



DATA NOTE

# The genome sequence of the Bird-cherry Ermine moth, *Yponomeuta evonymella* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Douglas Boyes<sup>1+</sup>, Callum Murray<sup>ib</sup><sup>2</sup>,  
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

<sup>1</sup>UK Centre for Ecology & Hydrology, Wallingford, England, UK

<sup>2</sup>DNA Pipelines, Wellcome Sanger Institute, Hinxton, England, UK

+ Deceased author

---

**V1** First published: 21 Oct 2024, 9:618  
<https://doi.org/10.12688/wellcomeopenres.23211.1>  
Latest published: 21 Oct 2024, 9:618  
<https://doi.org/10.12688/wellcomeopenres.23211.1>

---

## 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.

## Abstract

We present a genome assembly from an individual female *Yponomeuta evonymella* (the Bird-cherry Ermine; Arthropoda; Insecta; Lepidoptera; Yponomeutidae). The genome sequence has a total length of 572.70 megabases. Most of the assembly is scaffolded into 32 chromosomal pseudomolecules, including the trivalent sex chromosomes Z1, Z2 and W. The mitochondrial genome has also been assembled and is 16.16 kilobases in length.

## Keywords

*Yponomeuta evonymella*, Bird-cherry Ermine moth, genome sequence, chromosomal, Lepidoptera



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; **Murray C:** 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:** © 2024 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, Murray C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Bird-cherry Ermine moth, *Yponomeuta evonymella* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, 9:618 <https://doi.org/10.12688/wellcomeopenres.23211.1>

**First published:** 21 Oct 2024, 9:618 <https://doi.org/10.12688/wellcomeopenres.23211.1>

## Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Yponomeutoidea; Yponomeutidae; Yponomeutinae; *Yponomeuta*; *Yponomeuta evonymella* (Linnaeus, 1758) (NCBI:txid2567737).

## Background

*Yponomeuta evonymella* (Linnaeus, 1758), is a species of ermine moth identifiable by its white forewings adorned with five longitudinal rows of small black spots and a wingspan of 16–25 mm (UK Moths, no date). The ermine moth family is named due to the resemblance of the spot pattern to ermine fur. *Y. evonymella* caterpillars are yellowish-grey with black spots and live in communal webs on bird-cherry trees (*P. padus*); the host species in which the species is named.

*Y. evonymella* is native to the temperate zone in Europe, Siberia, India and China (Łukowski *et al.*, 2019) and is widely distributed across the Palaearctic region, from east Ireland all the way to Japan (GBIF Secretariat, 2023; Le & Pemberton, 2009). *Y. evonymella* is typically monophagous, feeding only on bird-cherry. However, larvae have been observed to feed on the invasive North American black-cherry (*P. serotina*) when in close proximity to bird-cherry (Łukowski *et al.*, 2019). *Y. evonymella* and other *Yponomeuta* species have interestingly been observed to utilise acoustic mullerian mimicry; using their hind wings to produce ultrasonic clicking noises when under attack from bats (O'Reilly *et al.*, 2019).

Adult *Y. evonymella* lay their eggs on the small twigs and buds of *P. padus*, covering the egg clusters with a fluid that hardens into a protective shield-like structure (Leather & Mackenzie, 1994). Once they emerge, the larvae spin large communal web-like nests, often encasing entire trees in a silken netting (Yonemura & Sehna, 2006). This can act as a barrier from possible predators such as parasitic wasps, as well as insulation from the weather, allowing large quantities of the caterpillars to participate in group feeding unhindered (Leather & Mackenzie, 1994; Łukowski *et al.*, 2019). The gregarious feeding often results in significant defoliation of the host plant, especially in years of high population density (Åström & Haeggström, 2018).

The production of high-quality reference genomes can refine our understanding of phylogenetic relationships and increase the accuracy of phylogenomic analyses (Triant *et al.*, 2018). Furthermore, it can deepen our understanding of pheromone production and its impact on insect-host relationships (Liénard *et al.*, 2010) and improve our knowledge of the genes involved in silk production; These are long and difficult to assemble, thus limiting studies on silk composition in lepidopteran species (Volenikova *et al.*, 2022).

