alsophila%20aescularia/dataset/CAMIUD01/cumulative.

High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from abdomen tissue of ilAlsAesc1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was 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 the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing were performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from head tissue of ilAlsAesc1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly of PacBio HiFi reads was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described

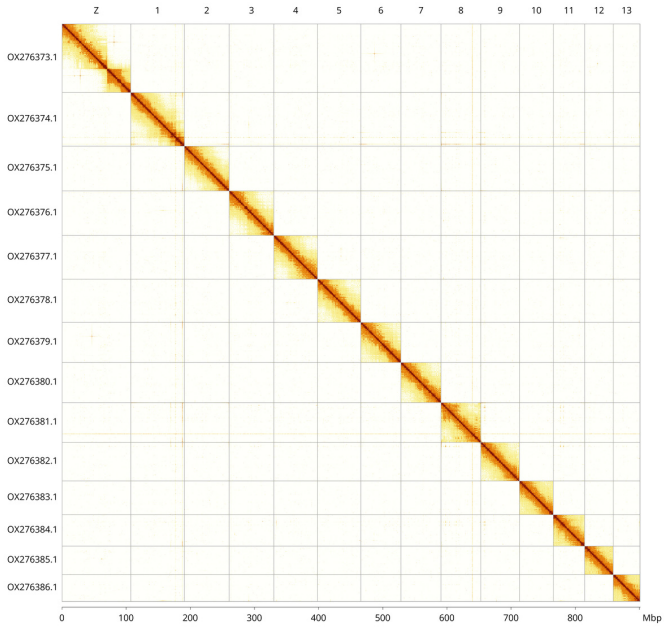

Figure 5. Genome assembly of *Alsophila aescularia*, ilAlsAesc1.1: Hi-C contact map of the ilAlsAesc1.1 assembly, visualised using PretextView and Pretext Snapshot. Chromosomes are shown in decreasing order of size from left to right and top to bottom and are labelled along the axes, with a megabase scale on the bottom axis. An interactive version of this figure in HiGlass may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Zzm2cXQWQAyyvZr4TFTIsA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Alsophila aescularia*, ilAlsAesc1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX276374.1	1	83.69	37.5
OX276375.1	2	69.7	37.0
OX276376.1	3	69.65	37.0
OX276377.1	4	68.12	37.0
OX276378.1	5	67.36	37.0
OX276379.1	6	62.72	37.5
OX276380.1	7	62.37	37.0
OX276381.1	8	62.06	37.0
OX276382.1	9	60.34	37.0
OX276383.1	10	52.48	37.0
OX276384.1	11	49.12	37.0
OX276385.1	12	44.52	37.5
OX276386.1	13	41.58	37.0
OX276373.1	Z	107.31	37.5
OX276387.1	MT	0.02	17.0

previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Chromosomal painting was performed using lep_busco_painter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths calculated using NCBI datasets. Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

