



DATA NOTE

The genome sequence of the Pied Smudge, *Ypsolopha sequella* (Clerck, 1759) [version 1; peer review: 2 approved]

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, William B.V. Langdon²,
Darwin Tree of Life Consortium

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

²University of Oxford, Oxford, Oxfordshire, UK

+ Deceased author

v1 First published: 19 Jan 2023, 8:32
<https://doi.org/10.12688/wellcomeopenres.18768.1>
Latest published: 19 Jan 2023, 8:32
<https://doi.org/10.12688/wellcomeopenres.18768.1>

Abstract

We present a genome assembly from an individual *Ypsolopha sequella* (the Pied Smudge; Arthropoda; Insecta; Lepidoptera; Ypsolophidae). The genome sequence is 867 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules with the Z sex chromosome assembled. The mitochondrial genome has also been assembled and is 15.3 kilobases in length. Gene annotation of this assembly on Ensembl identified 20,394 protein coding genes.

Keywords

Ypsolopha sequella, Pied Smudge, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1 19 Jan 2023	 view	 view

1. **Kay Lucek** , University of Basel, Basel, Switzerland
2. **Guang Yang**, Fujian Agriculture and Forestry University, Fuzhou, China

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.*

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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the Pied Smudge, *Ypsolopha sequella* (Clerck, 1759) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:32 <https://doi.org/10.12688/wellcomeopenres.18768.1>

First published: 19 Jan 2023, 8:32 <https://doi.org/10.12688/wellcomeopenres.18768.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Yponomeutoidea; Ypsolophidae; *Ypsolopha*; *Ypsolopha sequella* (Clerck, 1759) (NCBI:txid1870436).

Background

Ypsolopha sequella (Clerck, 1759) is a micro moth of the family Yponomeutidae and the genus *Ypsolopha*. Adults in this group have relatively elongated forewings that are held close to the body at rest, which is often in a declining position. Of the 13 species in the genus known in Great Britain and Ireland (Agassiz *et al.*, 2013), the adult *Y. sequella* is distinctive for the strongly contrasting black and white colouring in most specimens. Some have a suffusion of black scales in among the white ground colour, and can be quite dark, but the black markings along the dorsum remain visible in most specimens, including a rabbit-shaped blotch below the thorax, for which this moth is often affectionately known as the ‘bunny moth’ (Wheeler, n.d.), although (Porter, 2002) gives it the vernacular ‘Pied Smudge’.

In the UK, adults are on the wing between July and October (Sterling & Parsons, 2018), with a peak in sightings in August (Wheeler, no date). Adults are mostly nocturnal and are attracted to light, though rarely in numbers (Sterling & Parsons, 2012). They can also occasionally be found around the larval foodplants by day, resting high up on the leaves (Langmaid *et al.*, 2018), or dislodged lower down on trunks (Agassiz, 1996). These foodplants are primarily Field maple (*Acer campestre*) but also Sycamore (*Acer pseudoplatanus*). While some *Ypsolopha* species overwinter as adults, *Y. sequella* hibernates as an egg on the twigs of these trees (Agassiz, 1996). The greenish larva, which has the characteristic spindle shape of the genus, can then be found in a flimsy spinning on the leaves in May and June (Agassiz, 1996; Langmaid *et al.*, 2018).

Both *A. campestre* and *A. pseudoplatanus* are frequently grown ornamentally, and the moth can therefore be found where they occur in woodland and where they have been planted in suburban areas (Agassiz, 1996). This planting may have aided the recent expansion of the moth, as it is thought to have for other species feeding on the same plants, such as Maple Prominent (*Ptilodon cucullina*) (Randle *et al.*, 2019; Waring *et al.*, 2017). Indeed, *Y. sequella* is one of several moth species expanding rapidly northwards into Scotland (Emmet *et al.*, no date). Having first been recorded in the country in 1975 and then again in 1997 (Bland, 1998), it is now resident in a few southern counties (Knowler, 2012). Field Maple is scarcer in Scotland than in southern England, and here *Y. sequella* may depend on the more widespread Sycamore (Sterling & Parsons, 2018). Away from Scotland, the moth is common in southern England (Davis, 2012) and found throughout northern and central Europe into the Middle East (Agassiz, 1996).