The production of high-quality reference genomes can refine our understanding of *Yponomeuta evonymella*'s phylogenetic relationships and improve phylogenomic analyses (Triant *et al.*, 2018). This approach can also enhance our insights into

pheromone production and its role in insect-host dynamics (Liénard *et al.*, 2010), as well as facilitate studies on silk production genes, which are often complex and challenging to assemble (Volenikova *et al.*, 2022).

## Genome sequence report

The genome of an adult female *Yponomeuta evonymella* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 48.16 Gb (gigabases) from 4.30 million reads, providing approximately 79-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 85.78 Gb from 568.05 million reads, yielding an approximate coverage of 150-fold. Specimen and sequencing details are summarised in Table 1.

Manual assembly curation corrected 65 missing joins or mis-joins, reducing the scaffold number by 50.43%. The final assembly has a total length of 572.70 Mb in 56 sequence scaffolds, with 140 gaps, and a scaffold N50 of 20.0 Mb (Table 2). 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 (99.86%) of the assembly sequence was assigned to 32 chromosomal-level scaffolds, representing 29 autosomes and the Z<sub>1</sub>, Z<sub>2</sub> and W sex chromosomes. This heterogametic female sample appears to exhibit the sex chromosome trivalent system of  $n = 29A + AA^WZ$  as described in Nilsson *et al.* (1988). The half-coverage A component of the trivalent in iYpoEvo2 has been assigned as chromosome 27Z, this is based on its alignment to chromosome 27Z in the closely related species *Yponomeuta cagnagella* (GCA\_947310995.1). All other chromosomes also are named based on synteny with *Yponomeuta cagnagella* (Figure 5; Table 3). The order and orientation of contigs on the W chromosome between 9 Mb and 12.6 Mb is uncertain. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype



**Figure 1.** Photograph of the *Yponomeuta evonymella* (iYpoEvo2) specimen used for genome sequencing.

**Table 1. Specimen and sequencing data for *Yponomeuta evonymella*.**

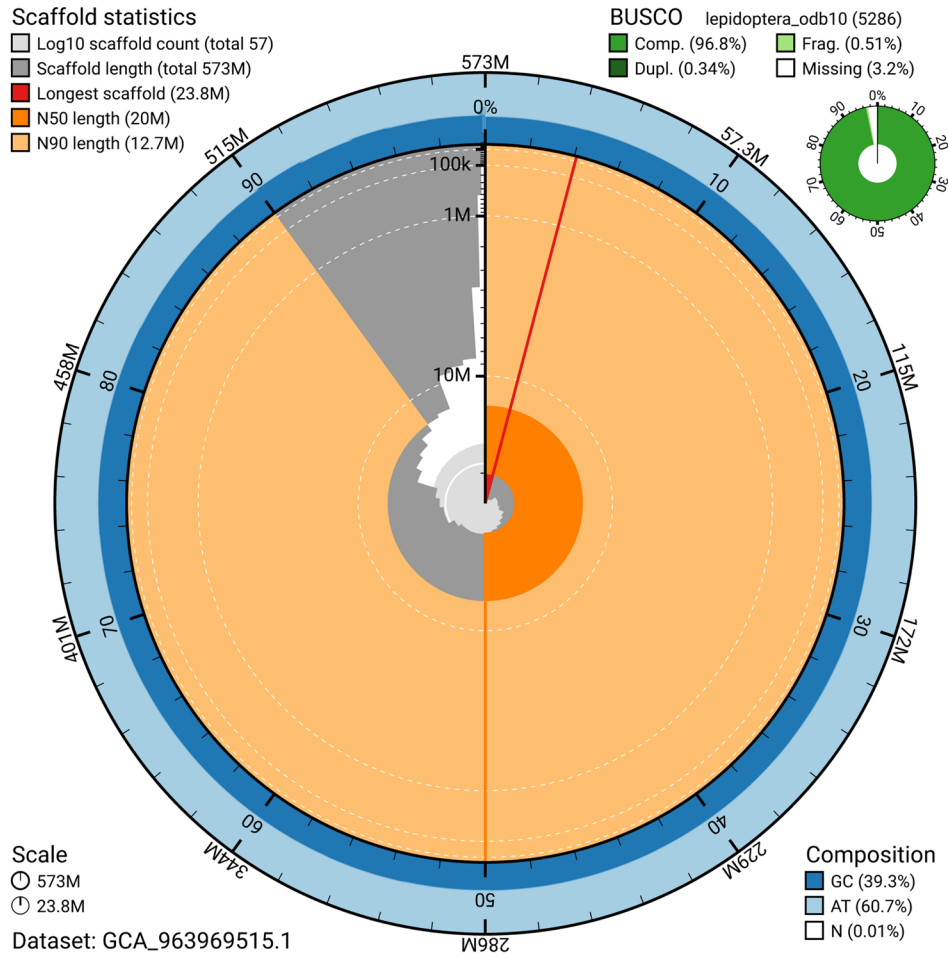
Project information			
Study title	<i>Yponomeuta evonymella</i> (bird-cherry ermine)		
Umbrella BioProject	PRJEB67425		
BioSample	SAMEA7701507		
NCBI taxonomy ID	2567737		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	ilYpoEvon2	SAMEA7701689	Whole organism
Hi-C sequencing	ilYpoEvon2	SAMEA7701689	Whole organism
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR12121877	5.68e+08	85.78
PacBio Revio	ERR12120049	4.30e+06	48.16

