DATA NOTE

The genome sequence of the Shaded Pug, *Eupithecia subumbrata* (Denis & Schiffermüller, 1775) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Denise C. Wawman[2],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]Edward Grey Institute, Department of Biology, University of Oxford, Oxford, England, UK

[+] Deceased author

v1 **First published:** 12 Feb 2024, **9**:29
https://doi.org/10.12688/wellcomeopenres.20829.1
Latest published: 12 Feb 2024, **9**:29
https://doi.org/10.12688/wellcomeopenres.20829.1

Abstract

We present a genome assembly from an individual male *Eupithecia subumbrata* (the Shaded Pug; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 496.2 megabases in span. Most of the assembly is scaffolded into 24 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.17 kilobases in length. Gene annotation of this assembly on Ensembl identified 17,426 protein coding genes.

Keywords

Eupithecia subumbrata, shaded pug, 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; **Wawman DC**: 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) and the Darwin Tree of Life Discretionary Award (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, Wawman DC, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Shaded Pug, *Eupithecia subumbrata* (Denis & Schiffermüller, 1775) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:29 https://doi.org/10.12688/wellcomeopenres.20829.1

First published: 12 Feb 2024, **9**:29 https://doi.org/10.12688/wellcomeopenres.20829.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; Eupithecia (Denis & Schiffermüller, 1775) (NCBI:txid986984).

Background

The Shaded Pug *Eupithecia subumbrata* is a small moth, with a forewing 10–12 mm in length, that can have dark forms which are difficult to distinguish from other members of the family Geometridae (Waring *et al.*, 2017). A typical specimen has a narrow forewing with a straight or slightly curved leading edge and a small white spot placed centrally, although this spot may be absent. Darker edges surround a chalky-white ground colour, and there are multiple dark cross-lines and a light brown streak along each of the three main radial veins, however, these markings can be large and obscure the ground colour (Waring *et al.*, 2017).

The core United Kingdom range of *Eupithecia subumbrata* is limited to south-eastern and southern England excluding Cornwall, but there are small local populations elsewhere including the Isles of Scilly, and parts of northern England, Wales and western Scotland. It is most frequently found on chalk grassland but can be found on roadside verges, salt-marshes, sea-cliffs and in woodland glades (Skinner & Wilson, 2009; Waring *et al.*, 2017).

Eupithecia subumbrata overwinters as a pupa in loose soil and the adults are on the wing from July to September (Waring *et al.*, 2017). The larvae, found from July to September, feed on the flowers of various herbaceous plants, including field scabious *Knautia arvensis*, ragwort *Jacobaea vulgaris*, syn. *Senecio jacobaea*, St John's-wort *Hypericum perforatum*, hawk's-beards *Crepis* spp., flixweed *Descurainia sophia*, dark mullein *Verbascum nigrum*, Spanish catchfly *Silene otites,* and wild marjoram *Origanum vulgare* (Skinner & Wilson, 2009; Waring *et al.*, 2017).

We present a chromosomally complete genome sequence for *Eupithecia subumbrata*, based on one male specimen collected in Wytham Woods, Oxfordshire for the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from one male *Eupithecia subumbrata* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 50-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 4 missing joins or mis-joins and removed 2 haplotypic duplications, reducing the assembly length by 0.99% and increasing the scaffold number by 3.03%.

Figure 1. Photograph of the *Eupithecia subumbrata* (ilEupSubu1) specimen used for genome sequencing.

The final assembly has a total length of 496.2 Mb in 33 sequence scaffolds with a scaffold N50 of 24.5 Mb (Table 1). The snailplot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.89%) of the assembly sequence was assigned to 24 chromosomal-level scaffolds, representing 23 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). Chromosome Z was assigned by alignment to *Eupithecia dodoneata* (GCA_947044415.1) (Boyes *et al.*, 2023), *Eupithecia exiguata* (GCA_947086465.1) and *Eupithecia insigniata* (GCA_947859395.1) (Holland *et al.*, 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 68.2 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 97.9% (single = 97.1%, duplicated = 0.8%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/986984.

