DATA NOTE

The genome sequence of the Grey Ermine, *Yponomeuta sedella* (Treitschke, 1832) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+],
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, William B.V. Langdon[2],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]University of Oxford, Oxford, Oxfordshire, UK

[+] Deceased author

v1 **First published:** 02 Feb 2023, **8**:50
https://doi.org/10.12688/wellcomeopenres.18898.1
Latest published: 02 Feb 2023, **8**:50
https://doi.org/10.12688/wellcomeopenres.18898.1

Abstract

We present a genome assembly from an individual male *Yponomeuta sedella* (the Grey Ermine; Arthropoda; Insecta; Lepidoptera; Yponomeutidae). The genome sequence is 658 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the assembled Z sex chromosome. The mitochondrial genome has also been assembled and is 16.4 kilobases in length. Gene annotation of this assembly on Ensembl has identified 13,010 protein coding genes.

Keywords

Yponomeuta sedella, Grey Ermine, 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; **Langdon WBV**: 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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the Grey Ermine, *Yponomeuta sedella* (Treitschke, 1832) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:50 https://doi.org/10.12688/wellcomeopenres.18898.1

First published: 02 Feb 2023, **8**:50 https://doi.org/10.12688/wellcomeopenres.18898.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Yponomeutoidea; Yponomeutidae; Yponomeutinae; *Yponomeuta*; *Yponomeuta sedella* (Treitschke, 1832) (NCBI: txid263933).

Background

Yponomeuta sedella (Treitschke, 1832) is a moth in the family Yponomeutidae and the genus *Yponomeuta*, commonly known as the 'small ermine moths' for their black and white colouring. In the UK, *Y. sedella* is rather local, and while it has been recorded from most counties south of a line between the Humber and the Severn (Emmet *et al.*, no date), it was classified as nationally scarce in the last review of UK micro-moth statuses (Davis, 2012). Other *Yponomeuta* species are regarded as occasional migrants to the UK from Europe (e.g. Clancy, 2013), and it is likely that this species is too (Sterling & Parsons, 2018), though in lesser numbers. It is found across mainland Europe and east into Russia, China and Japan (Agassiz, 1996).

The adult moth is nocturnal and flies in two broods in spring and late summer when it is attracted to light (Agassiz, 1996). Like other species in the genus, *Y. sedella* larvae live gregariously, initially mining the leaves of Orpine (*Sedum telephium*) (Pitkin *et al.*, no date), before feeding externally on the leaves and flowers of the plant in a small web (Agassiz, 1996). These webs are not as extensive as those of other species in the group feeding on woody plants, which can blanket whole bushes (Sterling & Parsons, 2018). *Y. sedella* has also been bred from cultivated species of *Sedum*, for example *Sedum cauticolum* 'Ruby Glow' (Halstead, 1985) and may well feed on these regularly (Sterling & Parsons, 2018). Transportation of larvae on horticultural varieties, as has been recorded in the UK (Hall *et al.*, no date), may therefore explain occasional records outside its known range.

Other species in the genus feed on woody plants and in general show much overlap in their biology, so much so that there has been great interest in their use as a system for studying mating systems and sympatric speciation, reviewed in detail by Lienard and Löfstedt, 2016. Each species appears to have adopted a unique combination of host plant, pheromone composition and activity pattern to ensure reproductive isolation. For example, although the sex pheromones produced by female *Y. sedella* are almost identical to those of *Y. evonymella* (Löfstedt *et al.*, 1991), the two species occur in different habitats and fly at different times of the year, while *Y. sedella* females also call much earlier in the night (Hendrikse, 1979).

Recent work has also shown that organs in the hindwings which can produce ultrasonic clicks are widespread in the Yponomeutidae (Agassiz, 2017), and are present on *Y. sedella* (Agassiz, 2022, pers. comm.). These clicks are similar to those produced by Arctiid moths to warn their main predators (bats) of the noxious compounds they contain. *Yponomeuta* species also contain such compounds, and this has been suggested as an example of acoustic Müllerian mimicry (O'Reilly *et al.*, 2019).

