DATA NOTE

The genome sequence of the Dingy Shears, *Fissipunctia ypsillon* (Denis & Schiffermüller, 1775) [version 1; peer review: 1 approved]

Douglas Boyes[1+], Peter W.H. Holland[2],
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, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK

[+] Deceased author

v1 **First published:** 19 Jun 2023, **8**:253
https://doi.org/10.12688/wellcomeopenres.19524.1
Latest published: 19 Jun 2023, **8**:253
https://doi.org/10.12688/wellcomeopenres.19524.1

Abstract

We present a genome assembly from an individual female *Fissipunctia ypsillon* (the Dingy Shears; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 715.1 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the W and Z sex chromosomes. The mitochondrial genome has also been assembled and is 15.46 kilobases in length.

Keywords

Fissipunctia ypsillon, Dingy Shears, genome sequence, chromosomal, Lepidoptera

Open Peer Review

Approval Status ✓

	1
version 1 19 Jun 2023	✓ view

1. **Sivasankaran Kuppusamy**, Loyola College, Chennai, India

Any reports and responses or comments on the article can be found at the end of the article.

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; **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.

How to cite this article: Boyes D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Dingy Shears, *Fissipunctia ypsillon* (Denis & Schiffermüller, 1775) [version 1; peer review: 1 approved]** Wellcome Open Research 2023, **8**:253 https://doi.org/10.12688/wellcomeopenres.19524.1

First published: 19 Jun 2023, **8**:253 https://doi.org/10.12688/wellcomeopenres.19524.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Noctuidae; Xyleninae; *Fissipunctia*; *Fissipunctia ypsillon* (Denis & Schiffermüller, 1775) (NCBI:txid987458).

Background

The Dingy Shears, *Fissipunctia ypsillon*, is a moth found primarily around damp woodlands, riverbanks and wetlands across most of northern and central Europe, with additional scattered records east through Russia and Kazakhstan (GBIF Secretariat, 2022). The larvae are specialist feeders on the leaves of willow (S*alix* sp.) and poplar (*Populus* sp.), spending the daytime hidden beneath loose bark and ascending trees to reach foliage at night (Bretherton *et al.*, 1983; South, 1961). Pupation occurs in soil or in bark crevices. In Britain, the species is widely distributed but rarely abundant across England and Wales, with fewer records from Scotland or Northern Ireland (Randle *et al.*, 2019). In Ireland, there is a small number of records from the east coast (MothsIreland, 2023).

As the common name suggests, the forewings of the adult moth are generally dull greyish-brown, marked with a series of black triangles that, with imagination, resemble a pair of cutting blades. The forewing ground colouration in some individuals has a reddish tinge, while others are dark brown or almost black (Bretherton *et al.*, 1983; South, 1961). The adult is attracted to light and strongly attracted to sugary substances (South, 1961).

A complete genome sequence for *Fissipunctia ypsillon* will facilitate research into the biochemical adaptations underpinning larval feeding habits and contribute to wider comparative studies into lepidopteran evolution.

Genome sequence report

The genome was sequenced from one *Fissipunctia ypsillon* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.31). A total of 40-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 60-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 55 missing joins or mis-joins and removed 4 haplotypic duplications, reducing the assembly length by 1.87% and the scaffold number by 47.62%, and increasing the scaffold N50 by 0.46%.

The final assembly has a total length of 715.1 Mb in 43 sequence scaffolds with a scaffold N50 of 25.1 Mb (Table 1). Most (99.94%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the W and Z sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2).: While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding

Figure 1. Photograph of the *Fissipunctia ypsillon* (ilFisYpsi1) specimen used for genome sequencing.

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 63.5 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.2% (single = 97.7%, duplicated = 0.6%), 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/987458.

Methods

Sample acquisition and nucleic acid extraction

A female *Fissipunctia ypsillon* specimen (ilFisYpsi1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.31) on 2020-06-25, using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilFisYpsi1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Abdomen 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 the 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

Table 1. Genome data for *Fissipunctia ypsillon*, ilFisYpsi1.1.

