DATA NOTE

The genome sequence of the Black-tipped Ermine, *Yponomeuta plumbella* (Denis & Schiffermüller, 1775) [version 1; peer review: awaiting peer review]

Douglas Boyes\(^1\), Ian Sims\(^2\), David Lees\(^3\), Peter W.H. Holland\(^4\),
University of Oxford and Wytham Woods Genome Acquisition Lab,
Natural History Museum Genome Acquisition Lab,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

\(^1\)UK Centre for Ecology & Hydrology, Wallingford, England, UK
\(^2\)The British Entomological and Natural History Society (BENHS), Reading, England, UK
\(^3\)Natural History Museum, London, England, UK
\(^4\)University of Oxford, Oxford, England, UK

* Deceased author

**Abstract**

We present a genome assembly from an individual male *Yponomeuta plumbella* (the Black-tipped Ermine; Arthropoda; Insecta; Lepidoptera; Yponomeutidae). The genome sequence is 636.6 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.5 kilobases in length.

**Keywords**

Yponomeuta plumbella, Black-tipped Ermine, 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)

Author roles: Boyes D: Investigation, Resources; Sims I: Investigation, Resources; Lees 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.

Copyright: © 2023 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, Sims I, Lees D et al. The genome sequence of the Black-tipped Ermine, Yponomeuta plumbella (Denis & Schiffermüller, 1775) [version 1; peer review: awaiting peer review] Wellcome Open Research 2023, 8:246 https://doi.org/10.12688/wellcomeopenres.19368.1

First published: 14 Jun 2023, 8:246 https://doi.org/10.12688/wellcomeopenres.19368.1
**Species taxonomy**
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Yponomeutoidea; Yponomeutidae; Yponomeutinae; *Yponomeuta*; *Yponomeuta plumbellus* (Denis & Schiffermüller, 1775) (NCBI:txid1594356).

**Background**
The genus *Yponomeuta* contains several small moths with white forewings patterned with black speckles, collectively known as ‘small ermine’ moths. *Y. plumbella* (synonym *Y. plumbellus*), the Black-tipped Ermine or Large-spot Ermine, can be distinguished from similar species by a large black smudge halfway along the speckled forewing plus a dark mark at the wing apex (Asher et al., 2013). The ground colour of the wings is slightly off-white which may be the origin of the specific name *plumbella* meaning lead-like (Emmet, 1991). *Y. plumbella* is widespread across northern, central and eastern Europe, southern parts of Scandinavia, southwest regions of Ireland, and central and southern counties of England and Wales (GBIF Secretariat, 2023; Sterling & Parsons, 2018). In Britain, the moth is commonest on chalk-rich areas of southern England (Sterling & Parsons, 2018). The moth has also been introduced to the United States with the first record being a 1949 specimen from Martha’s Vineyard, Massachusetts (Hoebeke, 1987). It does not seem to be spreading from probable sites of introduction and has only been recorded from Massachusetts and Rhode Island (MassMoths, 2023).

There has been confusion over whether *Y. plumbella* has one or two generations per year in Britain (Agassiz, 1996; Kimber, 2023; Sterling & Parsons, 2018); records that report the life cycle stage are clearly consistent with a univoltine life cycle, with larvae found predominantly from April to June and adults in July and August (Perry, 2023). The adults lay eggs on twigs of spindle *Euonymus europaeus* and the hatched larvae overwinter just below the empty eggshell. Development continues in spring with the larvae burrowing into fresh shoots, before spinning a silken web on the food plant where they live eating spindle leaves (Agassiz, 1996). The webs spun by *Y. plumbella* are smaller and less conspicuous than those made by the speckled ermine *Y. cagnagella* on the same plant and contain just a few larvae (Agassiz, 1996; Sterling & Parsons, 2018).

A genome sequence for *Y. plumbella* will facilitate research into the evolution of host plant specificity in herbivorous insects and contribute to the growing set of resources for studying the evolution of Lepidoptera.

**Genome sequence report**
The genome was sequenced from one male *Yponomeuta plumbella* (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude −1.34). A total of 30-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 10 missing joins or mis-joins, reducing the scaffold count by one.

The final assembly has a total length of 636.6 Mb in 51 sequence scaffolds with a scaffold N50 of 22.9 Mb (Table 1). Most (99.81%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). *ilYpoPlub1* is a male specimen, exhibiting the typical ZZ chromosome system. There is a possibility that chromosome 27 in this assembly may be an additional sex chromosome based on its alignment to 27Z in the assemblies of the female specimens of *Yponomeuta cagnagella* (GCA_947310995.1) and *Y. rorrellus* (GCA_947308005.1), both of which have been found to exhibit the sex chromosome trivalent system of *n* = 29A + A A° Z as described in (Nilsson et al., 1988). In the two female genome assemblies, the A component of the trivalent was assigned as chromosome 27Z based on its alignment to chromosome 27 in the closely related species *Y. sedellus* (iLYpoSed1 GCA_934045075.1), a ZZ male specimen (Boyes & Langdon, 2023).

While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype 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 with k-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 97.6% (single = 97.2%, duplicated = 0.4%), using the lepidoptera_odb10 reference set (*n* = 5,286).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/1594356.