The genome of *Y. sequella* 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 *Ypsolopha sequella*, based on two specimens from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Y. sequella* (Figure 1) collected from Wytham, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude -1.33). A total of 27-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 or mis-joins and removed seven haplotypic duplications, reducing the assembly length by 0.53% and the scaffold number by 6.83%.

The final assembly has a total length of 866.9 Mb in 150 sequence scaffolds with a scaffold N50 of 28.9 Mb (Table 1). Most (96.87%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome (Figure 2–Figure 5; Table 2). Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size. The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98% (single 96.9%, duplicated 1.2%) using the OrthoDB v10 lepidoptera reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report

The *Y. sequella* genome assembly (GCA_934047225.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Ypsolopha_sequella_GCA_934047225.1/). The resulting annotation includes 20,557 transcribed mRNAs from 20,394 protein-coding genes.



Figure 1. Image of the *Ypsolopha sequella* (ilYpsSequ2) specimen used for genome sequencing.

Table 1. Genome data for *Ypsolopha sequella*, ilYpsSequ2.1.

Project accession data		
Assembly identifier	ilYpsSequ2.1	
Species	<i>Ypsolopha sequella</i>	
Specimen	ilYpsSequ2	
NCBI taxonomy ID	1870436	
BioProject	PRJEB50740	
BioSample ID	SAMEA7519929	
Isolate information	ilYpsSequ2 (PacBio), ilYpsSequ1 (Hi-C, 10X)	
Assembly metrics*		Benchmark
Consensus quality (QV)	65.2	≥ 50
<i>k</i> -mer completeness	100	≥ 95%
BUSCO**	C:98.0%[S:96.9%,D:1.2%], F:0.5%,M:1.5%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	96.87%	≥ 95%
Sex chromosomes	Z	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR9081700, ERR9081701	
10X Genomics Illumina	ERR8571658–ERR8571661	
Hi-C Illumina	ERR8571662	
Genome assembly		
Assembly accession	GCA_934047225.1	
<i>Accession of alternate haplotype</i>	GCA_934041175.1	
Span (Mb)	866.9	
Number of contigs	228	
Contig N50 length (Mb)	20.8	
Number of scaffolds	150	
Scaffold N50 length (Mb)	28.9	
Longest scaffold (Mb)	52.0	
Genome annotation		
Number of protein-coding genes	20,394	

* 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).

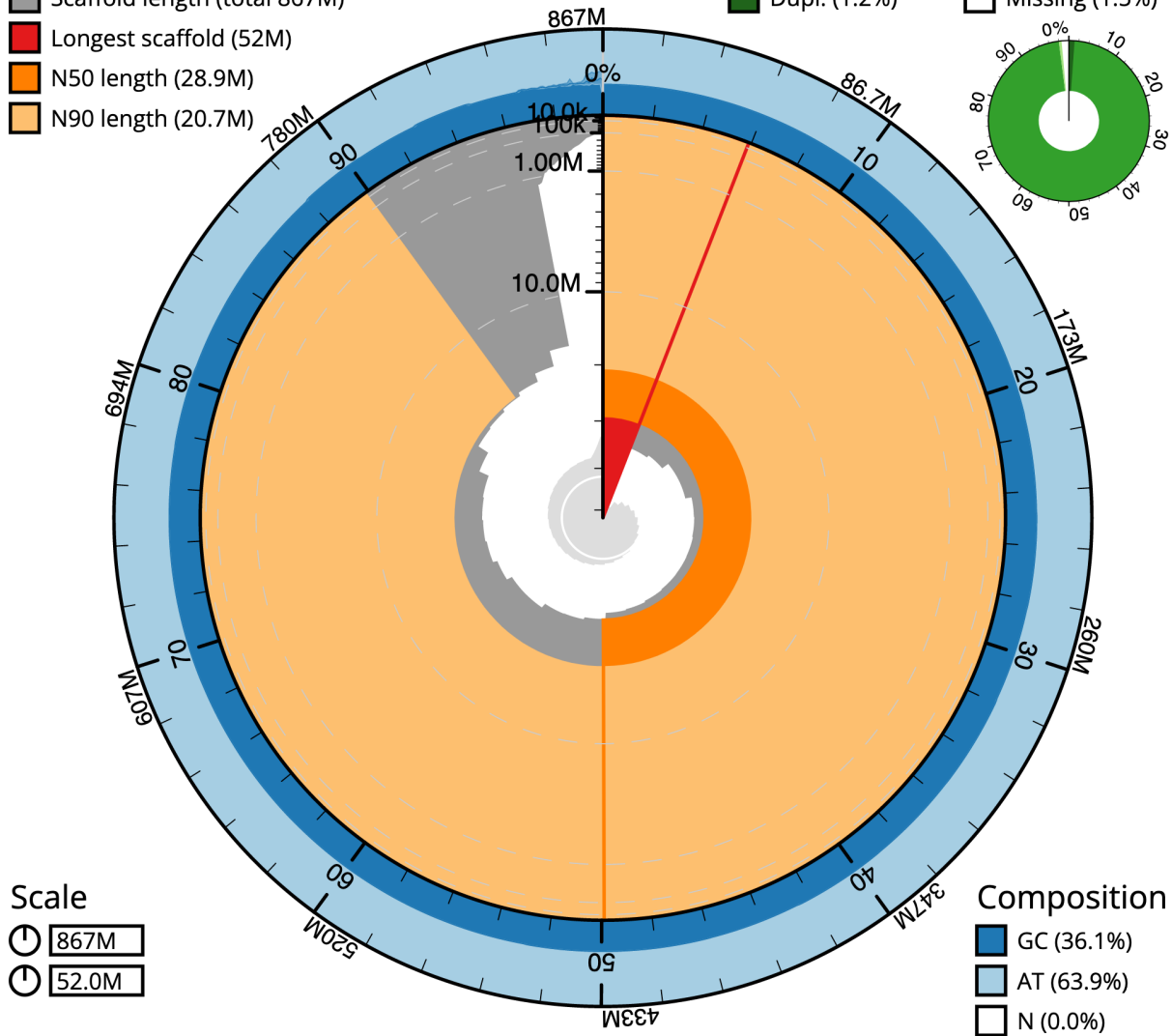
** BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at <https://blobtoolkit.genomehubs.org/view/ilYpsSequ2.1/dataset/CAKOHE01/busco>.