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

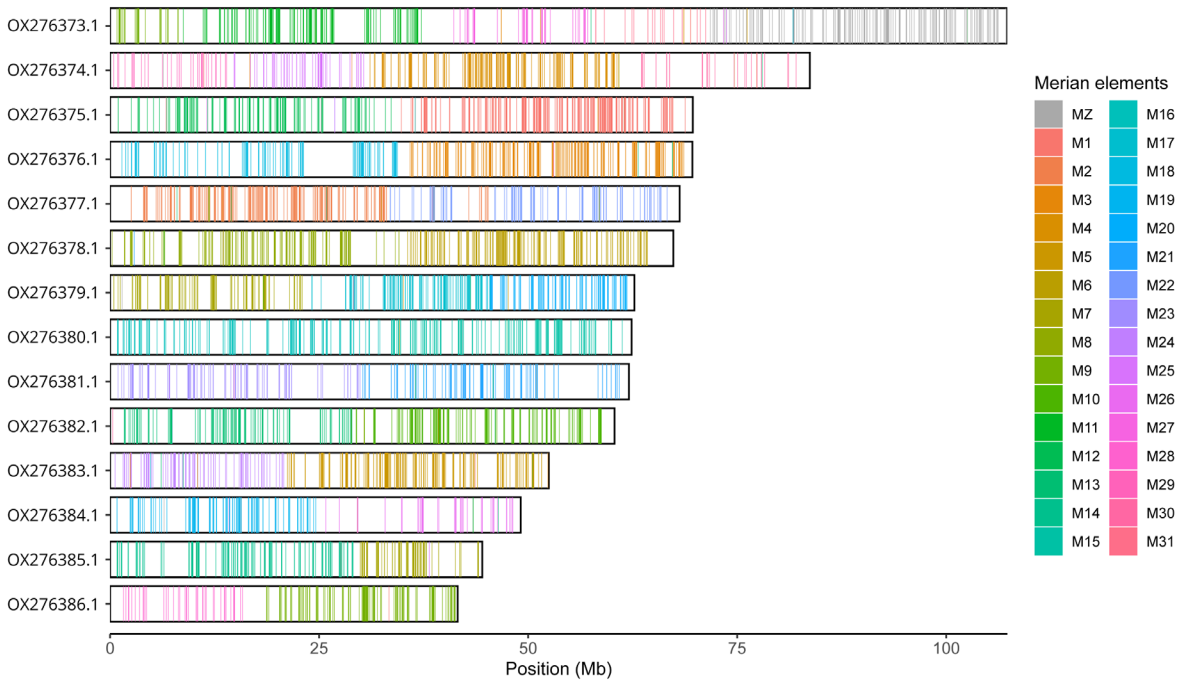

Figure 6. Merian elements painted across the chromosomes in the primary assembly of *Alsophila aescularia*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Lep_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter
Merqury.FK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2.5	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Alsophila aescularia* assembly (GCA_946251855.1). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website here. By agreeing with and signing up

to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

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

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

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

Data availability

European Nucleotide Archive: *Alsophila aescularia* (March moth). Accession number PRJEB54803; https://identifiers.org/ena.embl/PRJEB54803 (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The *Alsophila aescularia* 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.

Author information

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

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

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

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

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

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

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

References

Abdennur N, Mirny LA: **Cooler: Scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
PubMed Abstract | Publisher Full Text | Free Full Text

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): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

do Amaral RJV, Denton A, Yatsenko H, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

GBIF Secretariat: ***Alsophila aescularia* (Denis & Schiffermüller) 1775**. *GBIF Backbone Taxonomy.* 2023; [Accessed 15 November 2023].
Reference Source

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Lawrence Sterne Trust: **A wingless moth**. 2014; [Accessed 15 November 2023].
Reference Source

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**

NBN Atlas Partnership: ***Alsophila aescularia* ([Denis & Schiffermüller], 1775)**. March Moth, *NBN Atlas*, 2023; [Accessed 15 November 2023].
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): 1665–1680.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | **Publisher Full Text**

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: Manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *BMC Bioinformatics.* 2023**; 24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Velcheva N: **Lepidopterian Fauna in New Planted Apple Orchard in West Bulgaria.** *Biotechnol Biotechnol Equip.* 2009; **23**(sup1): 127–131.
Publisher Full Text

Wahlberg N, Snäll N, Viidalepp J**,** *et al.*: **The evolution of female flightlessness among Ennominae of the Holarctic forest zone (Lepidoptera, Geometridae).** *Molec Phylogenet Evol.* 2010; **55**(3): 929–938.
PubMed Abstract | **Publisher Full Text**

Wellcome Sanger Institute: **The genome sequence of the March moth, *Alsophila aescularia* (Denis & Schiffermüller).** European Nucleotide Archive. [dataset], accession number PRJEB54803, 2022.

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–790.
PubMed Abstract | **Publisher Full Text**| **Free Full Text**

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ? ✓ ✓ ✓

Version 2

Reviewer Report 06 November 2025

https://doi.org/10.21956/wellcomeopenres.27545.r134781

✓ **Daniel Doucet**
Natural Resources Canada, Ontario, Canada

The authors describe the complete sequencing and assembly of the genome from the March moth, *Alsophila aescularia.* The moth belongs to one of the most speciose family of the order Lepoidoptera, the geometridae, and as such this genome provides critical insight into the evolution of this important taxa.