**Table 2. Genome assembly data for *Yponomeuta evonymella*, ilYpoEvon2.1.**

Genome assembly		
Assembly name	ilYpoEvon2.1	
Assembly accession	GCA_963969515.1	
Accession of alternate haplotype	GCA_963969545.1	
Span (Mb)	572.70	
Number of contigs	197	
Contig N50 length (Mb)	8.5	
Number of scaffolds	56	
Scaffold N50 length (Mb)	20.0	
Longest scaffold (Mb)	23.83	
Assembly metrics*	Benchmark	
Consensus quality (QV)	63.8	≥ 50
k-mer completeness	100.0%	≥ 95%
BUSCO**	C:96.7%[S:96.3%,D:0.4%], F:0.6%,M:2.7%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.86%	≥ 95%
Sex chromosomes	Z <sub>1</sub> , Z <sub>2</sub> , W	localised homologous pairs
Organelles	Mitochondrial genome: 16.16 kb	complete single alleles

\* 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 version 5.4.3. 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/Yponomeuta\\_evonymella/dataset/GCA\\_963969515.1/busco](https://blobtoolkit.genomehubs.org/view/Yponomeuta_evonymella/dataset/GCA_963969515.1/busco).



**Figure 2. Genome assembly of *Yponomeuta evonymella*, iYpoEvon2.1: metrics.** The BlobToolKit snail plot 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 572,742,383 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 (23,825,693 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (19,958,162 and 12,683,257 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/GCA\\_963969515.1/dataset/GCA\\_963969515.1/snail](https://blobtoolkit.genomehubs.org/view/GCA_963969515.1/dataset/GCA_963969515.1/snail).

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

The estimated Quality Value (QV) of the final assembly is 63.8 with  $k$ -mer completeness of 100.0%, and the assembly has a BUSCO v5.4.3 completeness of 96.7% (single = 96.3%, duplicated = 0.4%), using the lepidoptera\_odb10 reference set ( $n = 5,286$ ).