Scaffold statistics

- Log10 scaffold count (total 151)
- Scaffold length (total 867M)
- Longest scaffold (52M)
- N50 length (28.9M)
- N90 length (20.7M)

BUSCO lepidoptera_odb10 (5286)

- Comp. (98.0%)
- Frag. (0.5%)
- Dupl. (1.2%)
- Missing (1.5%)



Dataset: CAKOHE01

Figure 2. Genome assembly of *Ypsolopha sequella*, ilYpsSequ2.1: 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 866,879,050 bp assembly. The distribution of sequence lengths is shown in dark grey with the plot radius scaled to the longest sequence present in the assembly (51,978,369 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 sequence lengths (28,919,749 and 20,718,692 bp), respectively. The pale grey spiral shows the cumulative sequence 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/ilYpsSequ2.1/dataset/CAKOHE01/snail>.

Methods

Sample acquisition and nucleic acid extraction

Two *Y. sequella* specimens (ilYpsSequ1 and ilYpsSequ2) were collected from the main track in Wytham, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude

-1.33) by Douglas Boyes, using a light trap. The specimens were identified by Douglas Boyes and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilYpsSequ1 and ilYpsSequ2

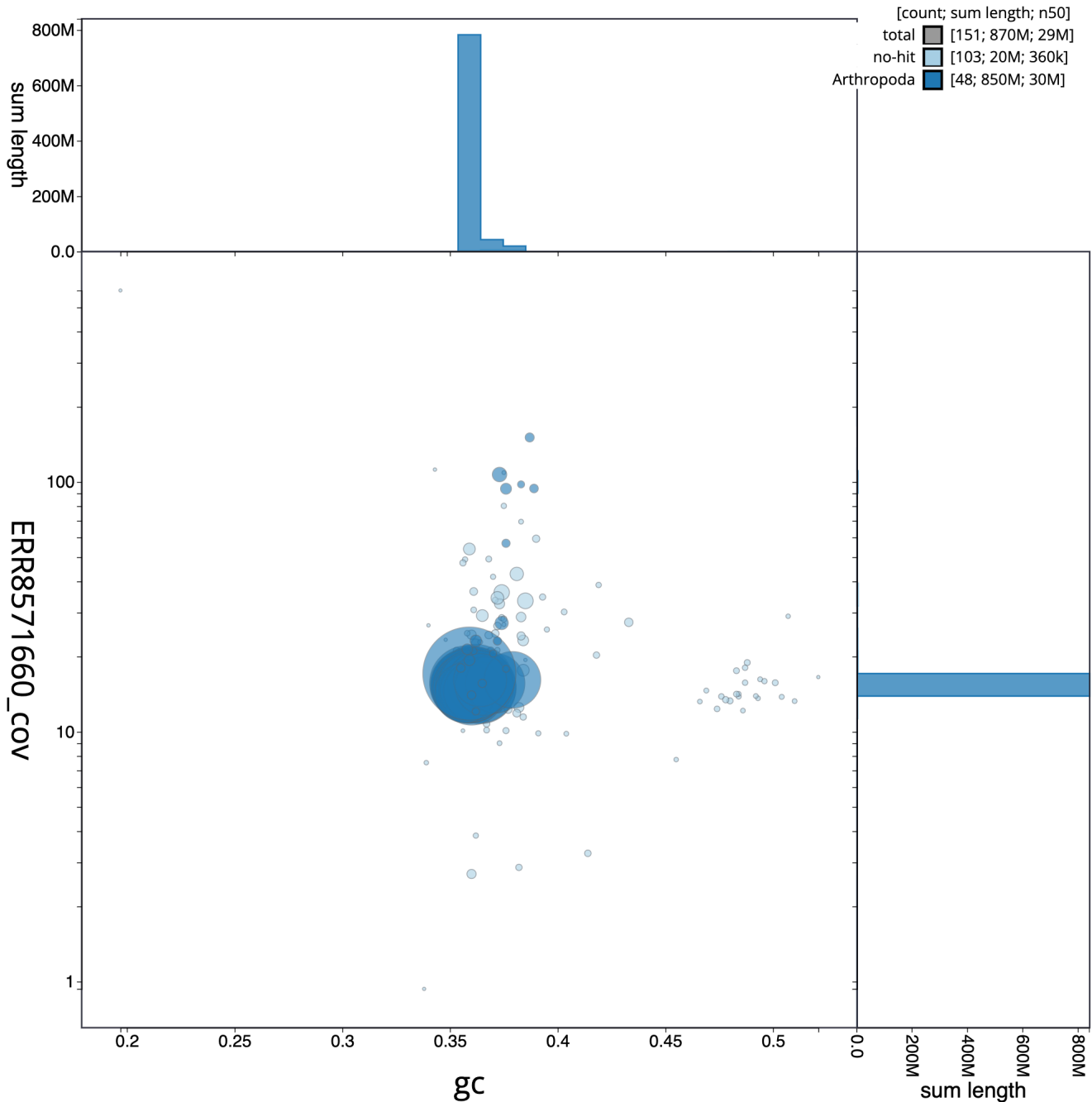


Figure 3. Genome assembly of *Ypsolopha sequella*, ilYpsSeq2.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilYpsSeq2.1/dataset/CAKOHE01/blob>.

specimens were weighed and dissected on dry ice, with tissue from ilYpsSeq1 set aside for Hi-C sequencing. Whole body 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. Low molecular weight DNA was removed

from a 20 ng aliquot of extracted DNA using 0.8X AMPure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. 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