The rationale for creating the dataset is clearly described. *A. aescularia* is not only present in the UK but also widespread across Europe, thus providing foundational data for future population genomics studies that factor in variables such as climate and geography. The species being polyphagous, host range, host range x climate etc. interactions can also be studied through the lens of population genomics.

Adequate details on the material and methods of genome sequencing and assembly have been given by the authors. Sequencing methods (PacBio HiFi and Hi-C sequencing), assembly software and workflows are described succinctly, but with enough details such that other researchers with bioinformatics training can replicate the study.

The datasets provided are standard for Wellcome Open genome report publications, including basic genome assembly statistics (genome length, scaffold N50, BUSCO score, etc.). The mapping of the Merian elements is a nice addition to this type of report, as they allow quick comparison of ancestral linkage groups between lepidopteran sequenced genomes.

One suggestion I would have for the authors, if they wanted to extend the reach of their report, is to compare the number of *roadkill* (*rdx*) genes in the genome of *A. aescularia* with that of another geometrid with wingless females, the winter moth (*Operophtera brumata*). Four *rdx* orthology groups have been identified in insects, and *O. brumea* displays a large and specific expansion of 25 genes in one particular group, possibly linking *rdx* function with brachyptery. It would be interesting to know if such an expansion is conserved among brachypterous Geometrids. In *Drosophila melanogaster*, *rdx* genes play a role in organ growth and patterning, particularly in the eye and *rdx* overexpression has been observe to lead to smaller wing size (see citation).

References
1. Derks M, Smit S, Salis L, Schijlen E, et al.: The Genome of Winter Moth (Operophtera brumata) Provides a Genomic Perspective on Sexual Dimorphism and Phenology. *Genome Biology and Evolution*. 2015; **7** (8): 2321-2332 Publisher Full Text

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

Are the protocols appropriate and is the work technically sound?
Yes

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Insect cell 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 06 November 2025

https://doi.org/10.21956/wellcomeopenres.27545.r134780

Sophie Tandonnet
[1] Universitat de Barcelona, Barcelona, Catalonia, Spain
[2] Universitat de Barcelona Institut de Recerca de la Biodiversitat (Ringgold ID: 535620), Barcelona, Catalonia, Spain

This note presents a high-quality, chromosome-scale genome assembly of the unusual moth *Alsophila aescularia*, a species notable for its extreme sexual dimorphism and its reduced chromosome number (14 compared to the typical 31 in other lepidopteran species). The mechanisms of sex determination and differentiation, as well as the broader biology of this species, are likely to be particularly intriguing, and this genome will represent a valuable resource for the research community.
In agreement with the other reviewers, I regret that the heterogametic sex was not sequenced. The Z chromosome was identified based on “homologous pairs,” and while it does appear to

contain the majority of Merian element MZ, chromosome assignment by homology (the precise method of which is not clearly described here) can be unreliable. I would therefore advise caution and recommend that the authors explicitly state that the sex chromosome assignment remains unverified and may be incorrect.

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: Evolutionary Biology, non-model organisms, genomics, sex determination systems

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 29 September 2025

https://doi.org/10.21956/wellcomeopenres.27545.r134416

Jeffrey Marcus
University of Manitoba, Winnipeg, Manitoba, Canada

I thank the authors for responding to the reviews. They appear to have adequately addressed the critiques of my previous review.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary biology of insects, phylogenomics

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.

Version 1

Reviewer Report 05 July 2024

https://doi.org/10.21956/wellcomeopenres.22854.r86299

?

Jeffrey Marcus
University of Manitoba, Winnipeg, Manitoba, Canada

In this manuscript, the authors describe the sequencing and assembly of the *Alsophila aescularia* genome using DNA from a single male specimen collected in the UK. The primary genome sequence assembly includes proposed chromosomal pseudomolecule sequences for 13 autosomes, the Z sex chromosome, and a complete mitochondrial genome. No females were collected, so the W sex chromosome was not sequenced. Overall, this is a reasonably solid contribution to the scientific literature, but due the absence of a W chromosome assembly, the data presented will be limited in its application to "facilitate research into the evolution of sexual dimorphism in wind development", one of the purported rationales for assembling this genome.