Genome annotation report

The *Eupithecia subumbrata* genome assembly (GCA_949316285.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Eupithecia_subumbrata_GCA_949316285.1/Info/Index). The resulting annotation includes 17,607 transcribed mRNAs from 17,426 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Eupithecia subumbrata specimens were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire),

Table 1. Genome data for *Eupithecia subumbrata*, ilEupSubu1.1.

Project accession data		
Assembly identifier	ilEupSubu1.1	
Species	*Eupithecia subumbrata*	
Specimen	ilEupSubu1	
NCBI taxonomy ID	986984	
BioProject	PRJEB59792	
BioSample ID	SAMEA10979184	
Isolate information	ilEupSubu1, male: whole organism (DNA sequencing) ilEupSubu2: whole organism (Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	68.2	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:97.9%[S:97.1%,D:0.8%], F:0.6%,M:1.5%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.89%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 16.17 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10879933	
Hi-C Illumina	ERR10890736	
Genome assembly		
Assembly accession	GCA_949316285.1	
Accession of alternate haplotype	GCA_949316275.1	
Span (Mb)	496.2	
Number of contigs	70	
Contig N50 length (Mb)	13.6	
Number of scaffolds	33	
Scaffold N50 length (Mb)	24.5	
Longest scaffold (Mb)	35.7	
Genome annotation		
Number of protein-coding genes	17,426	
Number of gene transcripts	17,607	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from Rhie *et al.* (2021).

** BUSCO scores based on the lepidoptera_odb10 BUSCO set using version 5.3.2. 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/ilEupSubu1_1/dataset/ilEupSubu1_1/busco.

