Check for updates

DATA NOTE

The genome sequence of the Pinion-spotted Pug, *Eupithecia insigniata* (Hübner, 1790)

[version 1; peer review: 1 approved]

Peter W.H. Holland[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, Darwin Tree of Life Consortium

[1]University of Oxford, Oxford, England, UK

v1 **First published:** 12 Jul 2023, **8**:305
https://doi.org/10.12688/wellcomeopenres.19652.1
Latest published: 12 Jul 2023, **8**:305
https://doi.org/10.12688/wellcomeopenres.19652.1

Abstract

We present a genome assembly from an individual male *Eupithecia insigniata* (the Pinion-spotted Pug; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 438.1 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.52 kilobases in length.

Keywords

Eupithecia insigniata, Pinion-spotted Pug, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓

	1
version 1 12 Jul 2023	✓ view

1. **Hans-Peter Fuehrer**, University of Veterinary Medicine Vienna, Vienna, Austria

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: Holland PWH: Investigation, Resources, 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 Holland PWH *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: 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: 1 approved]** Wellcome Open Research 2023, **8**:305 https://doi.org/10.12688/wellcomeopenres.19652.1

First published: 12 Jul 2023, **8**:305 https://doi.org/10.12688/wellcomeopenres.19652.1

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; Geometroidea; Geometridae; Larentiinae; *Eupithecia*; *Eupithecia insigniata* (Hübner, 1790). (NCBI:txid986977).

Background

The genus *Eupithecia*, colloquially known as the 'pug moths', contains over a thousand species of geometrid moth found across Eurasia, Africa and the Americas, and is possibly the largest genus in the family Geometridae (Mironov & Galsworthy, 2014). The genus has a reputation for taxonomic confusion, identification errors and under-recording, since many species are small and drab-coloured with similar markings. Even in a limited geographic region such as Britain and Ireland, with fewer than 50 species, several are difficult to distinguish based on wing pattern alone. One of the striking exceptions is the Pinion-spotted Pug, *Eupithecia insigniata* (known in older literature by the junior synonym *Eupithecia consignata*). The moth is distinctive, with silvery wings crossed by scalloped grey-brown lines, three chocolate brown patches on the leading edge of the forewings and a prominent, elongate, discal spot. There is very little variation between individuals.

E. insigniata is rare or very local across southern and central England with recent records coming from a few widely dispersed sites; there are three historic Scottish specimens (NBN Atlas Partnership, 2021). The scarcity is puzzling since the larval food plants are common and widespread. The species also has a sparse and scattered distribution across mainland Europe and the southern parts of Scandinavia; records from further east should be treated with caution due to possible confusion with *E. insignioides* (GBIF Secretariat, 2022; Mironov & Galsworthy, 2014). A description of the larvae and food plants of *E. insigniata* was first published by Harpur Crewe (1868), although primary credit should be given to Emma Sarah Hutchinson who supplied Henry Harpur Crewe with the eggs which made his study possible. Hutchinson first caught a live specimen of this moth struggling with a spider in her Herefordshire garden in 1864, and from around that time she bred the species continuously in captivity for over 30 years, studying its life-cycle and supplying professional and amateur entomologists with specimens on which they then published (Holland, 2009; Hutchinson, 1892). In captivity, the moth will lay eggs on apple leaves (*Malus* sp.) or hawthorn flowers (*Crataegus* sp.), and the larvae can be reared on the leaves of either plant (Harpur Crewe, 1868; Jackson, 1979; Riley & Prior, 2003). In the wild, larvae are found most frequently on hawthorn, but much remains to be confirmed about ecological requirements of the species (Haggett, 1980). The specimen of *E. insigniata* from which the current genomic work was undertaken was obtained from a garden habitat where apple and hawthorn grow next to each other; the species is recorded regularly at this site. The adult moth flies for a short period in spring after apple blossom has appeared, usually the end of April or start of May in southern England (NBN Atlas Partnership, 2021); larvae feed until July and the pupae overwinter, with the possibility of overwintering for two years (Sterling, 1985).

The complete genome of *Eupithecia insigniata* was determined as part of the Darwin Tree of Life project. The assembled genome will contribute to the growing set of resources for studying insect ecology and evolution.

Genome sequence report

The genome was sequenced from one male *Eupithecia insigniata* (Figure 1) collected from Wallingford, Oxfordshire, UK (51.6, –1.14). A total of 46-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 37 missing joins or mis-joins and removed 9 haplotypic duplications, reducing the scaffold number by 29.31%, and increasing the scaffold N50 by 0.47%.

The final assembly has a total length of 438.1 Mb in 40 sequence scaffolds with a scaffold N50 of 15.4 Mb (Table 1). Most (99.95%) 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). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been

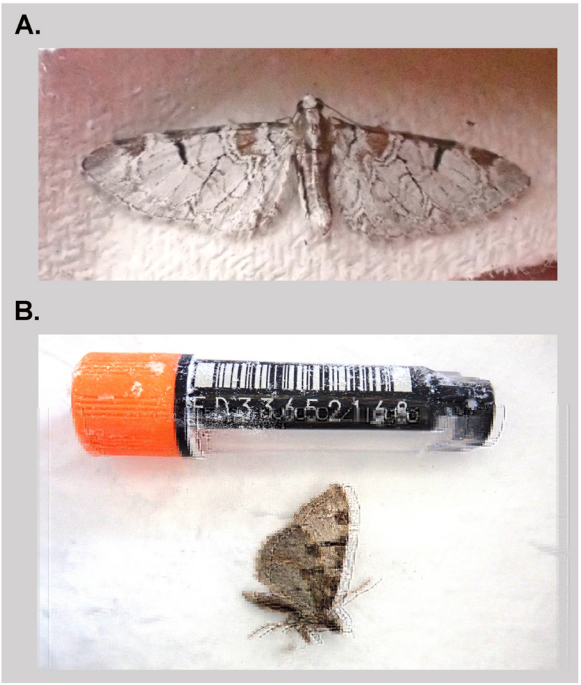


Figure 1. Photograph of the *Eupithecia insigniata* (ilEupInsi1) specimen used for genome sequencing.