The genome of *Y. sedella* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *Y. sedella*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Y. sedella* (Figure 1) collected from Wytham Woods (51.772, –1.338). A total of 31-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 39 missing joins or mis-joins, reducing the scaffold number by 23.29%, and increasing the scaffold N50 by 0.71%.

The final assembly has a total length of 657.6 Mb in 56 sequence scaffolds with a scaffold N50 of 23.9 Mb (Table 1). Most (99.8%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 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). The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 97.5% (single 97%, duplicated 0.5%) using the lepidoptera_odb10 reference set (n = 5,286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Figure 1. Photograph of the *Yponomeuta sedella* (ilYpoSedl1) specimen used for genome sequencing.

Table 1. Genome data for *Yponomeuta sedella*, ilYpoSedl1.1.

Project accession data		
Assembly identifier	ilYpoSedl1.1	
Species	*Yponomeuta sedella*	
Specimen	ilYpoSedl1	
NCBI taxonomy ID	263933	
BioProject	PRJEB50749	
BioSample ID	SAMEA7746622	
Isolate information		
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.7	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:97.5%[S:97.0%,D:0.5%], F:0.6%, M:1.9%, n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.8%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR8575389	
Hi-C Illumina	ERR8571675	
Genome assembly		
Assembly accession	GCA_934045075.1	
Accession of alternate haplotype	GCA_934044855.1	
Span (Mb)	657.6	
Number of contigs	453	
Contig N50 length (Mb)	2.8	
Number of scaffolds	56	
Scaffold N50 length (Mb)	23.9	
Longest scaffold (Mb)	27.3	
Genome annotation		
Number of protein-coding genes	13,010	
Number of non-coding genes	2,198	
Number of gene transcripts	22,856	

* 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/ilYpoSedl1.1/dataset/CAKOGX01/busco.

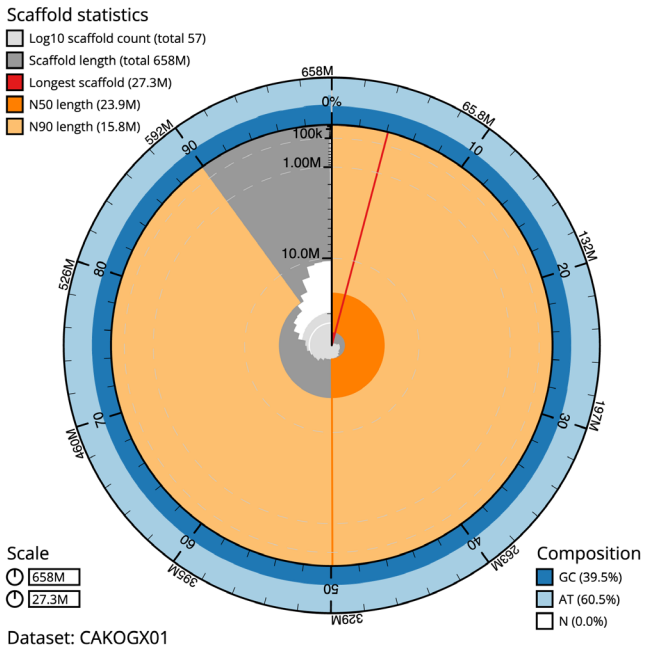

Figure 2. Genome assembly of *Yponomeuta sedella*, ilYpoSedl1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 657,583,382 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (27,294,112 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (23,909,508 and 15,780,076 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/ilYpoSedl1.1/dataset/CAKOGX01/snail.