**Methods**
**Sample acquisition and nucleic acid extraction**
A male *Yponomeuta plumbella* specimen (*ilYpoPlub1*) was collected from Wytham Woods, Oxfordshire (biological
A second male *Yponomeuta plumbella* specimen (ilYpoPlub2) was collected in Hartslock Nature Reserve, UK (latitude 51.51, longitude –1.11) on 29 July 2021 by Ian Sims (British Entomological and Natural History Society) using a light trap. This specimen was identified by David Lees (Natural History Museum).
History Museum) and dry frozen at –80°C. The specimen ilYpoPlub2 was used for Hi-C scaffolding.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilYpoPlub1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Whole organism tissue was using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from abdomen tissue of ilYpoPlub2 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer’s instructions. RNA was then eluted in 50 μl RNase-free water and its concentration assessed using

---

**Figure 2. Genome assembly of *Yponomeuta plumbeola*, ilYpoPlub1.1: 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 636,643,948 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 (27,348,352 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (22,918,849 and 14,468,000 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/ilYpoPlub1.1/dataset/CAMZJP01/snail.
a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing
Pacific Biosciences HiFi circular consensus 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 thorax tissue of ilYpoPlub2 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation
Assembly was carried out with Hifiasm (Cheng et al., 2021) and haplotypic duplication was identified and removed with

Figure 3. Genome assembly of Yponomeuta plumbella, ilYpoPlub1.1: 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/ilYpoPlub1.1/dataset/CAMZJP01/blob.
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 as described 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., 2022), which runs MitoFinder (Allio et al., 2020) or MITOS (Bernt et al., 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence. To evaluate the assembly, MerquryFK was used to estimate consensus quality (QV) scores and k-mer completeness (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. Table 3 contains a list of software tool versions and sources.

Figure 4. Genome assembly of *Yponomeuta plumbella*, ilYpoPlub1.1: 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/ilYpoPlub1.1/dataset/CAMZJP01/cumulative.
Table 2. Chromosomal pseudomolecules in the genome assembly of *Yponomeuta plumbella*, iYpoPlub1.

<table>
<thead>
<tr>
<th>INSDC accession</th>
<th>Chromosome</th>
<th>Size (Mb)</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OX371192.1</td>
<td>1</td>
<td>26.37</td>
<td>39</td>
</tr>
<tr>
<td>OX371193.1</td>
<td>2</td>
<td>26.08</td>
<td>38.9</td>
</tr>
<tr>
<td>OX371194.1</td>
<td>3</td>
<td>23.82</td>
<td>39</td>
</tr>
<tr>
<td>OX371195.1</td>
<td>4</td>
<td>25.43</td>
<td>39.2</td>
</tr>
<tr>
<td>OX371196.1</td>
<td>5</td>
<td>24.4</td>
<td>38.8</td>
</tr>
<tr>
<td>OX371197.1</td>
<td>6</td>
<td>24.98</td>
<td>38.8</td>
</tr>
<tr>
<td>OX371198.1</td>
<td>7</td>
<td>24.65</td>
<td>39.3</td>
</tr>
<tr>
<td>OX371199.1</td>
<td>8</td>
<td>24.37</td>
<td>39.3</td>
</tr>
<tr>
<td>OX371200.1</td>
<td>9</td>
<td>23.58</td>
<td>38.8</td>
</tr>
<tr>
<td>OX371201.1</td>
<td>10</td>
<td>22.92</td>
<td>39.1</td>
</tr>
<tr>
<td>OX371202.1</td>
<td>11</td>
<td>23.73</td>
<td>39.3</td>
</tr>
<tr>
<td>OX371203.1</td>
<td>12</td>
<td>22.62</td>
<td>39.2</td>
</tr>
<tr>
<td>OX371204.1</td>
<td>13</td>
<td>22.17</td>
<td>39.1</td>
</tr>
<tr>
<td>OX371205.1</td>
<td>14</td>
<td>22.97</td>
<td>38.9</td>
</tr>
<tr>
<td>OX371206.1</td>
<td>15</td>
<td>21.22</td>
<td>39.3</td>
</tr>
</tbody>
</table>

Figure 5. Genome assembly of *Yponomeuta plumbella*, iYpoPlub1.1: Hi-C contact map of the iYpoPlub1.1 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/l/?d=Y5Kqph9NR0CKSw-LVD24Vg.
### Table 3. Software tools: versions and sources.

<table>
<thead>
<tr>
<th>Software tool</th>
<th>Version</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlobToolKit</td>
<td>4.0.7</td>
<td><a href="https://github.com/blobtookit/blobtoolkit">https://github.com/blobtookit/blobtoolkit</a></td>
</tr>
<tr>
<td>BUSCO</td>
<td>5.3.2</td>
<td><a href="https://gitlab.com/ezlab/busco">https://gitlab.com/ezlab/busco</a></td>
</tr>
<tr>
<td>Hifiasm</td>
<td>0.16.1-r375</td>
<td><a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a></td>
</tr>
<tr>
<td>HiGlass</td>
<td>1.11.6</td>
<td><a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a></td>
</tr>
<tr>
<td>Merquury</td>
<td>MerquuryFK</td>
<td><a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a></td>
</tr>
<tr>
<td>MitoHiFi</td>
<td>2</td>
<td><a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a></td>
</tr>
<tr>
<td>PretextView</td>
<td>0.2</td>
<td><a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a></td>
</tr>
<tr>
<td>purge_dups</td>
<td>1.2.3</td>
<td><a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a></td>
</tr>
<tr>
<td>YaHS</td>
<td>yahs-1.1.91eebc2</td>
<td><a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a></td>
</tr>
</tbody>
</table>

### Ethics and compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code 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. All efforts are undertaken to minimise the suffering of animals used for sequencing. 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 plumbellus* (black-tipped ermine). Accession number PRJEB55978; https://identifiers.org/ena.embl/PRJEB55978. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Yponomeuta plumbella* 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.

### Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789928.

Members of the Natural History Museum Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4790043.

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 programme are listed here: https://doi.org/10.5281/zenodo.4783585.


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


Boyes D, Langdon WB: *The genome sequence of the Grey Ermine, Yponomeuta sedella (Treitschke, 1832) [version 1; peer review: awaiting peer review]*. Wellcome Open Res. 2023; 8: 50. Publisher Full Text


