




DATA NOTE

The genome sequence of the mottled umber, *Erannis defoliaria* (Clerk, 1759) [version 1; peer review: awaiting peer review]

Douglas Boyes¹⁺,
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, Thomas Lewin ²,
Darwin Tree of Life Consortium

¹UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK

²Department of Zoology, University of Oxford, Oxford, UK

+ Deceased author

V1 First published: 11 Aug 2022, 7:208
<https://doi.org/10.12688/wellcomeopenres.17815.1>
Latest published: 11 Aug 2022, 7:208
<https://doi.org/10.12688/wellcomeopenres.17815.1>

Abstract

We present a genome assembly from an individual male *Erannis defoliaria* (the mottled umber; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 522 megabases in span. The majority of the assembly (99.93%) is scaffolded into 29 chromosomal pseudomolecules, with the Z sex chromosome assembled. The mitochondrial genome was also assembled, and is 13.1 kilobases in length.

Keywords

Erannis defoliaria, mottled umber, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status Awaiting Peer Review

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; **Lewin T:** Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust 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.

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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the mottled umber, *Erannis defoliaria* (Clerk, 1759) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2022, 7:208 <https://doi.org/10.12688/wellcomeopenres.17815.1>

First published: 11 Aug 2022, 7:208 <https://doi.org/10.12688/wellcomeopenres.17815.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Geometroidea; Geometridae; Ennominae; Erannis; *Erannis defoliaria* (Clerk, 1759) (NCBI:txid104474).

Background

The mottled umber (*Erannis defoliaria*), first described in 1759 by Carl Alexander Clerk, is a moth of the family Geometridae (Skinner & Wilson, 2009). It is distributed across much of the Palearctic, including throughout Britain and Ireland, where it is most commonly observed in southern England and Wales (Randle *et al.*, 2019), and is found on the wing from October to December (Skinner & Wilson, 2009). The species is of present interest because large declines are being observed in Britain, with a dramatic 86% abundance decrease being reported from 1970 to 2016 (Randle *et al.*, 2019).

Notably, female *E. defoliaria* are micropterous and flightless, retaining only vestigial wings which manifest as small lobes (Sattler, 1991) and, though they are not completely apterous, they are commonly referred to as wingless. These wingless females initiate reproduction by ascending trees and producing a potent male-attracting pheromone (Hansson *et al.*, 1990). The evolution of wing reduction and winglessness has occurred independently many times across the insects and particularly within the Lepidoptera, with at least 20% of moth families containing flightless, wing-reduced members (Sattler, 1991; Wagner & Liebherr, 1992); the wealth of data produced by the Darwin Tree of Life project may be used to uncover the genetic factors regulating this wingless phenotype.

Like many flightless moth species, the mottled umber is polyphagous, and it utilises a wide range of habitats including woodland and suburban areas. In much of Europe it is considered a pest, and is a key contributor to the defoliation of oak (*Quercus*) forests (Blažek *et al.*, 2021; Glavendekić & Medarević, 2010; Netoiu *et al.*, 2015; Sarvašová *et al.*, 2020); dramatic overpopulation and defoliation events occur on 7 to 11 year cycles, and these are becoming increasingly severe with climate change (Hittenbeck *et al.*, 2019).

Genome sequence report

The genome was sequenced from one male *E. defoliaria* (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude -1.338). A total of 23-fold coverage in Pacific Biosciences single-molecule long reads and 75-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 62 missing/misjoins and removed 17 haplotypic duplications, reducing the assembly size by 0.43% and the scaffold number by 41.27%, and increasing the scaffold N50 by 8.57%.

The final assembly has a total length of 522 Mb in 37 sequence scaffolds with a scaffold N50 of 19.2 Mb (Table 1). The majority of the assembly sequence (99.93%) was assigned to 29 chromosomal-level scaffolds, representing 28 autosomes



Figure 1. Image of the *Erannis defoliaria* (ilEraDefo1) specimen taken during preservation and processing.

Table 1. Genome data for *Erannis defoliaria*, ilEraDefo1.1.