Project accession data		
Assembly identifier	ilFisYpsi1.1	
Species	*Fissipunctia ypsillon*	
Specimen	ilFisYpsi1	
NCBI taxonomy ID	987458	
BioProject	PRJEB56257	
BioSample ID	SAMEA7701297	
Isolate information	ilFisYpsi1; female: abdomen (DNA sequencing), head and thorax (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	63.2	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.2%[S:97.7%,D:0.6%], F:0.2%,M:1.6%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.94%	*≥ 95%*
Sex chromosomes	W and Z chromosomes	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10355974, ERR10355975, ERR10355976 (ULI)	
10X Genomics Illumina	ERR10297876–ERR10297879	
Hi-C Illumina	ERR10297880	
Genome assembly		
Assembly accession	GCA_947568875.1	
Accession of alternate haplotype	GCA_947568855.1	
Span (Mb)	715.1	
Number of contigs	221	
Contig N50 length (Mb)	11.5	
Number of scaffolds	43	
Scaffold N50 length (Mb)	25.1	
Longest scaffold (Mb)	58.6	

* 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/ilFisYpsi1.1/dataset/CANOQK01/busco.

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.

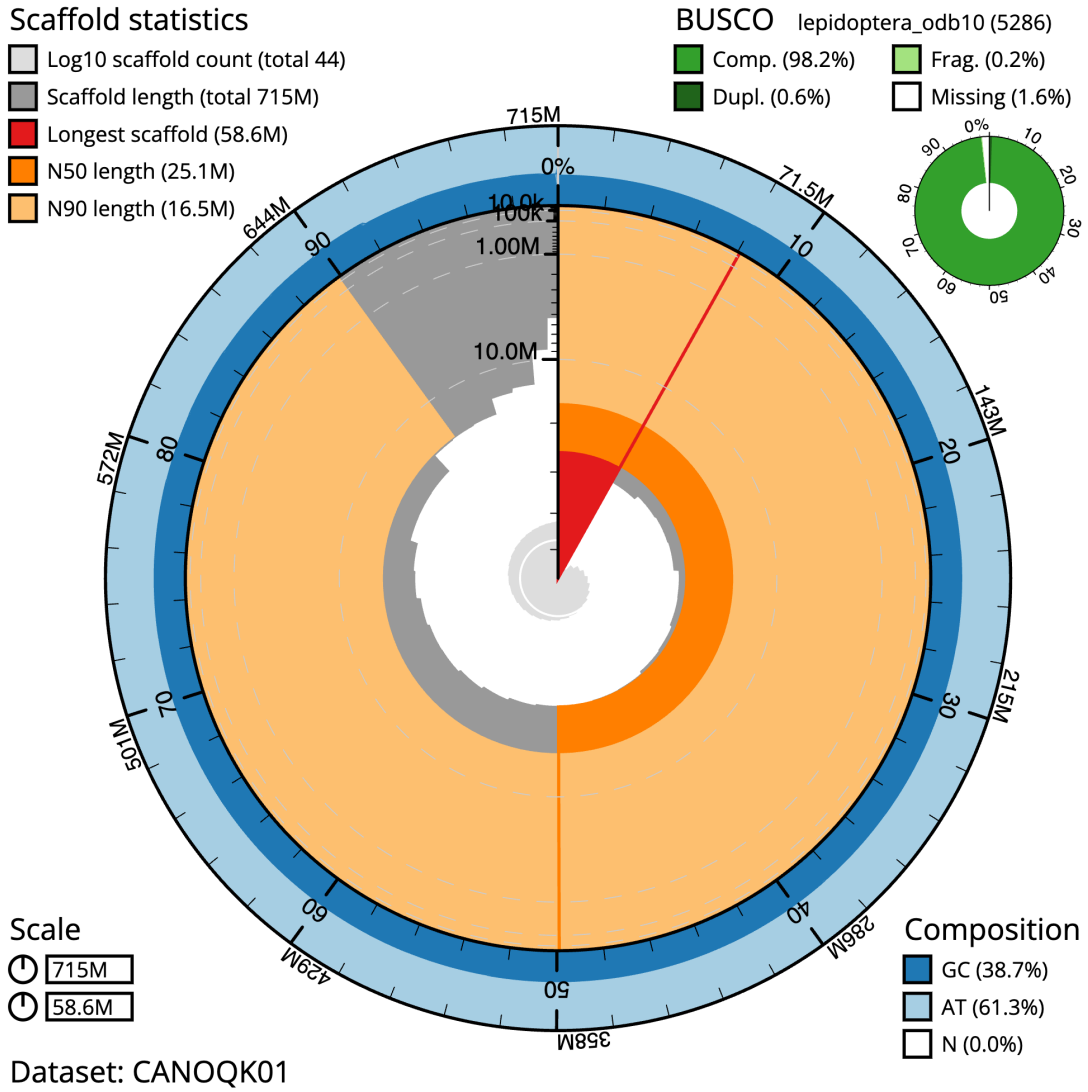

Figure 2. Genome assembly of *Fissipunctia ypsillon*, ilFisYpsi1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 715,143,095 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (58,639,465 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (25,108,161 and 16,474,186 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/ilFisYpsi1.1/dataset/CANOQK01/snail.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud 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 Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from head and thorax tissue of ilFisYpsi1 using the Arima2 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 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.*, 2023). The assembly was checked for contamination and