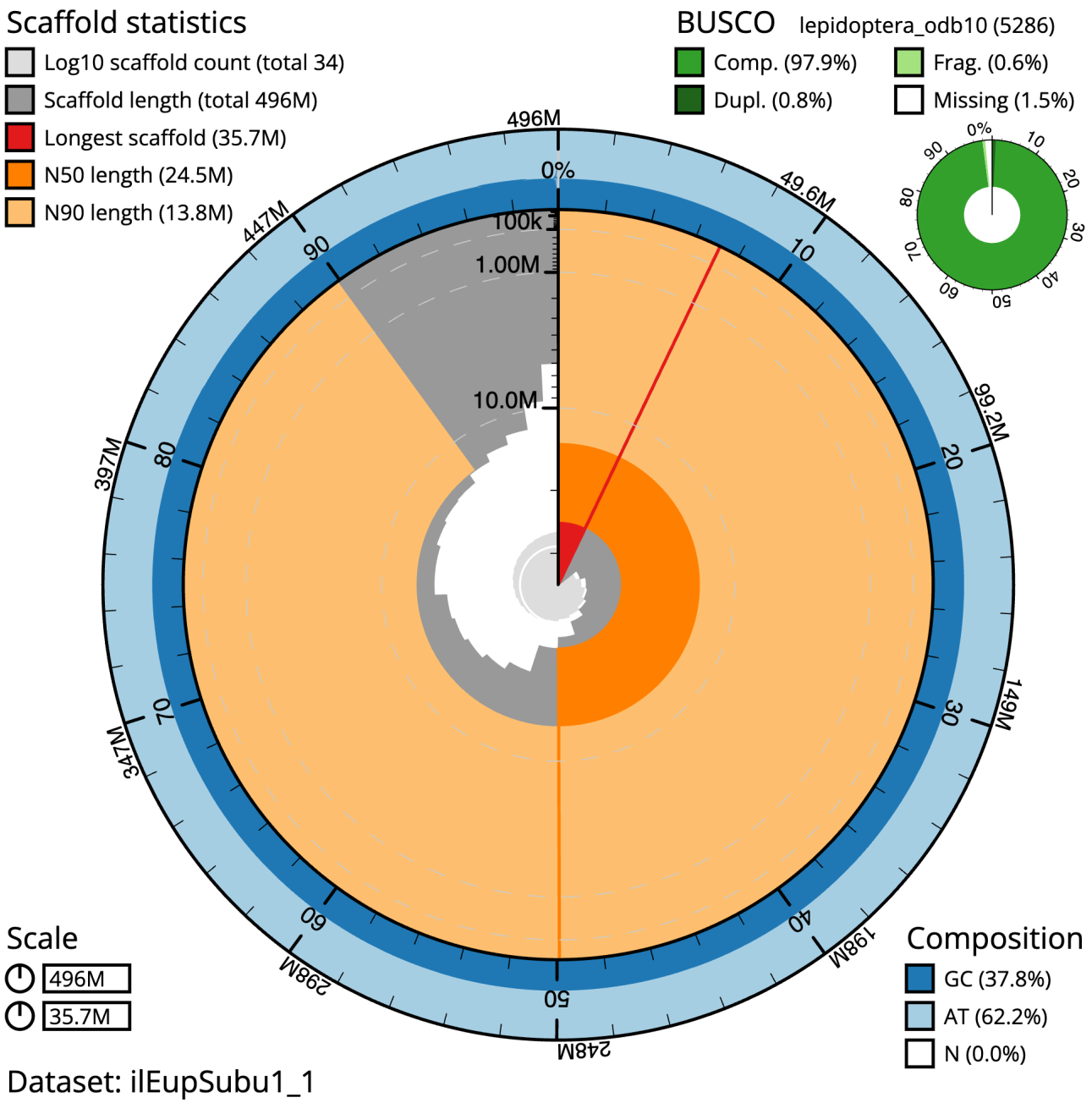

Figure 2. Genome assembly of *Eupithecia subumbrata*, ilEupSubu1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 496,202,682 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (35,696,783 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (24,477,164 and 13,819,472 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/ilEupSubu1_1/dataset/ilEupSubu1_1/snail.

UK (latitude 51.77, longitude –1.34) on 2021-06-16 using a light trap. The specimens were collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice. The specimen with ID Ox001921 (ToLID ilEupSubu1) was used for DNA sequencing, and the specimen with ID Ox001922 (ToLID ilEupSubu2) was used for Hi-C sequencing.