Some suggestions to the authors:

1. Methods: The date of collection for the sequenced *Alsophila aescularia* specimen is not included in the manuscript. As I have previously noted in a review of a different paper by these same authors [Marcus J. Peer Review Report For: The genome sequence of the bramble shoot moth, Notocelia uddmanniana (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Res 2021, 6:348 (https://doi.org/10.21956/wellcomeopenres.19338.r79177)], "it is best practice to include collection dates for all wild-caught insect specimens used for genome sequencing because it is not unusual for cryptic species pairs to be distinguishable by differences in phenology of adult emergence. If it should come to pass that genus [*Alsophila*] includes cryptic species, knowing the specimen collection date can help future researchers determine to which taxon this sequenced genome should be assigned". The collection date of the specimen should be added in a revision of this manuscript.
2. And as I have also recommended before, "For future work, I suggest that the researchers preferentially sequence the heterogametic sex when assembling genomes for previously unstudied species (in the case of Lepidoptera, the heterogametic sex is female), so that draft assembles can be prepared for both sex chromosomes." This is particularly important if authors hope that their work will be used by those who wish to conduct meaningful studies of the genetic basis of sexual dimorphism.
3. I am in agreement with another review of this manuscript by DePrins [De Prins J. Peer Review Report For: The genome sequence of the March moth, Alsophila aescularia (Denis & Schiffermüller) [version 1; peer review: 1 approved with reservations]. Wellcome Open Res 2024, 9:50 (https://doi.org/10.21956/wellcomeopenres.22854.r78079)] which describes the reconstruction of a genome with 13 autosomes, plus the Z chromosome to be an exceptionally noteworthy finding. Most Lepidoptera have a haploid chromosome number

of 29 (28 autosomes, plus the Z/W sex chromosome), with deeply conserved synteny across lepidopteran taxonomy [Wright, C.J., Stevens, L., Mackintosh, A. *et al.* Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* **8**, 777–790 (2024). https://doi.org/10.1038/s41559-024-02329-4]. DePrins has suggested imaging a meiotic plate from this species to confirm the genome structure. That would be helpful, but even a bioinformatics approach to analyzing the existing data set, mapping the sequence synteny between the assembled chromosomes of *Alsophila aescularia* and those of a lepidopteran with a more conventional haploid chromosome number would do much to further our understanding of the chromosomal fusions that have occurred in this instance. I encourage the authors to conduct this synteny analysis in a revision of this manuscript.

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 biology of insects, phylogenomics, evolution and development of insect wings

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.

Author Response 21 Sep 2025

Tree of Life Team Sanger

Thank you for reviewing this data note. Your comments are addressed below.

1. **Methods: The date of collection for the sequenced *Alsophila aescularia* specimen is not included in the manuscript. As I have previously noted in a review of a different paper by these same authors [Marcus J. Peer Review Report For: The genome sequence of the bramble shoot moth, Notocelia uddmanniana (Linnaeus, 1758) [version 1; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Res 2021, 6:348 (https://doi.org/10.21956/wellcomeopenres.19338.r79177)], "it is best practice to include collection dates for all wild-caught insect specimens used for genome sequencing because it is not unusual for cryptic species pairs to be distinguishable by differences in phenology of adult emergence. If it should**

come to pass that genus [*Alsophila*] includes cryptic species, knowing the specimen collection date can help future researchers determine to which taxon this sequenced genome should be assigned". The collection date of the specimen should be added in a revision of this manuscript.

Response: The collection date is given in the Methods section already. Note that the Methods comes after the Genome sequence report and the sample collection information is given in detail under the subheading "Sample acquisition and nucleic acid extraction".

2. And as I have also recommended before, "For future work, I suggest that the researchers preferentially sequence the heterogametic sex when assembling genomes for previously unstudied species (in the case of Lepidoptera, the heterogametic sex is female), so that draft assembles can be prepared for both sex chromosomes." This is particularly important if authors hope that their work will be used by those who wish to conduct meaningful studies of the genetic basis of sexual dimorphism.

Response: We agree that it is most useful to sequence the heterogametic sex, and request that collectors provide these as far as possible.