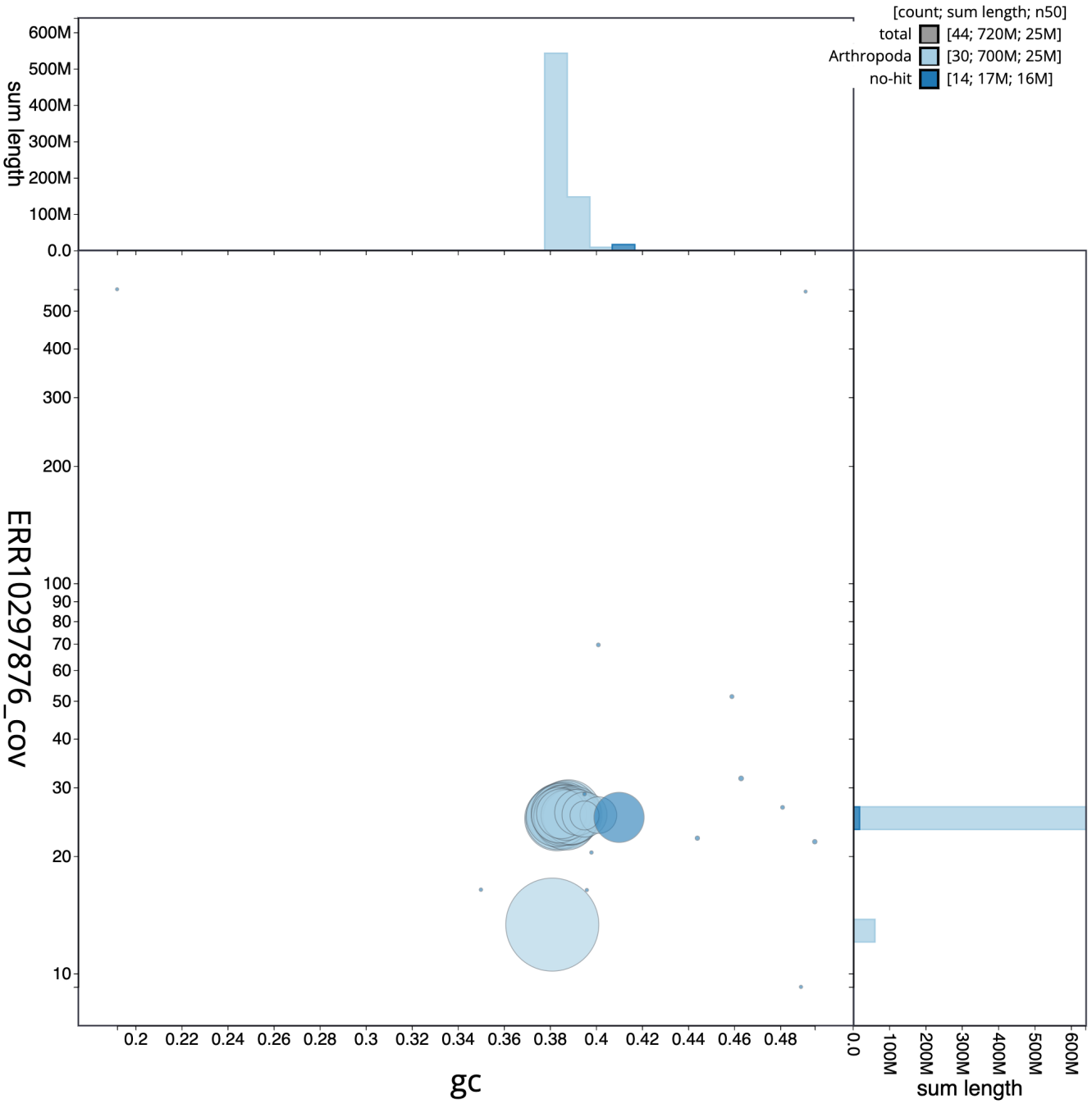

Figure 3. Genome assembly of *Fissipunctia ypsillon*, ilFisYpsi1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilFisYpsi1.1/dataset/CANOQK01/blob.

corrected 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.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics,