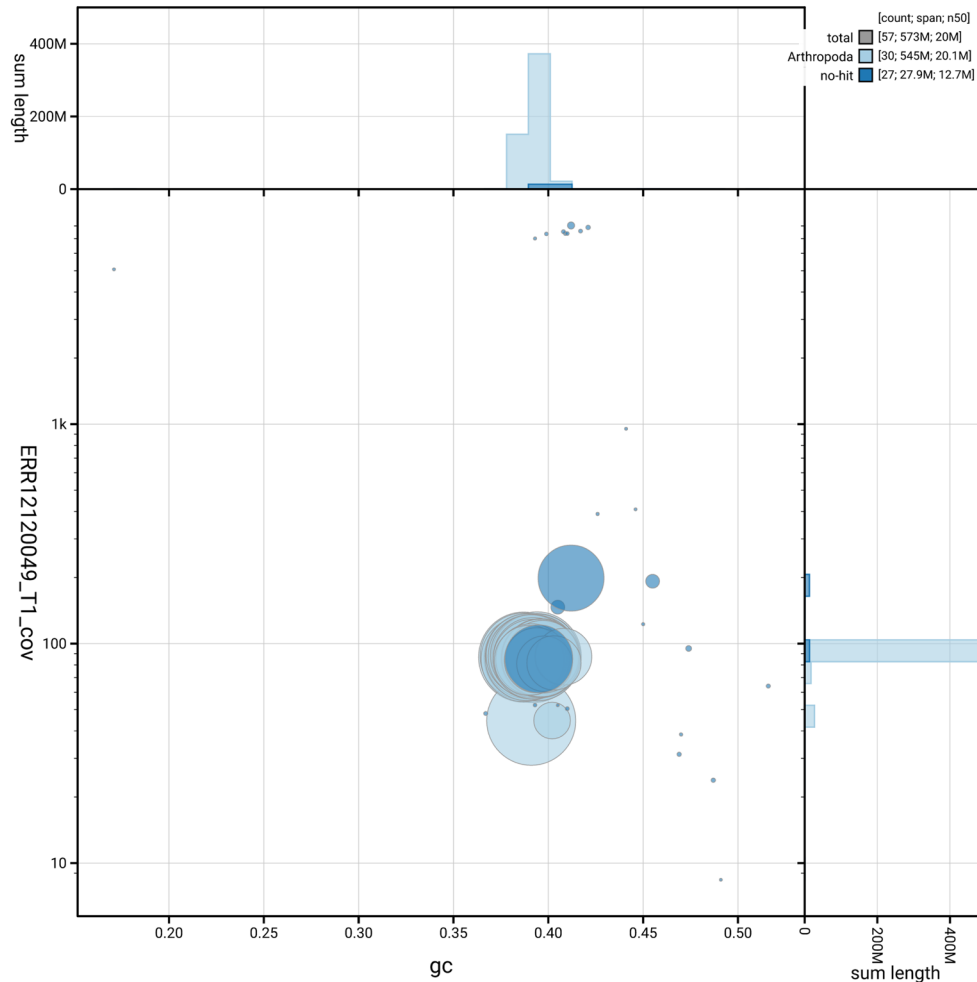
Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at <https://links.tol.sanger.ac.uk/species/2567737>.

## Methods

### Sample acquisition

An adult female *Yponomeuta evonymella* (specimen ID Ox000645, ToLID iYpoEvon2) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude  $-1.34$ ) on 2020-07-20 by light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by [Twyford et al. \(2024\)](#). A small sample was dissected from the specimens and stored in ethanol, while the remaining parts of the specimen were shipped on dry ice to the Wellcome



**Figure 3. Genome assembly of *Yponomeuta evonymella*, iYpoEvo2.1: BlobToolKit GC-coverage plot.** Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/GCA\\_963969515.1/dataset/GCA\\_963969515.1/blob](https://blobtoolkit.genomehubs.org/view/GCA_963969515.1/dataset/GCA_963969515.1/blob).