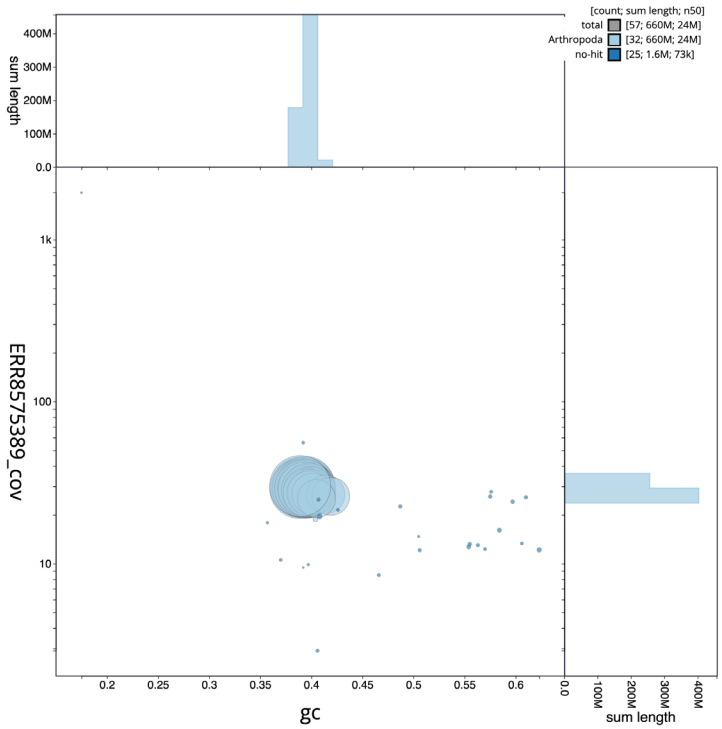

Figure 3. Genome assembly of *Yponomeuta sedella*, ilYpoSedl1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilYpoSedl1.1/dataset/CAKOGX01/blob.

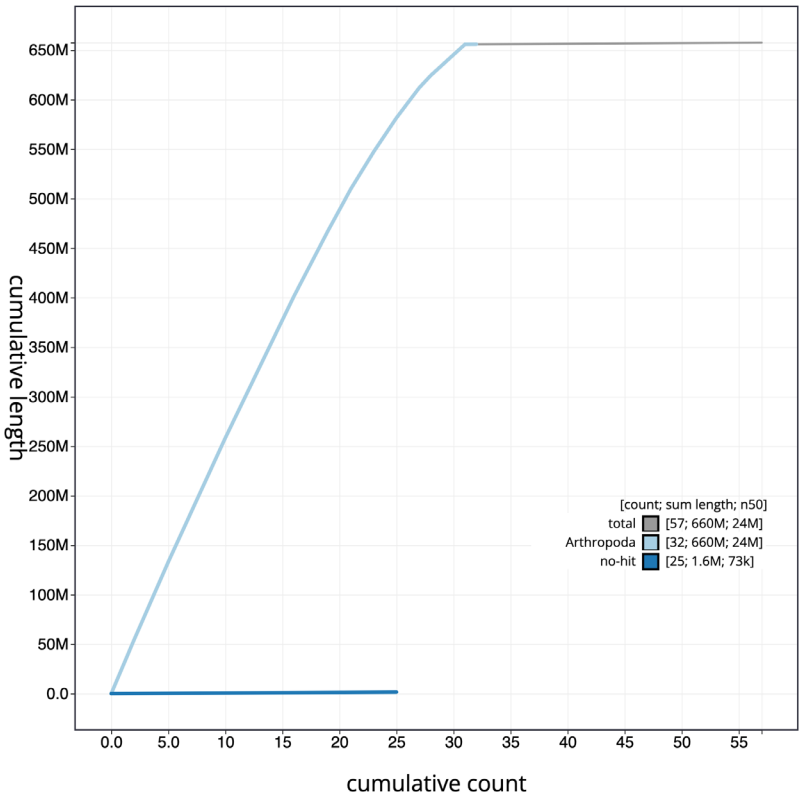

Figure 4. Genome assembly of *Yponomeuta sedella*, ilYpoSedl1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilYpoSedl1.1/dataset/CAKOGX01/cumulative.