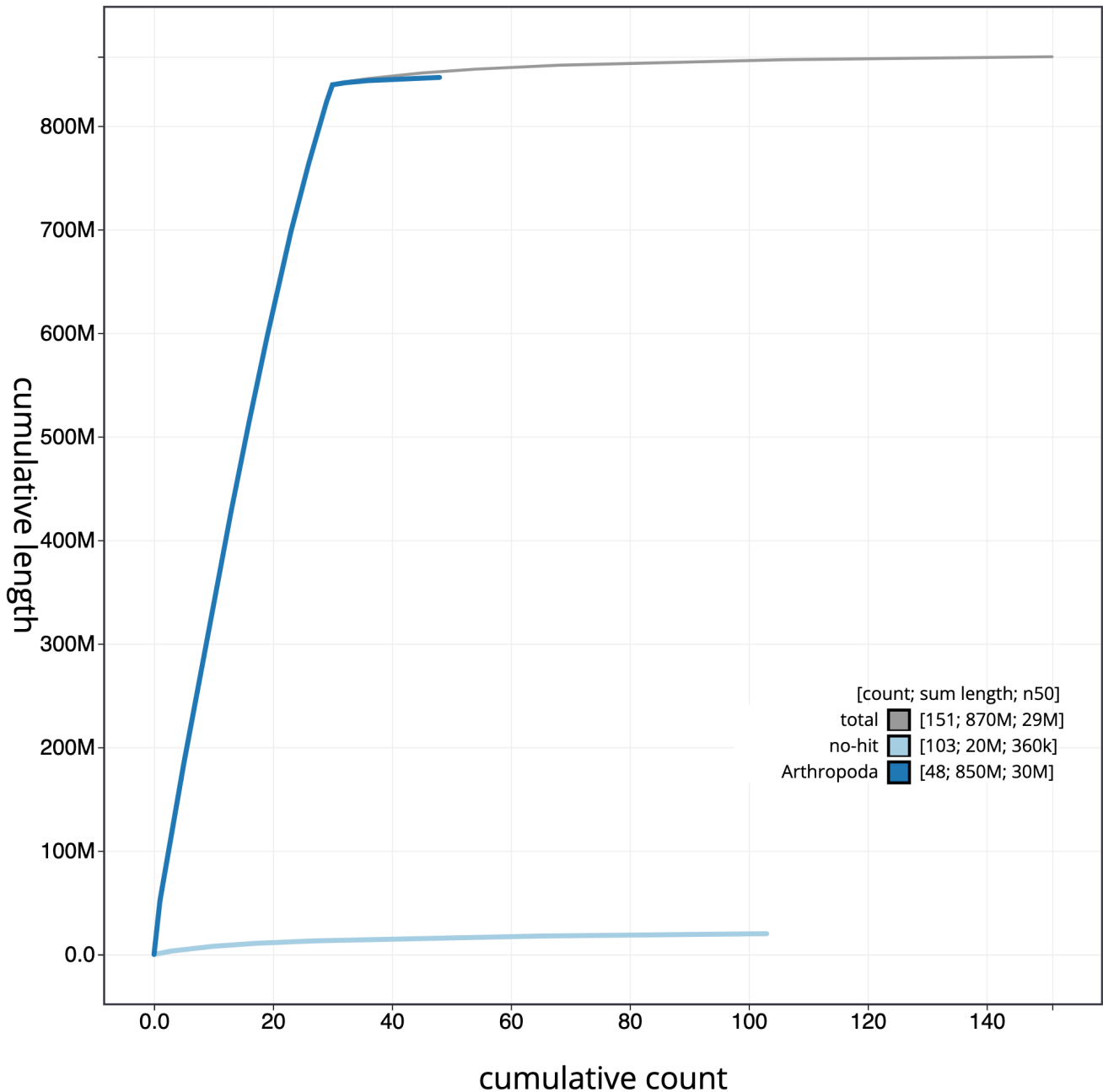


Figure 4. Genome assembly of *Ypsolopha sequella*, iLYpsSequ2.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iLYpsSequ2.1/dataset/CAKOE01/cumulative>.

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 (iLYpsSequ2) and 10X Genomics read cloud (iLYpsSequ1) 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) and HiSeq X Ten (10X) instruments.