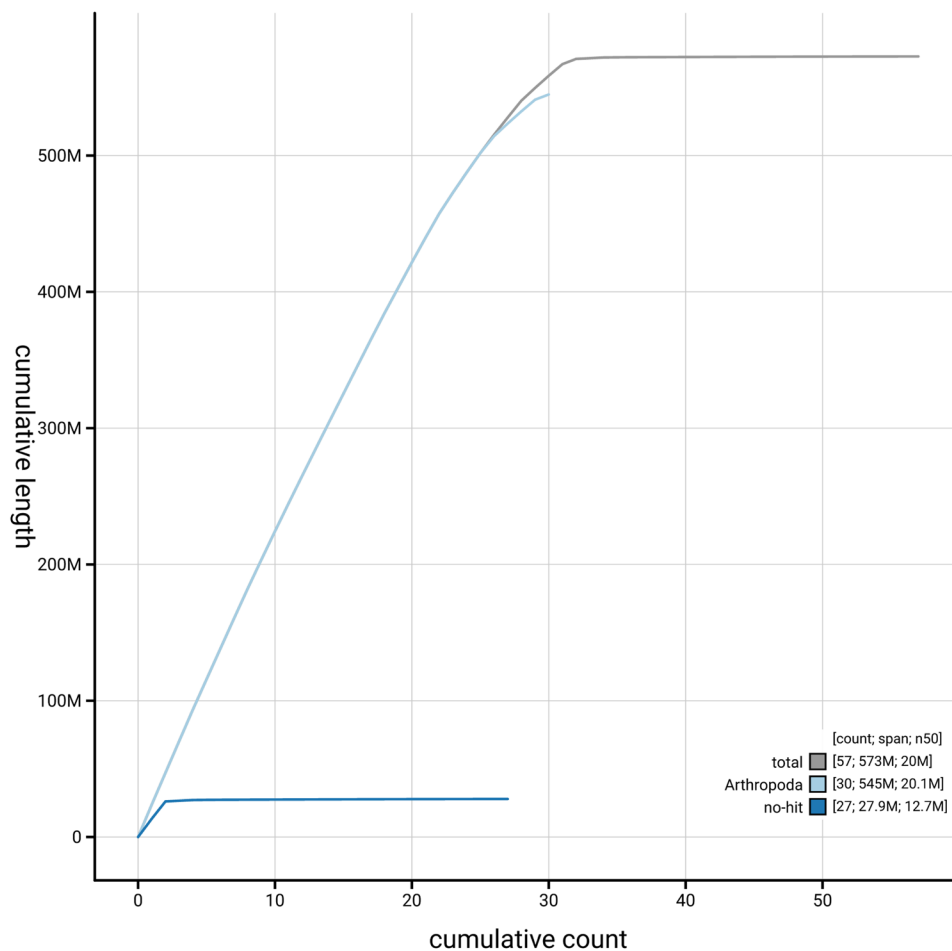
Sanger Institute (WSI). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io (Beasley *et al.*, 2023).

#### Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b).

The iYpoEvo2 sample was weighed and dissected on dry ice (Jay *et al.*, 2023), and tissue derived from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.



**Figure 4. Genome assembly of *Yponomeuta evonymella* iYpoEvoN2.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscodegenes taxrule. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/GCA\\_963969515.1/dataset/GCA\\_963969515.1/cumulative](https://blobtoolkit.genomehubs.org/view/GCA_963969515.1/dataset/GCA_963969515.1/cumulative).

### Library preparation and sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences Revio instrument.

Hi-C data were generated from remaining whole organism tissue of iYpoEvoN2 using the Arima-HiC v2 kit. In brief, frozen tissue ( $-80^{\circ}\text{C}$ ) was fixed, and the DNA crosslinked using a TC buffer containing formaldehyde. The crosslinked DNA was then digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with a biotinylated nucleotide and proximally ligated. The biotinylated DNA construct was fragmented to a fragment size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit, following manufacturers' instructions. The Hi-C sequencing was performed

using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

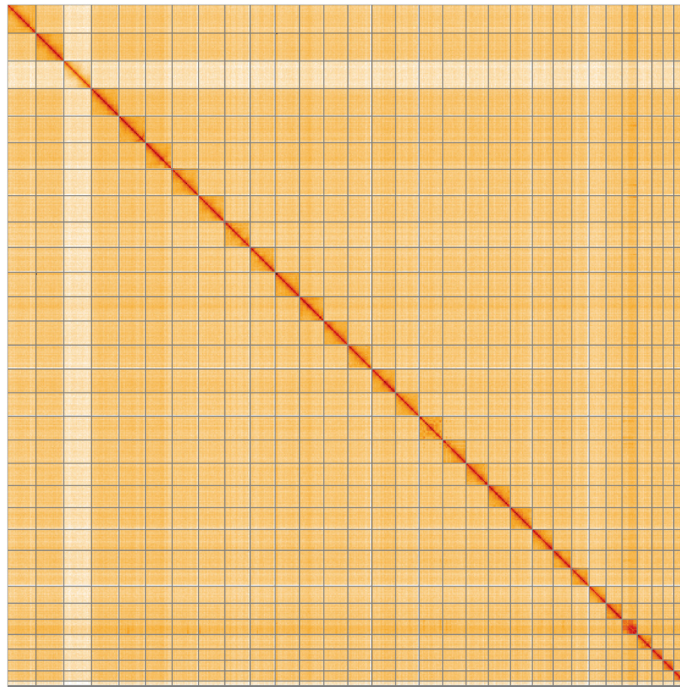
### Genome assembly, curation and evaluation

#### Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the `--primary` option. Haplotypic duplications were identified and removed using `purge_dups` (Guan *et al.*, 2020). The Hi-C reads were mapped to the primary contigs using `bwa-mem2` (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the `--break` option. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

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





**Figure 5. Genome assembly of *Yponomeuta evonymella* iYpoEvon2.1: Hi-C contact map of the iYpoEvon2.1 assembly, visualised using HiGlass.** Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at [https://genome-note-higlass.tol.sanger.ac.uk/I/?d=DNhfm4aBR22\\_LKgvr0dqMg](https://genome-note-higlass.tol.sanger.ac.uk/I/?d=DNhfm4aBR22_LKgvr0dqMg).

**Table 3. Chromosomal pseudomolecules in the genome assembly of *Yponomeuta evonymella*, iYpoEvon2.**

INSDC accession	Name	Length (Mb)	GC%
OZ018342.1	1	23.83	38.5
OZ018343.1	2	23.29	39.0
OZ018352.1	3	20.6	39.0
OZ018345.1	4	23.08	39.5
OZ018349.1	5	21.96	39.0
OZ018348.1	6	22.22	38.5
OZ018347.1	7	22.22	39.5
OZ018346.1	8	22.28	39.5
OZ018350.1	9	21.24	38.5
OZ018353.1	10	20.33	39.0
OZ018354.1	11	20.11	39.0
OZ018356.1	12	19.89	39.5
OZ018357.1	13	19.86	39.0
OZ018351.1	14	20.98	39.0
OZ018358.1	15	19.84	39.0

INSDC accession	Name	Length (Mb)	GC%
OZ018355.1	16	19.96	39.0
OZ018361.1	17	18.61	39.0
OZ018359.1	18	19.41	39.0
OZ018360.1	19	18.66	39.0
OZ018362.1	20	18.18	39.5
OZ018363.1	21	17.67	39.5
OZ018365.1	22	14.71	39.0
OZ018364.1	23	15.38	39.5
OZ018366.1	24	14.28	39.0
OZ018368.1	25	13.42	39.5
OZ018369.1	26	12.37	39.5
OZ018370.1	28	9.32	41.0
OZ018372.1	29	8.53	40.5
OZ018371.1	30	8.95	40.0
OZ018367.1	W	12.68	41.0
OZ018374.1	MT	0.02	17.5
OZ018344.1	Z1	23.24	39.0
OZ018373.1	Z2	3.84	40.0



mitochondrial contig and to ensure the general quality of the sequence.

### Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and misjoins were corrected, and duplicate sequences were tagged and removed. The sex chromosomes were assigned based on synteny analysis. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation> (article in preparation).

### Evaluation of the final assembly

The final assembly was post-processed and evaluated using the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines: sanger-tol/readmapping (Surana *et al.*, 2023a), sanger-tol/genomnote (Surana *et al.*, 2023b), and sanger-tol/blobtoolkit (Muffato *et al.*, 2024). The readmapping pipeline aligns the Hi-C reads using bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The genomnote pipeline transforms the Hi-C alignments into a contact map with BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018). This pipeline also generates assembly statistics using the NCBI datasets report (Sayers *et al.*, 2024), computes

*k*-mer completeness and QV consensus quality values with FastK and MERQURY.FK, and runs BUSCO (Manni *et al.*, 2021) to assess completeness.

The sanger-tol/blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoAT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND (Buchfink *et al.*, 2021) blastp. The genome is also split into chunks according to the density of the BUSCO genes from the closest taxonomically lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Genome sequences without a hit are chunked with seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these outputs into a blobdir for visualisation.