Project accession data	
Assembly identifier	ilEraDefo1.1
Species	<i>Erannis defoliaria</i>
Specimen	ilEraDefo1
NCBI taxonomy ID	NCBI:txid104474
BioProject	PRJEB43796
BioSample ID	SAMEA7520367
Isolate information	Male, thorax/abdomen
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6560798
10X Genomics Illumina	ERR6054601-ERR6054604
Hi-C Illumina	ERR6054605-ERR6054607
Genome assembly	
Assembly accession	GCA_905404285.1
Accession of alternate haplotype	GCA_905404195.1
Span (Mb)	522
Number of contigs	90
Contig N50 length (Mb)	11.4
Number of scaffolds	37
Scaffold N50 length (Mb)	19.2
Longest scaffold (Mb)	26.9
BUSCO* genome score	C:98.4%[S:98.0%,D:0.4%], F:0.5%,M:1.1%,n:5286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.2.2. 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/ilEraDefo1.1/dataset/CAJQFK01.1/busco>.

(numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.2.2 (Manni *et al.*, 2021) completeness of 98.4% (single 98.0%, duplicated 0.4%) 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.

Methods

Sample acquisition and DNA extraction

A single male *E. defoliaria* (ilEraDefo1) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude -1.338) by Douglas Boyes, UKCEH, using a light trap in woodland. The sample was identified by the same individual, and preserved on dry ice.