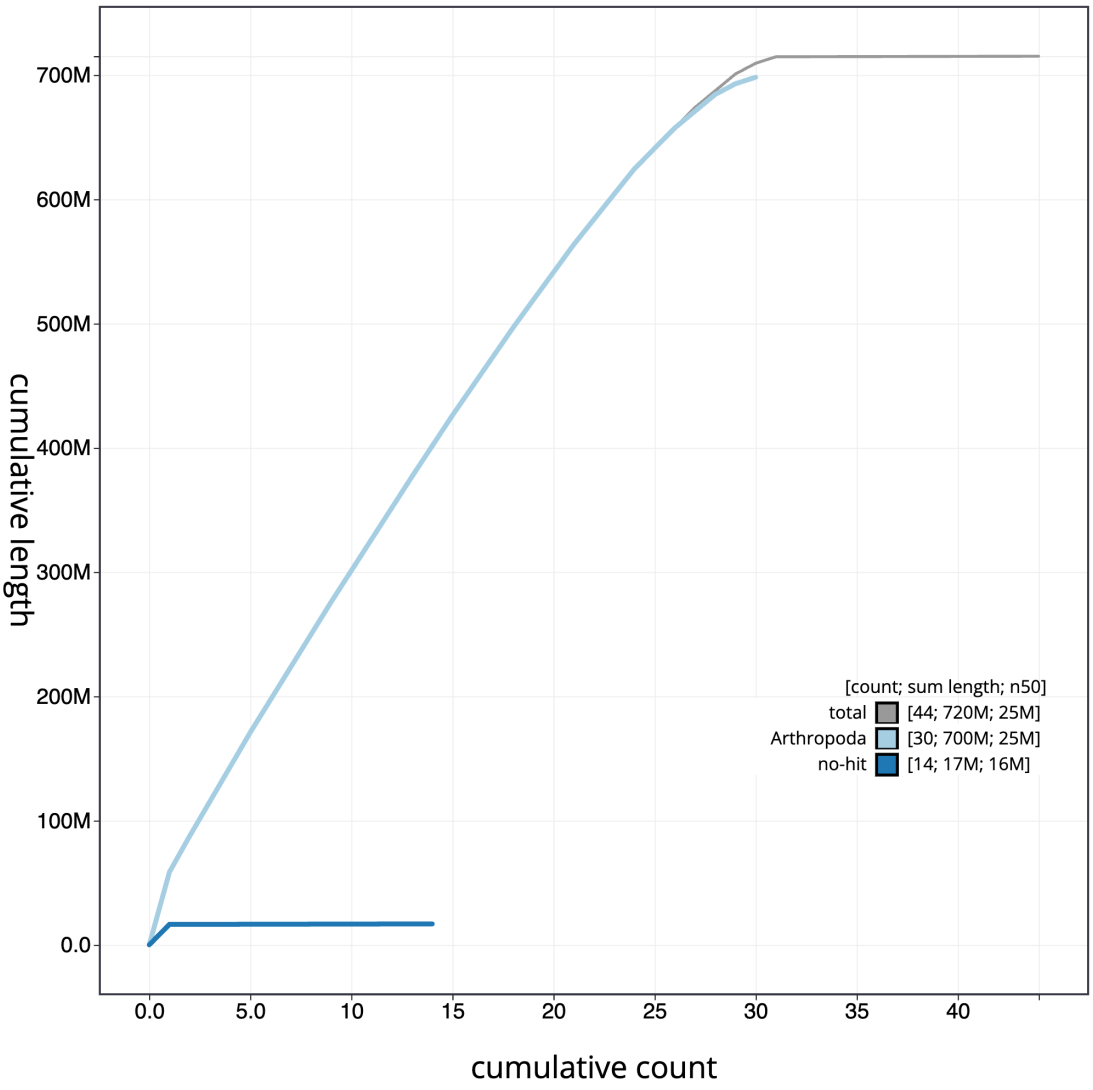

Figure 4. Genome assembly of *Fissipunctia ypsillon*, ilFisYpsi1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilFisYpsi1.1/dataset/CANOQK01/cumulative.

the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b). 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 relevant software tool versions and sources.

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

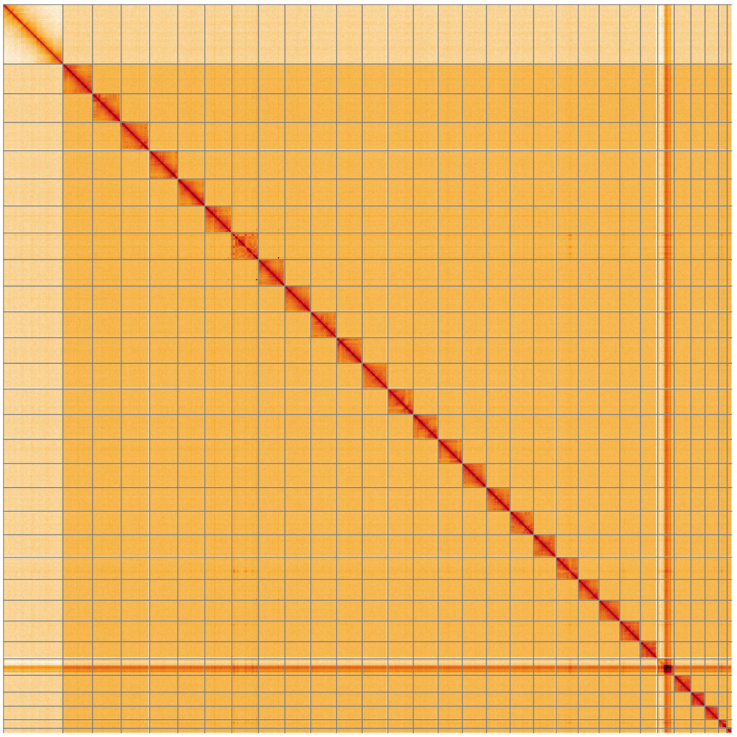

Figure 5. Genome assembly of *Fissipunctia ypsillon*, ilFisYpsi1.1: Hi-C contact map of the ilFisYpsi1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=ASEy_3IkQ6emTqCaFp_pzA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Fissipunctia ypsillon*, ilFisYpsi1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX387644.1	1	28.95	38.5