The genome assembly and evaluation pipelines were developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

**Table 4. Software tools: versions and sources.**

Software tool	Version	Source
BEDTools	2.30.0	<a href="https://github.com/arq5x/bedtools2">https://github.com/arq5x/bedtools2</a>
BLAST	2.14.0	<a href="ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/">ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/</a>
BlobToolKit	4.3.7	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.4.3 and 5.5.0	<a href="https://gitlab.com/e2lab/busco">https://gitlab.com/e2lab/busco</a>
bwa-mem2	2.2.1	<a href="https://github.com/bwa-mem2/bwa-mem2">https://github.com/bwa-mem2/bwa-mem2</a>
Cooler	0.8.11	<a href="https://github.com/open2c/cooler">https://github.com/open2c/cooler</a>
DIAMOND	2.1.8	<a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>
fasta_windows	0.2.4	<a href="https://github.com/tolkit/fasta_windows">https://github.com/tolkit/fasta_windows</a>
FastK	427104ea91c78c3b8b8b49f1a7d6b6beaa869ba1c	<a href="https://github.com/thegenemyers/FASTK">https://github.com/thegenemyers/FASTK</a>
Gfastats	1.3.6	<a href="https://github.com/vgl-hub/gfastats">https://github.com/vgl-hub/gfastats</a>
GoAT CLI	0.2.5	<a href="https://github.com/genomehubs/goat-cli">https://github.com/genomehubs/goat-cli</a>
Hifiasm	0.19.5-r587	<a href="https://github.com/chhylyp123/hifiasm">https://github.com/chhylyp123/hifiasm</a>
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Merqury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>

Software tool	Version	Source
MitoHiFi	3	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
MultiQC	1.14, 1.17, and 1.18	<a href="https://github.com/MultiQC/MultiQC">https://github.com/MultiQC/MultiQC</a>
NCBI Datasets	15.12.0	<a href="https://github.com/ncbi/datasets">https://github.com/ncbi/datasets</a>
Nextflow	23.04.0-5857	<a href="https://github.com/nextflow-io/nextflow">https://github.com/nextflow-io/nextflow</a>
PretextView	0.2	<a href="https://github.com/sanger-tol/PretextView">https://github.com/sanger-tol/PretextView</a>
purge_dups	1.2.5	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
samtools	1.16.1, 1.17, and 1.18	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
sanger-tol/ascc	-	<a href="https://github.com/sanger-tol/ascc">https://github.com/sanger-tol/ascc</a>
sanger-tol/genomenote	1.1.1	<a href="https://github.com/sanger-tol/genomenote">https://github.com/sanger-tol/genomenote</a>
sanger-tol/readmapping	1.2.1	<a href="https://github.com/sanger-tol/readmapping">https://github.com/sanger-tol/readmapping</a>
Seqtk	1.3	<a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>
Singularity	3.9.0	<a href="https://github.com/sylabs/singularity">https://github.com/sylabs/singularity</a>
TreeVal	1.0.0	<a href="https://github.com/sanger-tol/treeval">https://github.com/sanger-tol/treeval</a>
YaHS	1.2a.2	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

## 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: *Yponomeuta evonymella* (bird-cherry ermine). Accession number PRJEB67425; <https://identifiers.org/ena.embl/PRJEB67425> (Wellcome Sanger Institute, 2024). The genome sequence is released openly for reuse. The *Yponomeuta evonymella* 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 available 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](#) and [Table 2](#).

## Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12157525>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