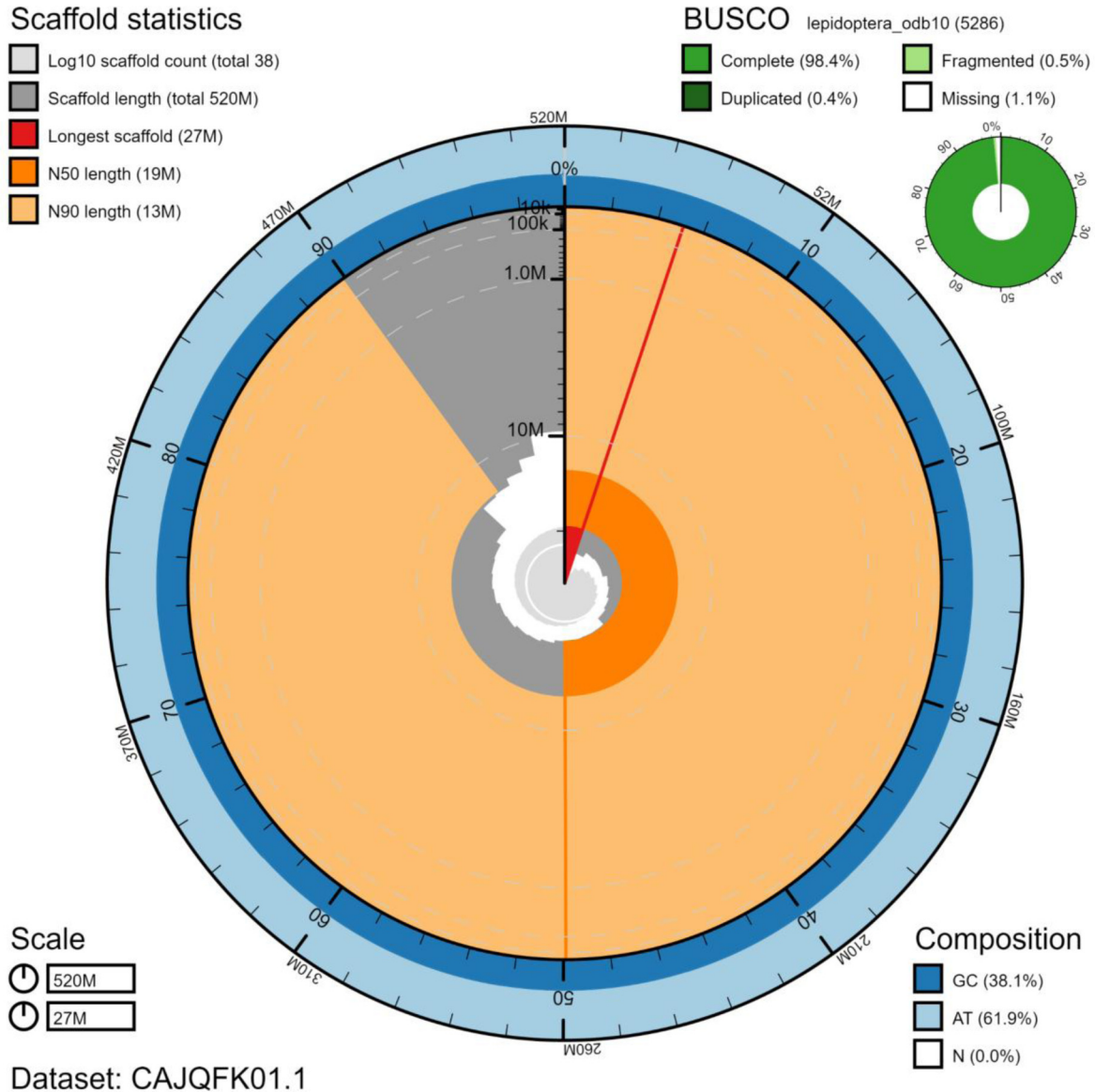


Figure 2. Genome assembly of *Erannis defoliaria*, ilEraDefo1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 522,369,616 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (26,948,889 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (19,179,966 and 13,177,765 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/ilEraDefo1.1/dataset/CAJQFK01.1/snail>.