OX387645.1	2	28.01	39.0
OX387646.1	3	27.8	38.5
OX387647.1	4	27.65	38.5
OX387648.1	5	26.58	38.5
OX387649.1	6	26.49	38.5
OX387650.1	7	26.04	38.5
OX387651.1	8	25.86	38.5
OX387652.1	9	25.38	38.5
OX387653.1	10	25.3	38.5
OX387654.1	11	25.16	38.5
OX387655.1	12	25.11	38.5
OX387656.1	13	24.84	38.5
OX387657.1	14	24.37	38.5
OX387658.1	15	23.86	38.5
OX387659.1	16	23.41	38.5
OX387660.1	17	23.35	38.5
OX387661.1	18	22.91	39.0
OX387662.1	19	22.1	39.0
OX387663.1	20	21.79	39.0
OX387664.1	21	20.35	38.5
OX387665.1	22	20.32	39.0
OX387666.1	23	20.27	38.5
OX387667.1	24	16.6	38.5
OX387669.1	25	16.33	38.5
OX387670.1	26	13.76	39.0
OX387671.1	27	13.17	39.5
OX387672.1	28	8.61	40.0
OX387673.1	29	5.24	39.5
OX387668.1	W	16.47	41.0
OX387643.1	Z	58.64	38.0
OX387674.1	MT	0.02	19.5