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](#)
- 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](#)
- Altschul SF, Gish W, Miller W, et al.: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Åström HM, Haeggström CA: **What happens with the tree rings when the bird-cherry (*Prunus padus* L.) is defoliated by the moth bird-cherry ermine (*Yponomeuta evonymellus* Linnæus, 1758)?** *Memoranda Societatis pro Fauna et Flora Fennica.* 2018; **94**: 78–83.  
[Reference Source](#)
- Bateman A, Martin MJ, Orchard S, et al.: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor<sup>®</sup>3 for LI PacBio.** *protocols.io.* 2023.  
[Publisher Full Text](#)
- Beasley J, Uhl R, Forrest LL, et al.: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023; (accessed 25 June 2024).  
[Publisher Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Kumar S, Sotero-Caio C, et al.: **Genomes on a Tree (GoAT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.  
[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](#)
- Crowley L, Allen H, Barnes I, et al.: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- da Veiga Leprevost F, Grüning BA, Alves Aflitos S, et al.: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danecek P, Bonfield JK, Liddle J, et al.: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.  
[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](#)
- Di Tommaso P, Chatzou M, Floden EW, et al.: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Diesch C, Stevens GJ, Xie P, et al.: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels P, Magnusson M, Lundin S, et al.: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ewels PA, Peltzer A, Fillinger S, et al.: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Formenti G, Abueg L, Brajuka A, et al.: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- GBIF Secretariat: ***Yponomeuta evonymella* (Linnaeus, 1758).** *GBIF Backbone Taxonomy.* 2023; [Accessed 1 October 2024].  
[Reference Source](#)
- Grüning B, Dale R, Sjödin A, et al.: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.  
[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](#)
- Harry E: **PretextView (Paired REAd TEXTure Viewer): a desktop application for viewing pretext contact maps.** 2022.  
[Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): gaa153.  
[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](#)
- Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Le JH, Pemberton RW: **Parasitoid complex of the bird cherry ermine moth *Yponomeuta evonymellus* in Korea.** *Entomol Res.* 2009; **39**(3): 201–206.  
[Publisher Full Text](#)
- Leather SR, Mackenzie GA: **Factors affecting the population development of the bird cherry ermine moth, *Yponomeuta evonymella*.** *The Entomologist.* 1994; **113**(2): 86–105.  
[Reference Source](#)
- Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liénard MA, Hagström AK, Lassance JM, et al.: **Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene.** *Proc Natl Acad Sci U S A.* 2010; **107**(24): 10955–10960.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Łukowski A, Janek W, Baraniak E, et al.: **Changing host plants causes structural differences in the parasitoid complex of the monophagous moth *Yponomeuta evonymella*, but does not improve survival rate.** *Insects.* 2019; **10**(7): 197.  
[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): 4647–4654.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].  
[Reference Source](#)
- Muffato M, Butt Z, Challis R, et al.: **sanger-tol/blobtoolkit: v0.3.0 – Poliwig.** 2024.  
[Publisher Full Text](#)
- Nilsson NO, Löfstedt C, Dävring L: **Unusual sex chromosome inheritance in six species of small ermine moths (*Yponomeuta*, Yponomeutidae,**

Lepidoptera). *Hereditas*. 1988; **108**(2): 259–265.

[Publisher Full Text](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2**. *protocols.io*. 2023.

[Publisher Full Text](#)

O'Reilly LJ, Agassiz DJL, Neil TR, *et al.*: **Deaf moths employ acoustic Müllerian mimicry against bats using wingbeat-powered tymbals**. *Sci Rep*. 2019; **9**(1): 1444.

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

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis**. 2023.

[Publisher Full Text](#)

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics*. 2010; **26**(6): 841–842.

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

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](#)

Sayers EW, Cavanaugh M, Clark K, *et al.*: **GenBank 2024 update**. *Nucleic Acids Res*. 2024; **52**(D1): D134–D137.

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

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

[Publisher Full Text](#)

Surana P, Muffato M, Qi G: **Sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0)**. *Zenodo*. 2023a.

[Publisher Full Text](#)

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev)**. *Zenodo*. 2023b.

[Publisher Full Text](#)

Triant DA, Cinel SD, Kawahara AY: **Lepidoptera genomes: current knowledge, gaps and future directions**. *Curr Opin Insect Sci*. 2018; **25**: 99–105.

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

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2024; **9**: 339.

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

UK Moths: **Bird-cherry ermine *Yponomeuta evonymella***. *UKMoths*. [no date]; [Accessed 22 August 2024].

[Reference Source](#)

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](#)

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](#)

Volenikova A, Nguyen P, Davey P, *et al.*: **Genome sequence and silkomics of the spindle ermine moth, *Yponomeuta cagnagella*, representing the early diverging lineage of the ditrysian Lepidoptera**. *Commun Biol*. 2022; **5**(1): 1281.

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

Wellcome Sanger Institute: **The genome sequence of the Bird-cherry Ermine moth, *Yponomeuta evonymella* (Linnaeus, 1758)**. European Nucleotide Archive. [dataset], accession number PRJEB67425, 2024.

Yonemura N, Sehna F: **The Design of Silk Fiber Composition in Moths Has Been Conserved for More Than 150 Million Years**. *J Mol Evol*. 2006; **63**(1): 42–53.

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