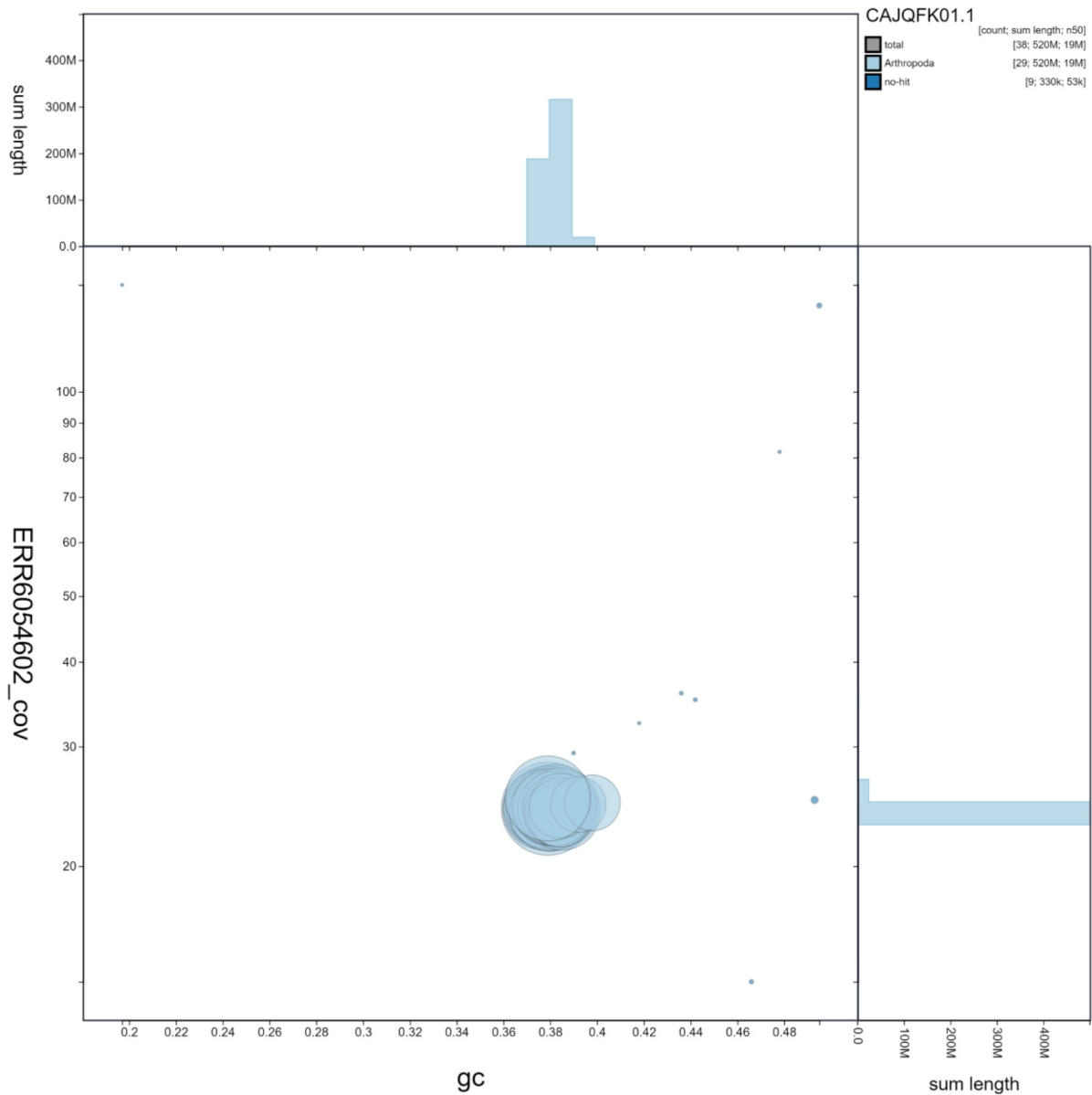


Figure 3. Genome assembly of *Erannis defoliaria*, iEraDefo1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iEraDefo1.1/dataset/CAJQFK01.1/blob>.

DNA was extracted from thorax/abdomen tissue of iEraDefo1 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.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome

Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina HiSeq X (10X) instruments. Hi-C data were generated from remaining thorax/abdomen tissue using the Arima Hi-C v1 kit and sequenced on a HiSeq X instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021); haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly