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
FreeBayes	1.3.1-17-gaa2ace8	https://github.com/freebayes/freebayes
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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: *Fissipunctia ypsillon* (dingy shears). Accession number PRJEB56257; https://identifiers.org/ena.embl/PRJEB56257. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Fissipunctia ypsillon* 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 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

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

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–9.
PubMed Abstract | Publisher Full Text

Bretherton RF, Goater B, Lorimer RI: **Noctuidae: Cucullinae to Hypeninae.** In: J. Heath and A.M. Emmet (eds) *The Moths and Butterflies of Great Britain and Ireland.* Colchester: Harley Books, 1983.
Reference Source

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

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

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.
Publisher Full Text

GBIF Secretariat: ***Apterogenum ypsillon* (Denis & Schiffermüller) 1775.** *GBIF Backbone Taxonomy.* 2022; (Accessed: 17 February 2023).
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): 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

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

MothsIreland: **73.222 Dingy Shears Apterogenum ypsillon, *MothsIreland*.** 2023; (Accessed: 17 February 2023).
Reference Source

Randle Z, Evans-Hill L, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths.** Newbury: NatureBureau, 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): 1665–80.
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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–2.
PubMed Abstract | Publisher Full Text

South R: **Moths of the British Isles. New edition.** London: Frederick Warne and Co, 1961.
Reference Source

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo*. 2023a; (Accessed: 17 April 2023).
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo*. 2023b; (Accessed: 17 April 2023).
Publisher Full Text

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

Vasimuddin Md, 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

Wellcome Sanger Institute: **The genome sequence of the Dingy Shears, *Fissipunctia ypsillon* (Denis & Schiffermüller, 1775).** European Nucleotide Archive. [dataset], accession number PRJEB56257, 2022.

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* Edited by C. Alkan, 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 03 July 2023

https://doi.org/10.21956/wellcomeopenres.21627.r61025

✔ **Sivasankaran Kuppusamy**
Division of Taxonomy and Biodiversity, Entomology Research Institute, Loyola College, Chennai, Tamil Nadu, India

I appreciate the author for assembling the genome sequence of *Fissipunctia ypsillon* (Denis & Schiffermuller, 1775). The standard methods for sequencing the genome, genome assembling technique and software were followed properly.

Comments

- In the abstract the first sentence can be modified as "The genome of *Fissipuntia ypsillon* (Arthropoda: Insecta: Lepidoptera: Noctuoidea: Noctuidae) was assembled".

- In the methods second paragraph fourth line the authors have mentioned "Abdomen tissue was disrupted...". Usually, the thorax tissue is used for DNA isolation. Why authors have used abdomen tissue for DNA isolation. I think it is a typo error? Authors need to clarify.

- In the whole genome sequence of *Fissipunctia ypsillon*, the authors have assembled the mitochondrial genome. Whole genome sequence of any species contains both mitochondrial and nuclear genes. Why haven't the authors annotated the assembled nuclear genomes? Authors have to give the justification.

- In the BioProject no PRJEB56256 of *Fissipunctia ypsillon*. The GC content of the chromosomes were observed. The decimal slightly varies in the chromosomes (Chr. 1,2,3, 4, 5, 8, 9, 10, 11, 12, 13, 15.... etc.). Authors can modify the decimal in the manuscript.

Overall, the manuscript can be accepted for indexing with minor revision.

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: Phylogenetic analysis of superfamily Noctuoidea moths using the complete mitochondrial genome sequence.

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.