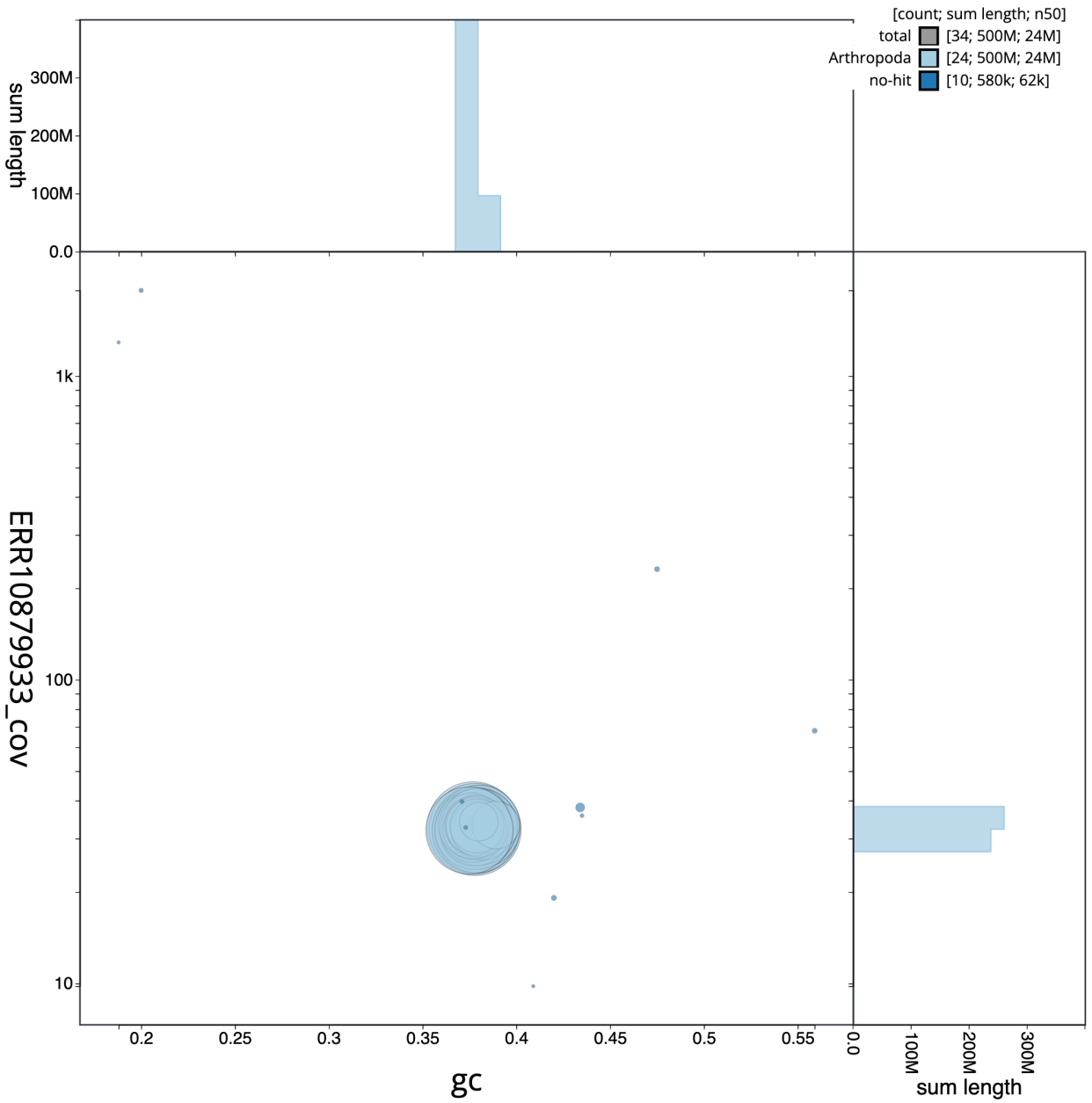

Figure 3. Genome assembly of *Eupithecia subumbrata*, ilEupSubu1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupSubu1_1/dataset/ilEupSubu1_1/blob.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilEupSubu1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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.