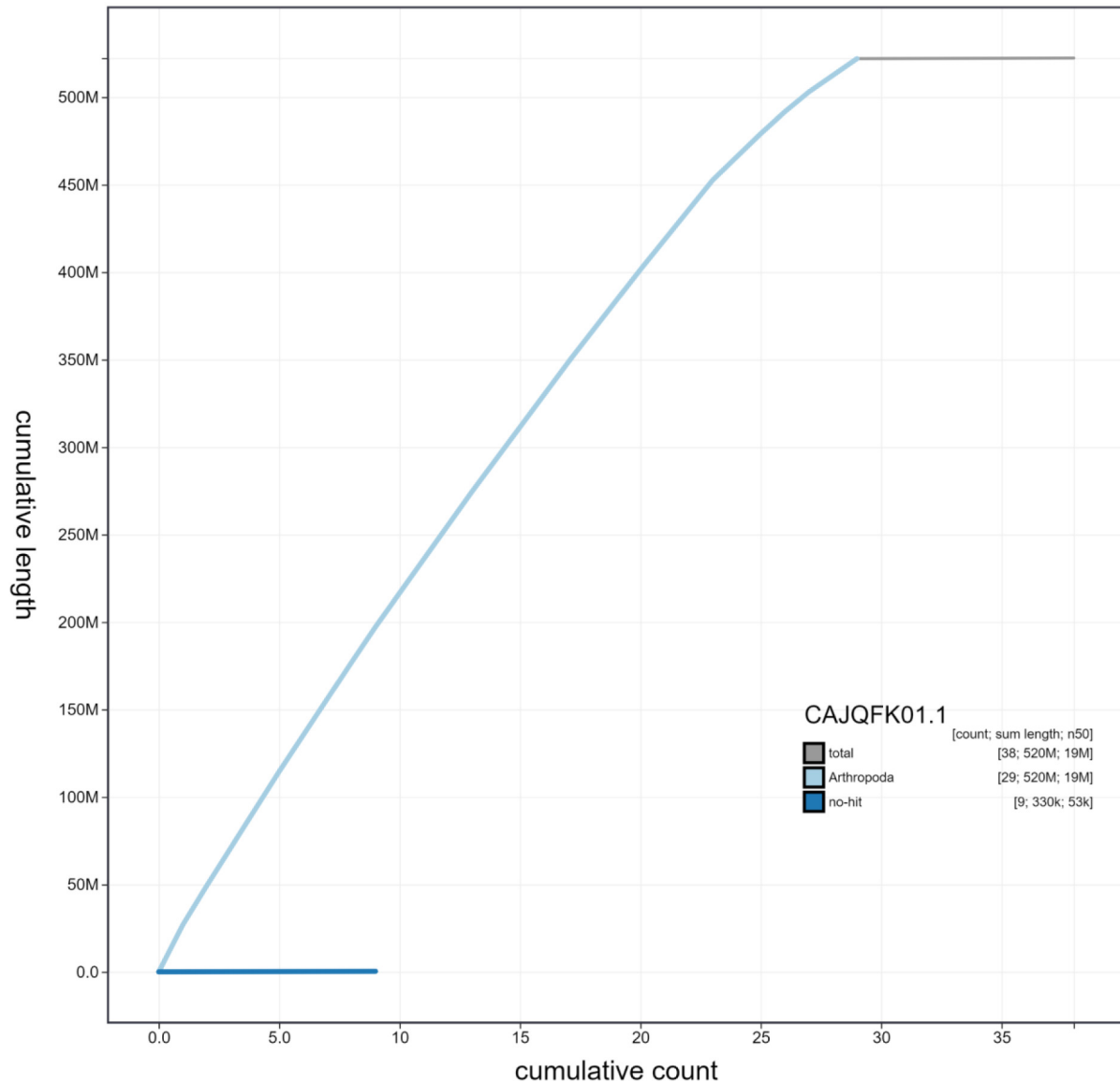


Figure 4. Genome assembly of *Erannis defoliaria*, ilEraDefo1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilEraDefo1.1/dataset/CAJQFK01.1/cumulative>.

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 gEVAL (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 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

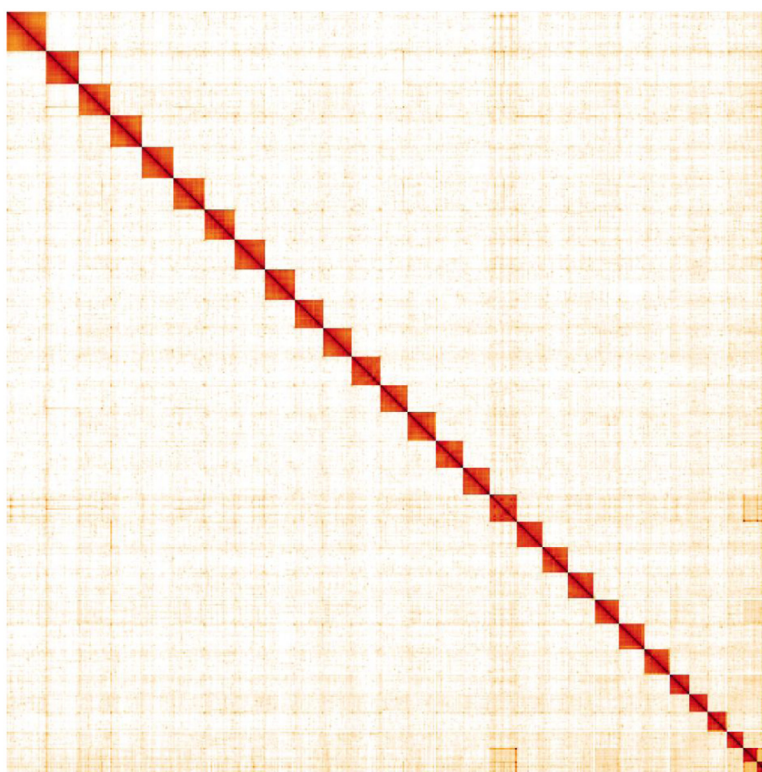


Figure 5. Genome assembly of *Erannis defoliaria*, iEraDefo1.1: Hi-C contact map. Hi-C contact map of the iEraDefo1.1 assembly, visualised in HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this map is available [here](#).