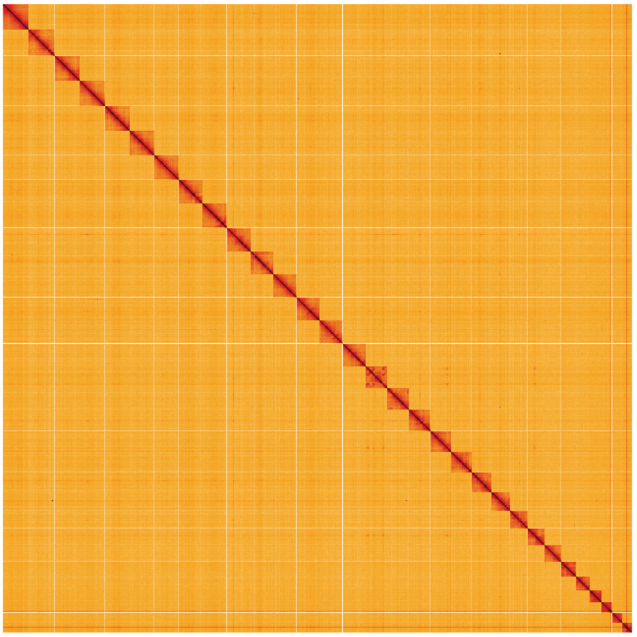

Figure 5. Genome assembly of *Yponomeuta sedella*, ilYpoSedl1.1: Hi-C contact map. Hi-C contact map of the ilYpoSedl1.1 assembly, visualised using HiGlass. The Hi-C data were obtained from specimen ilYpoSedl2. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=W2oFvC5fTO6upL-iiuiJ7A.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Yponomeuta sedella*, ilYpoSedl1.

INSDC accession	Chromosome	Size (Mb)	GC%
OW203722.1	1	27.08	39
OW203723.1	2	26.4	39.3
OW203724.1	3	26.28	39.1
OW203725.1	4	25.99	39.4
OW203726.1	5	25.33	39
OW203727.1	6	25.2	38.8
OW203728.1	7	25.06	39.5
OW203729.1	8	25.06	39.4
OW203730.1	9	24.7	39
OW203731.1	10	24.2	39.2
OW203732.1	11	24.05	39.5
OW203733.1	12	23.91	39.4
OW203734.1	13	23.78	39.4
OW203735.1	14	23.69	39.2
OW203736.1	15	23.13	39.6
OW203737.1	16	22.35	39.1
OW203738.1	17	22.17	39.4
OW203739.1	18	21.87	39.4
OW203740.1	19	21.3	39.4
OW203741.1	20	20.99	39.7
OW203742.1	21	19.36	39.6
OW203743.1	22	18.06	39.7
OW203744.1	23	17.31	40
OW203745.1	24	17.11	39.6
OW203746.1	25	15.78	40
OW203747.1	26	14.6	39.8
OW203748.1	27	12.37	40.1
OW203749.1	28	10.64	41.4
OW203750.1	29	10.33	41.9
OW203751.1	30	10.41	40.5
OW203721.1	Z	27.29	38.9
OW203752.1	MT	0.02	17.7
-	-	1.77	50.3

Genome annotation report

The *Y. sedella* (ilYpoSedl1.1) genome assembly was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Yponomeuta_sedellus_GCA_934045075.1/). The resulting annotation includes 22,856 transcribed mRNAs from 13,010 protein-coding and 2,198 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Y. sedella* specimen (ilYpoSedl1) was caught using a light trap in Wytham Great Wood, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.772, longitude –1.338). on 1 August 2020. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

A second *Y. sedella* specimen (ilYpoSedl2) was caught in Tonbridge, Kent, UK (latitude 51.19, longitude 0.29) on 11 August 2021 by Gavin Broad (Natural History Museum) and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilYpoSedl1 sample was weighed and dissected on dry ice. Whole organism tissue was disrupted 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.