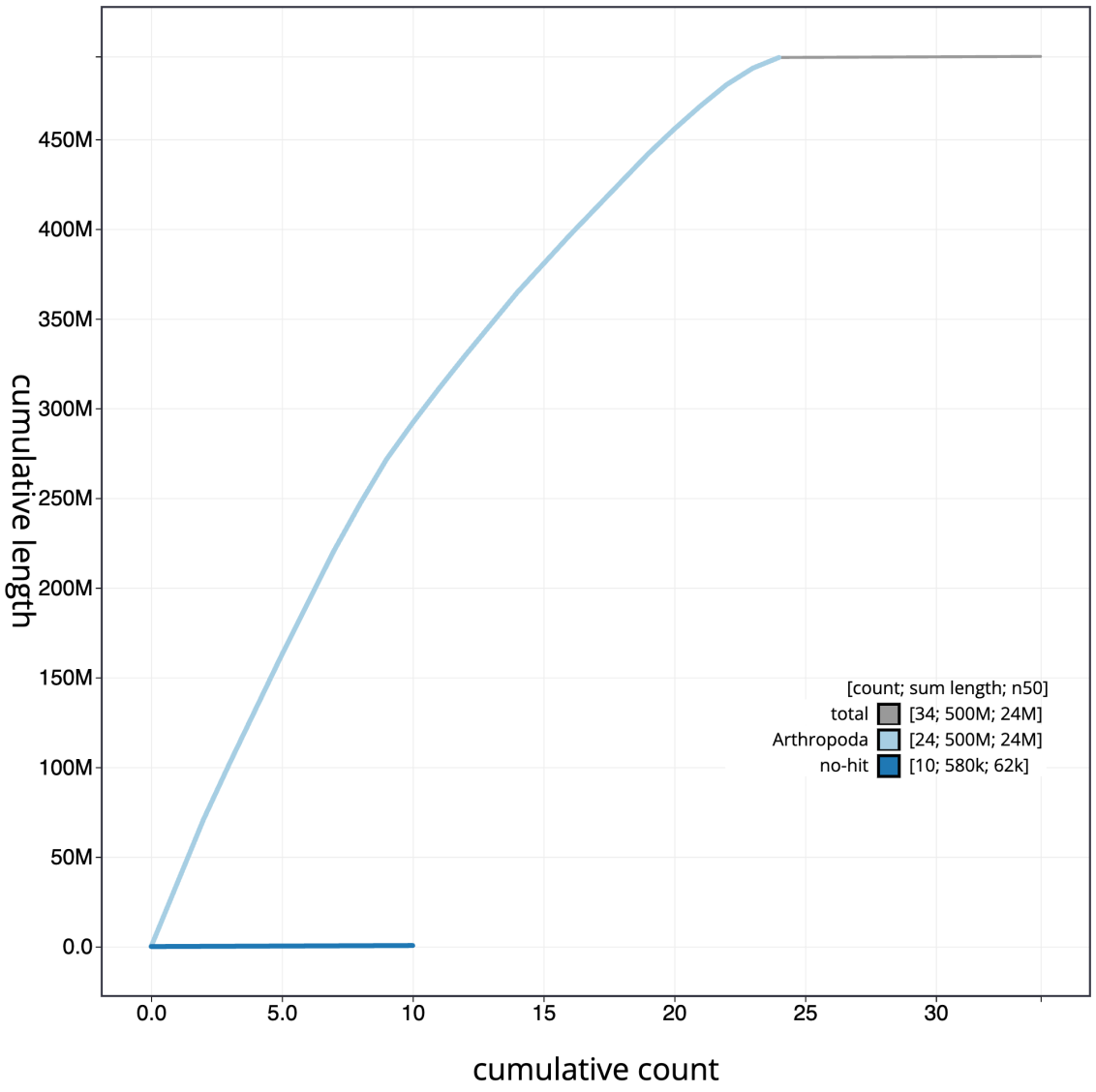

Figure 4. Genome assembly of *Eupithecia subumbrata*, ilEupSubu1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupSubu1_1/dataset/ilEupSubu1_1/cumulative.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL II (HiFi) instrument. Hi-C data were also generated from whole organism tissue of ilEupSubu2 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). 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 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.*, 2023), which runs MitoFinder (Allio *et al.*, 2020)