Table 2. Chromosomal pseudomolecules in the genome assembly of *Erannis defoliaria*, iEraDefo1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
FR990066.1	1	26.95	37.9
FR990068.1	2	21.76	38.2
FR990069.1	3	21.63	38.3
FR990070.1	4	21.62	38.0
FR990071.1	5	21.04	38.2
FR990072.1	6	20.86	37.9
FR990073.1	7	20.56	37.9
FR990074.1	8	20.34	37.7
FR990075.1	9	19.38	37.9
FR990076.1	10	19.34	38.0
FR990077.1	11	19.23	37.7
FR990078.1	12	19.18	37.8
FR990079.1	13	18.85	38.0
FR990080.1	14	18.58	38.0
FR990081.1	15	18.45	37.9

INSDC accession	Chromosome	Size (Mb)	GC%
FR990082.1	16	18.13	38.5
FR990083.1	17	18.00	38.1
FR990084.1	18	17.62	38.0
FR990085.1	19	17.46	38.4
FR990086.1	20	17.34	38.4
FR990087.1	21	17.21	38.2
FR990088.1	22	16.66	38.4
FR990089.1	23	13.38	38.5
FR990090.1	24	13.18	38.3
FR990091.1	25	12.47	38.4
FR990092.1	26	11.34	38.4
FR990093.1	27	9.50	39.2
FR990094.1	28	9.50	39.8
FR990067.1	Z	22.47	37.9
FR990095.1	MT	0.01	19.7
-	Unplaced	0.32	46.8

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.12	(Cheng <i>et al.</i> , 2021)
purge_dups	1.2.3	Guan <i>et al.</i> , 2020
SALSA	3.0	Ghurye <i>et al.</i> , 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
gEVAL	N/A	Chow <i>et al.</i> , 2016
HiGlass	1.11.6	(Kerpedjiev <i>et al.</i> , 2018)
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.0.5	Challis <i>et al.</i> , 2020

agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Each transfer of samples is further undertaken according to a Research Collaboration 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: Erannis defoliaria (mottled umber). Accession number [PRJEB43796](https://identifiers.org/ena.embl/PRJEB43796); <https://identifiers.org/ena.embl/PRJEB43796>.

The genome sequence is released openly for reuse. The *E. defoliaria* genome sequencing initiative is part of the [Darwin Tree of Life](#) (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

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

- Blažek J, Konečný A, Bartonička T: **Bat Aggregational Response to Pest Caterpillar Emergence.** *Sci Rep.* 2021; **11**(1): 13634. [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.** 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](#)

Glavendekić MM, Medarević MJ: **Insect Defoliators and Their Influence on Oak Forests in the Đerdap National Park, Serbia.** *Arch Biol Sci.* 2010; **62**(4): 1137–1141.

[Publisher Full Text](#)

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98.

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

Hansson BS, Szöcs G, Schmidt F, *et al.*: **Electrophysiological and Chemical Analysis of Sex Pheromone Communication System of the Mottled Umber, *Erannis Defoliaria* (Lepidoptera: Geometridae).** *J Chem Ecol.* 1990; **16**(6): 1887–97.

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

Hittenbeck A, Bialozyt R, Schmidt M: **Modelling the Population Fluctuation of Winter Moth and Mottled Umber Moth in Central and Northern Germany.** *For Ecosyst.* 2019; **6**(1): 4.

[Publisher Full Text](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience.* 2021; **10**(1): g1aa153.

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

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

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

Manni M, Berkeley MR, 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](#)

Netoiu C, Tomescu R, Vladescu D, *et al.*: **EVOLUTION OF INFESTATIONS WITH LOOPERMOTH (GEOMETRIDAE) IN OAK FORESTS FROM ROMANIA.** *Annals of the University of Craiova - Agriculture, Montanology, Cadastre Series.* 2015; **44**(1): 301–10.

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

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](#)

Sarvašová L, Kulfan J, Saniga M, *et al.*: **Winter Geometrid Moths in Oak Forests: Is Monitoring a Single Species Reliable to Predict Defoliation Risk?** *Forests.* 2020; **11**(3): 288.

[Publisher Full Text](#)

Sattler K: **A Review of Wing Reduction in Lepidoptera.** *Bulletin of the British Museum (Natural History) Entomology.* 1991; **60**: 243–88.

[Reference Source](#)

Skinner B, Wilson D: **Colour Identification Guide to the Moths of the British Isles.** 2009.

[Publisher Full Text](#)

Wagner DL, Liebherr JK: **Flightlessness in Insects.** *Trends Ecol Evol.* 1992; **7**(7): 216–20.

[Publisher Full Text](#)