3.I am in agreement with another review of this manuscript by DePrins [De Prins J. Peer Review Report For: The genome sequence of the March moth, Alsophila aescularia (Denis & Schiffermüller) [version 1; peer review: 1 approved with reservations]. Wellcome Open Res 2024, 9:50 (https://doi.org/10.21956/wellcomeopenres.22854.r78079)] which describes the reconstruction of a genome with 13 autosomes, plus the Z chromosome to be an exceptionally noteworthy finding. Most Lepidoptera have a haploid chromosome number of 29 (28 autosomes, plus the Z/W sex chromosome), with deeply conserved synteny across lepidopteran taxonomy [Wright, C.J., Stevens, L., Mackintosh, A. *et al.* Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* 8, 777–790 (2024). https://doi.org/10.1038/s41559-024-02329-4]. DePrins has suggested imaging a meiotic plate from this species to confirm the genome structure. That would be helpful, but even a bioinformatics approach to analyzing the existing data set, mapping the sequence synteny between the assembled chromosomes of *Alsophila aescularia* and those of a lepidopteran with a more conventional haploid chromosome number would do much to further our understanding of the chromosomal fusions that have occurred in this instance. I encourage the authors to conduct this synteny analysis in a revision of this manuscript.

Response: We have added a Merian plot (Figure 6) to show how ancestral single-copy orthologues map to the chromosomes here, illustrating the chromosome fusions that have taken place in this lineage.

Competing Interests: No competing interests were disclosed.

Reviewer Report 18 April 2024

https://doi.org/10.21956/wellcomeopenres.22854.r78079

?

urate De Prins
Royal Belgian Institute of Natural Sciences, Brussels, Belgium

The correct identification of the specimen that is sequenced with full genome methodology is the first, but essential step of any further study. We need correct evidence-based data, and not the long-read sequences that are obtained of doubtful or even misidentified species. Lepidoptera are identified by the combination of multi-evidence characters that are obtained from external morphology, internal micromorphology, bionomics, and DNA. Some species of Lepidoptera can be identified from external morphology only, but most of them are not. Therefore, it is a very desirable methodological approach in the Tree of Life initiative to include the full possible evidence of correct species identification.

In this particular case, the species *Alsophila aescularia* is correctly identified by external morphology only. The specimen presented in Figure1 certainly belongs to *Alsophila aescularia.* This species is particular by its extreme sexual dimorphism. Females are wingless. This phenomenon no doubt has an expression in genomics.
The other super rare fact is that the chromosome number of this species is n=14! This is very rare in moths (Heterocera) and there is no evidence presented in the paper except the part of the sentence "the assembly sequence was assigned to 14 chromosomal-level scaffolds, representing 13 autosomes and the Z sex chromosome".

Suggestions:
- Maybe the authors could present a meiotic plate or other visual evidence of such a rare case of chromosome arrangement in Lepidoptera. Is this related to extreme sexual dimorphism or do other mainly genetic factors play a role?

Questions:
- Is there any genetic factor found in full genome sequence data for extreme sexual dimorphism in *Alsophila aescularia?*
- Is the extreme fusion of chromosome set in *Alsophila aescularia* which is more than 2 times less than what is a usual case in moths, could be explained somehow by genetic data?

It is likely that these two important evolutionary questions can be answered in further studies.

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: Karyology of primitive moths, taxonomy, nomenclature, diversity of tropical taxa of primitive moths, moth-plant relationships, 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, however I have significant reservations, as outlined above.

Author Response 21 Sep 2025

Tree of Life Team Sanger

Thank you for reviewing this data note. Your comments are addressed below.

Suggestions:

- Maybe the authors could present a meiotic plate or other visual evidence of such a rare case of chromosome arrangement in Lepidoptera. Is this related to extreme sexual dimorphism or do other mainly genetic factors play a role?

Response*: We have added a Merian plot (Figure 6) to show how ancestral single-copy orthologues map to the chromosomes here, illustrating the chromosome fusions that have taken place in this lineage.*

Questions:

- Is there any genetic factor found in full genome sequence data for extreme sexual dimorphism in *Alsophila aescularia?*
- Is the extreme fusion of chromosome set in *Alsophila aescularia* which is more than 2 times less than what is a usual case in moths, could be explained somehow by genetic data?

It is likely that these two important evolutionary questions can be answered in further studies.

Response*: We agree, and hope that the dataset provided here will be a useful resource for these studies.*

Competing Interests: No competing interests were disclosed.