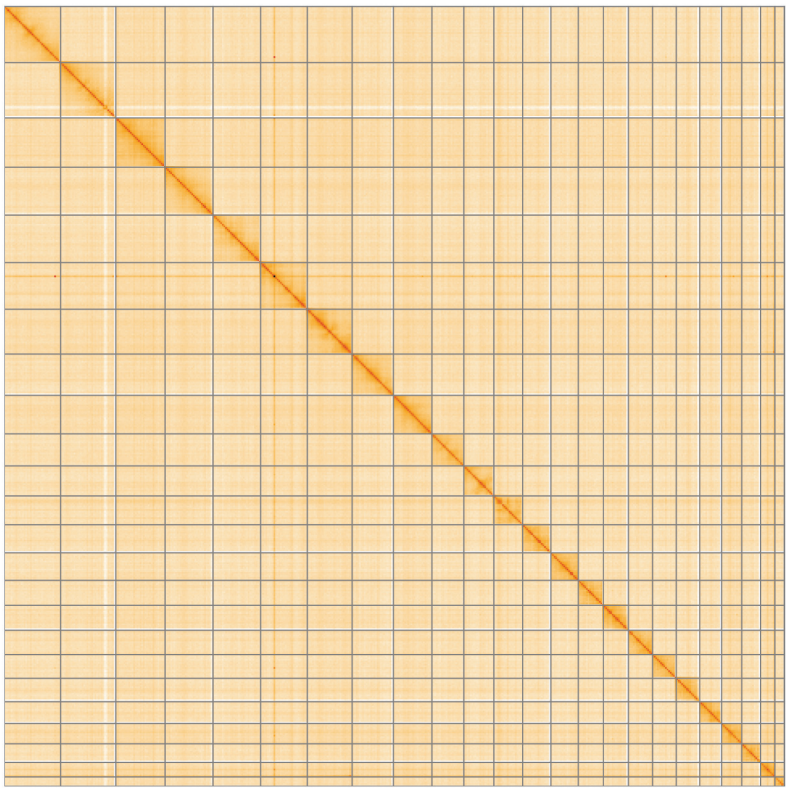

Figure 5. Genome assembly of *Eupithecia subumbrata*, ilEupSubu1.1: Hi-C contact map of the ilEupSubu1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=ABIN1E3KSkm4I0Ecz6Jwdg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eupithecia subumbrata*, ilEupSubu1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX438640.1	1	35.7	37.5
OX438641.1	2	35.09	38.0
OX438643.1	3	30.44	38.0
OX438644.1	4	30.2	38.0
OX438645.1	5	29.64	38.0
OX438646.1	6	28.5	38.0
OX438647.1	7	26.24	38.0
OX438648.1	8	24.48	38.0
OX438649.1	9	20.3	38.0
OX438650.1	10	19.1	38.0
OX438651.1	11	18.34	38.0
OX438627.1	12	17.85	37.5
OX438628.1	13	17.47	37.5
OX438629.1	14	15.89	38.0
OX438630.1	15	15.89	38.0
OX438631.1	16	15.38	38.0
OX438632.1	17	15.08	37.5
OX438633.1	18	14.94	38.0
OX438634.1	19	13.82	38.0
OX438635.1	20	12.89	38.0
OX438636.1	21	11.87	38.0
OX438637.1	22	9.17	39.0
OX438638.1	23	5.96	38.0
OX438642.1	Z	31.38	37.5
OX438639.1	MT	0.02	19.0

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, 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.

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Eupithecia subumbrata* assembly (GCA_949316285.1) in Ensembl Rapid Release.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website here. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
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	1.2a	https://github.com/c-zhou/yahs

Data availability

European Nucleotide Archive: *Eupithecia subumbrata* (shaded pug). Accession number PRJEB59792; https://identifiers.org/ena.embl/PRJEB59792 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Eupithecia subumbrata* 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.7125292.

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 Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

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): 313–319.
PubMed Abstract | Publisher Full Text

Boyes D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab, *et al.*: **The genome sequence of the Oak-tree Pug, *Eupithecia dodoneata* (Guenée, 1858) [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 133.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life Sample Homogenisation: PowerMash.** *Protocols.Io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life Wet Laboratory Protocol Collection V.1.** *Protocols.Io.* 2023b.
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

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

Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, *et al.*: **The genome sequence of the Pinion-spotted Pug, *Eupithecia insigniata* (Hübner, 1790) [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2023; **8**: 305.
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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life Sample Preparation: Triage and Dissection.** *Protocols.Io.* 2023.
Publisher Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

Sheerin E, Sampaio F, Oatley G, *et al.*: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.1.** *protocols.io.* 2023; **8**: 133.
Publisher 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): 3210–3212.
PubMed Abstract | Publisher Full Text

Skinner B, Wilson D: **Colour Identification Guide to the Moths of the British Isles**. Leiden, Netherlands: Brill, 2009.
Reference Source

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Manual SPRI.** *Protocols.Io.* 2023.
Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping**

v1.1.0 - Hebridean Black (1.1.0). *Zenodo.* 2023a; [Accessed 21 July 2023].
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023].
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 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.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free 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

Waring P, Townsend M, Lewington R: **Field Guide to the Moths of Great Britain and Ireland: Third Edition**. Bloomsbury Wildlife Guides, 2017.
Reference Source

Wellcome Sanger Institute: **The genome sequence of the Shaded Pug, *Eupithecia subumbrata* (Denis & Schiffermüller, 1775).** European Nucleotide Archive, [dataset], accession number PRJEB59792, 2023.

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