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 Pacific Biosciences SEQUEL II (HiFi) instrument. Hi-C data were also generated from head and thorax tissue of ilYpoSedl2 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly

Assembly 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 scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2022). 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 performed annotation

using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where relevant.

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Y. sedella* assembly (GCA_934045075.1). 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).

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. By agreeing with and signing up to the Sampling Code 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.

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	3.4.0	Challis *et al.*, 2020
Hifiasm	0.16.1-r375	Cheng *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
MitoHiFi	2	Uliano-Silva *et al.*, 2022
PretextView	0.2	Harry, 2022
purge_dups	1.2.3	Guan *et al.*, 2020
YaHS	yahs-1.1.91eebc2	Zhou *et al.*, 2022

Data availability

European Nucleotide Archive: *Yponomeuta sedellus* (grey ermine) Accession number PRJEB50749; https://identifiers.org/ena.embl/PRJEB50749 (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The *Y. sedella* 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.4789929.

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 Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

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

Agassiz DJL: **Yponomeutidae**. In: Emmet. A.M. (ed.) *The Moths of Great Britain and Ireland Volume 3 Yponomeutidae - Elachistidae.* Colchester: Harley Books, 1996; 39–114.

Agassiz DJL: **Do small ermine moths sing? Possible stridulatory sound production in Yponomeutidae (Lepidoptera).** *Journal of Natural History.* 2017; **51**(21–22): 1229–1236.
Publisher 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

Clancy SP: **The Immigration of Lepidoptera to the British Isles in 2010.** *The Entomologist's Record and Journal of Variation.* 2013; **125**(3): 81–107.

Davis T: **A review of the status of Microlepidoptera in Britain.** *Butterfly Conservation Report.* [Preprint], 2012; (Accessed: 6 December 2022).
Reference Source

Emmet AM, Langmaid J, Young M: **Micro-Moth Distribution Maps**. no date; (Accessed: 6 December 2022).
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**

Hall P, *et al.*: ***Yponomeuta sedella***. no date; (Accessed: 12 January 2023).
Reference Source

Halstead AJ: **Indoor Meetings.** *Proceedings and transactions of the British Entomological and Natural History Society.* 1985; **18**: 74–82.

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

Hendrikse A: **Activity patterns and sex pheromone specificity as isolating mechanisms in eight species of *Yponomeuta* (Lepidoptera: Yponomeutidae).** *Entomologia Experimentalis et Applicata.* 1979; **25**(2): 172–180.
Publisher Full Text

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**

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**

Liénard MA, Löfstedt C: **Small Ermine Moths: Role of Pheromones in Reproductive Isolation and Speciation**. In: J.D. Allison and R.T. Carde (eds) *Pheromone Communication in Moths: Evolution, Behavior and Application.* Oakland: University of California Press, 2016; 211–224.
Publisher Full Text

Löfstedt C, Herrebout WM, Menken SBJ: **Sex pheromones and their potential role in the evolution of reproductive isolation in small ermine moths (Yponomeutidae).** *Chemoecology.* 1991; **2**(1): 20–28.
Publisher Full Text

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**

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**

Pitkin B, Ellis W, Plant C, *et al.*: ***Yponomeuta sedella* Treitschke, 1832**. *The leaf and stem mines of British flies and other insects*. no date; (Accessed: 12 January 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**

Sterling P, Parsons M: **Field Guide to the Micro-moths of Great Britain and Ireland**. London: Bloomsbury, 2018.
Reference Source

Uliano-Silva M, Ferreira JG, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv.* [Preprint], 2022.
Publisher 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**

Wellcome Sanger Institute: **The genome sequence of the Grey Ermine, *Yponomeuta sedella* (Treitschke, 1832)**. European Nucleotide Archive, [Dataset], 2022.
https://identifiers.org/ena.embl/PRJEB50749

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *bioRxiv.* [Preprint], 2022.
Publisher Full Text