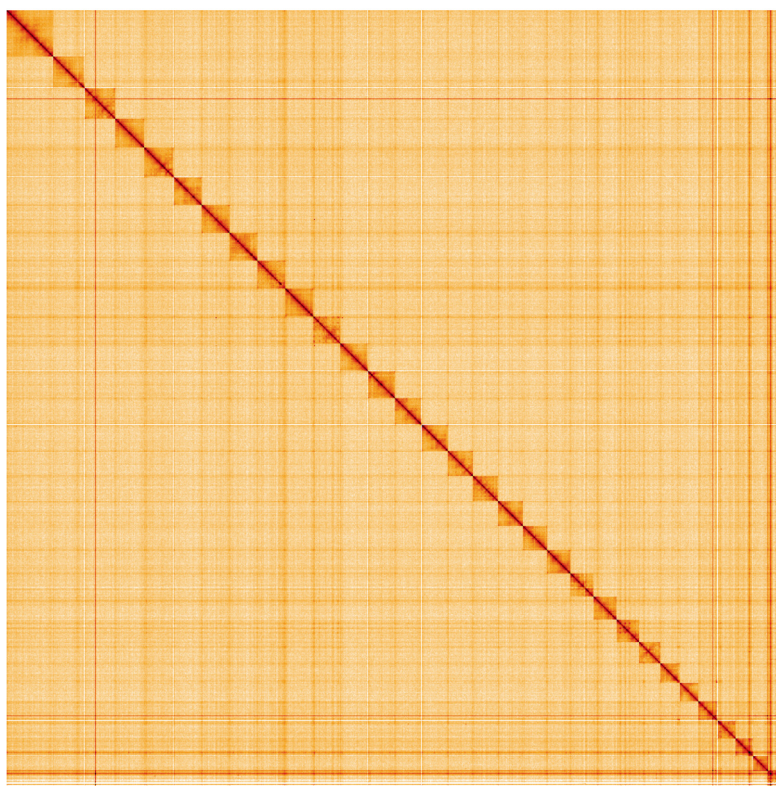


Figure 5. Genome assembly of *Ypsolopha sequella*, iYpsSequ2.1: Hi-C contact map. Hi-C contact map of the iYpsSequ2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/!/?d=A3UWotGzQ8WyAcqeFNsymw>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Ypsolopha sequella*, iYpsSequ2.

INSDC accession	Chromosome	Size (Mb)	GC%
OW203984.1	1	34.25	35.9
OW203985.1	2	33.72	35.8
OW203986.1	3	32.19	35.9
OW203987.1	4	32.19	36.2
OW203988.1	5	30.93	35.8
OW203989.1	6	30.85	36.1
OW203990.1	7	30.71	35.8
OW203991.1	8	30.67	36.4
OW203992.1	9	30.63	36.2
OW203993.1	10	30.5	36.4
OW203994.1	11	30.26	35.9
OW203995.1	12	30.09	36.1
OW203996.1	13	28.92	36.1
OW203997.1	14	28.51	35.7
OW203998.1	15	28.44	35.8

INSDC accession	Chromosome	Size (Mb)	GC%
OW203999.1	16	27.53	35.8
OW204000.1	17	27.51	35.9
OW204001.1	18	26.48	35.9
OW204002.1	19	25.68	36
OW204003.1	20	25.43	36
OW204004.1	21	25.27	36
OW204005.1	22	25.1	36.4
OW204006.1	23	23.3	36
OW204007.1	24	21.35	36.1
OW204008.1	25	21.15	35.7
OW204009.1	26	20.72	37.1
OW204010.1	27	20.15	36.6
OW204011.1	28	18.82	37.9
OW204012.1	29	16.43	36.4
OW203983.1	Z	51.98	35.9
OW204013.1	MT	0.02	19.9

Hi-C data were also generated from iYpsSequ1 using the Arima v2 kit and sequenced on the HiSeq X Ten 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). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then 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 PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), 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 appropriate.

Genome annotation

The BRAKER2 pipeline (Brùna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Ypsolopha sequella* assembly (GCA_934047225.1) in Ensembl Rapid Release.

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	3.4.0	Challis <i>et al.</i> , 2020
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
Hifiasm	0.16.1-r375	Cheng <i>et al.</i> , 2021
HiGlass	1.11.6	Kerpedjiev <i>et al.</i> , 2018
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
MitoHiFi	2	Uliano-Silva <i>et al.</i> , 2021
PretextView	0.2	Harry, 2022
purge_dups	1.2.3	Guan <i>et al.</i> , 2020
YaHS	yahs-1.1.91eebc2	Zhou <i>et al.</i> , 2022

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.

Data availability

European Nucleotide Archive: *Ypsolopha sequella* (pied smudge). Accession number PRJEB50740; <https://identifiers.org/ena.embl/PRJEB50740> (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The *Ypsolopha sequella* 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.4789928>.

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, Beavan SD, Heckford RJ: **Checklist of the Lepidoptera of the British Isles**. London: Royal Entomological Society, 2013; **40**(2): 487–488. [Publisher Full Text](#)
- Agassiz DJL: **Yponomeutidae**. In: A.M. Emmet (ed.), *The Moths of Great Britain and Ireland Volume 3 Yponomeutidae - Elachistidae*. Colchester: Harley Books, 1996; 39–114. [Reference Source](#)
- 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](#)
- Bland KP: **A second Scottish record of *Ypsolopha sequella* (Clerk) (Lep: Yponomeutidae)**. *Entomol's Rec J Var.* 1998; **110**: 30.
- Brůna T, Hoff KJ, Lomsadze A, et al.: **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database**. *NAR Genom Bioinform.* 2021; **3**(1): lqaa108. [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](#)
- 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: **(no date) Micro-Moth Distribution Maps**. (Accessed: 6 December 2022). [Reference Source](#)
- Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing**. 2012. [Reference Source](#)
- Guan D, McCarthy SA, Wood J, et al.: **Identifying and removing haplotypic duplication in primary genome assemblies**. *Bioinformatics.* 2020; **36**(9): 2896–2898. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps**. 2022; (Accessed: 19 October 2022). [Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation**. *GigaScience*. Oxford University Press, 2021; **10**(1): gaa153. [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](#)
- Knowler JT: **2010 Moth Records for Stirlingshire and West Perthshire**. *The Forth Naturalist and Historian.* 2012; **35**: 43–46.
- Langmaid JR, Palmer SM, Young MR: **A Field Guide to the Smaller Moths of Great Britain and Ireland**. 3rd edn. Wokingham: The British Entomological and Natural History Society, 2018. [Reference Source](#)
- Manni M, Berkeley MR, Seppay 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. [Publisher Full Text](#)
- Porter J: **A Label and Checklist of the British Micro-lepidoptera with Vernacular Names**. *Jim Porter Jim Porter.* 2002. [Reference Source](#)
- Randle Z, Evans-Hill L, Parsons MS, et al.: **Atlas of Britain and Ireland's Larger Moths**. Newbury: Pisces Publications, 2019. [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**. Gillingham: British Wildlife Publishing, 2012.
- Sterling P, Parsons M: **Field Guide to the Micro-moths of Great Britain and Ireland**. London: Bloomsbury, 2018. [Reference Source](#)
- Uliano-Silva M, et al.: **MitoHiFi**. 2021; (Accessed: 19 October 2022). [Reference Source](#)
- Waring P, Townsend M, Lewington R: **Field Guide to the Moths of Great Britain and Ireland**. 3rd edn. London: Bloomsbury Publishing, 2017. [Reference Source](#)
- Wellcome Sanger Institute: **The genome sequence of the Pied Smudge, *Ypsolopha sequella* (Clerck, 1759)**, European Nucleotide Archive [Dataset]. [identifiers.org]. 2022. <https://identifiers.org/ena.embl/PRJEB50740>
- Wheeler J: **(no date) *Ypsolopha sequella* (Pied Smudge)**. (Accessed: 6 December 2022). [Reference Source](#)
- Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool**. *bioRxiv [Preprint]*. 2022. [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 09 October 2023

<https://doi.org/10.21956/wellcomeopenres.20809.r65130>

© 2023 Yang G. This is an open access peer review report 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.



Guang Yang

Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

In the present article, the author provided the genome assembly of *Ypsolopha sequella* (the Pied Smudge; Arthropoda, Insecta, Lepidoptera, Ypsolophidae). The total genome size is 867 megabases with 30 chromosomes. Overall, the article is well-written and provides valuable information that could be further used. I just have a few minor suggestions.

1. It would be better to provide some more interpretation for the genome annotation either in terms of table(s) or figure(s).
2. It would be great if the author could provide the evolutionary relationship of *Ypsolopha sequella* with other lepidopteran species.
3. In the Background, please provide details on why the *Ypsolopha sequella* genome sequencing is necessary.
4. Please be consistent in using species names in the text, better to use abbreviated names such as *Y. sequella*.
5. Figures are blurry, better to use high-resolution pictures.
6. In the title of Table 2, the "iYpsSequ2" should be "iYpsSequ2.1".

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** insect molecular biology**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 02 June 2023

<https://doi.org/10.21956/wellcomeopenres.20809.r57965>

© 2023 Lucek K. This is an open access peer review report 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.

**Kay Lucek** 

University of Basel, Basel, Switzerland

The authors present the chromosome level genome assembly of the micro moth *Ypsolopha sequella*. The assembly consists of 30 chromosomes including the Z chromosome, suggesting that a male individual was sequenced. Interestingly, a different individual was used to generate the Hi-C and 10X Chromium data. My only question about the presented note concerns this second individual – was it also a male and did the use of a different individual interfere with the assembly?

Besides the W chromosome, the assembly is highly complete as suggested by the high BUSCO score. Overall, the presented assembly will be of great value to study genome as well as genome size evolution in Lepidoptera.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Speciation genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