Table 1. Genome data for *Eupithecia insigniata*, ilEupInsi1.1.

Project accession data		
Assembly identifier	ilEupInsi1.1	
Species	*Eupithecia insigniata*	
Specimen	ilEupInsi1	
NCBI taxonomy ID	986977	
BioProject	PRJEB57272	
BioSample ID	SAMEA110451585	
Isolate information	ilEupInsi1, male: whole organism (DNA sequencing and Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	62.4	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:97.8%[S:97.3%,D:0.6%], F:0.7%,M:1.5%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.95%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10462076	
Hi-C Illumina	ERR10466810	
Genome assembly		
Assembly accession	GCA_947859395.1	
Accession of alternate haplotype	GCA_947859295.1	
Span (Mb)	438.1	
Number of contigs	139	
Contig N50 length (Mb)	4.9	
Number of scaffolds	41	
Scaffold N50 length (Mb)	15.4	
Longest scaffold (Mb)	20.9	

* 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/ilEupInsi1.1/dataset/CANUEN01/busco.

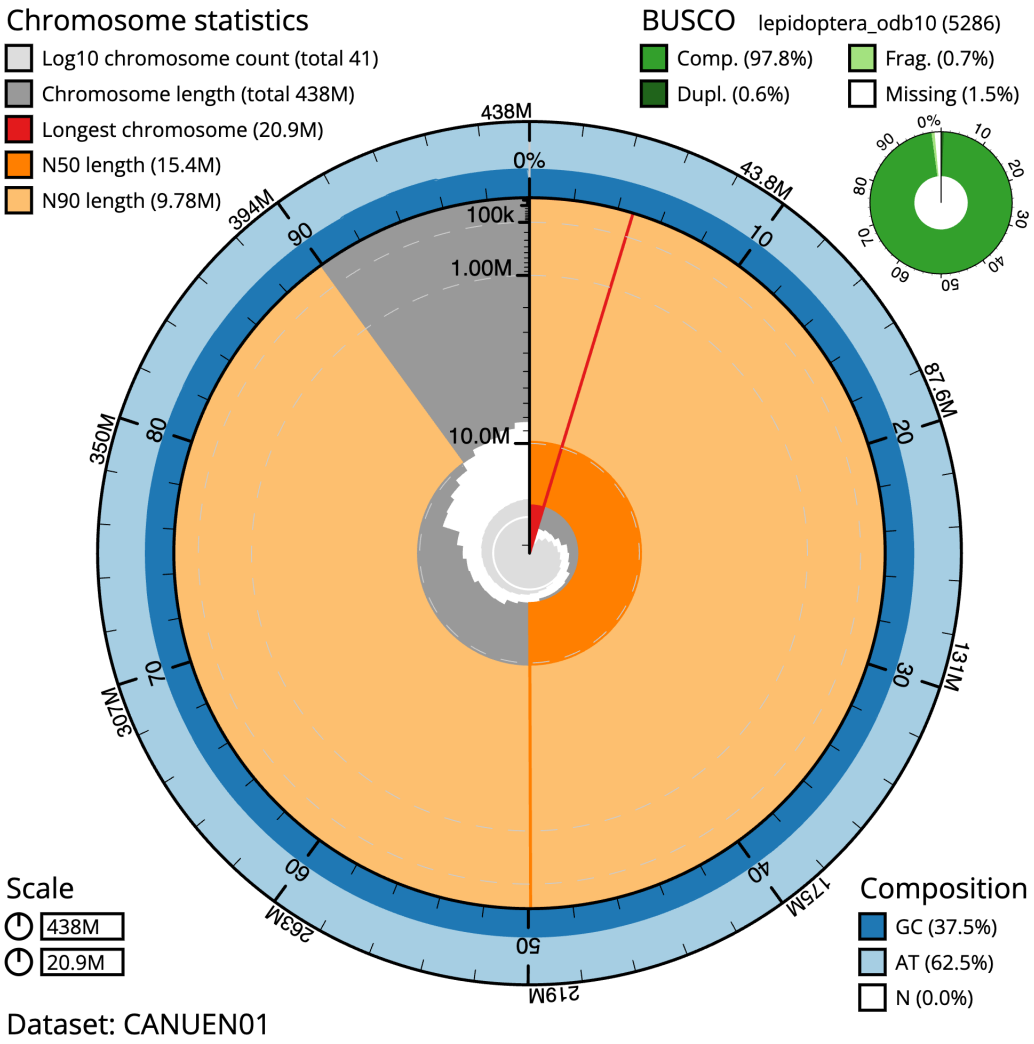


Figure 2. Genome assembly of *Eupithecia insigniata*, ilEupInsi1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 438,119,297 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (20,929,318 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (15,415,171 and 9,781,294 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/ilEupInsi1.1/dataset/CANUEN01/snail.

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 62.4 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 97.8% (single = 97.3%, 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/986977.

Methods

Sample acquisition and nucleic acid extraction

A male *Eupithecia insigniata* (specimen ID Ox002138, ilEupInsi1) was collected using a light trap from a garden in Wallingford, Oxfordshire (biological vice-county Berkshire),

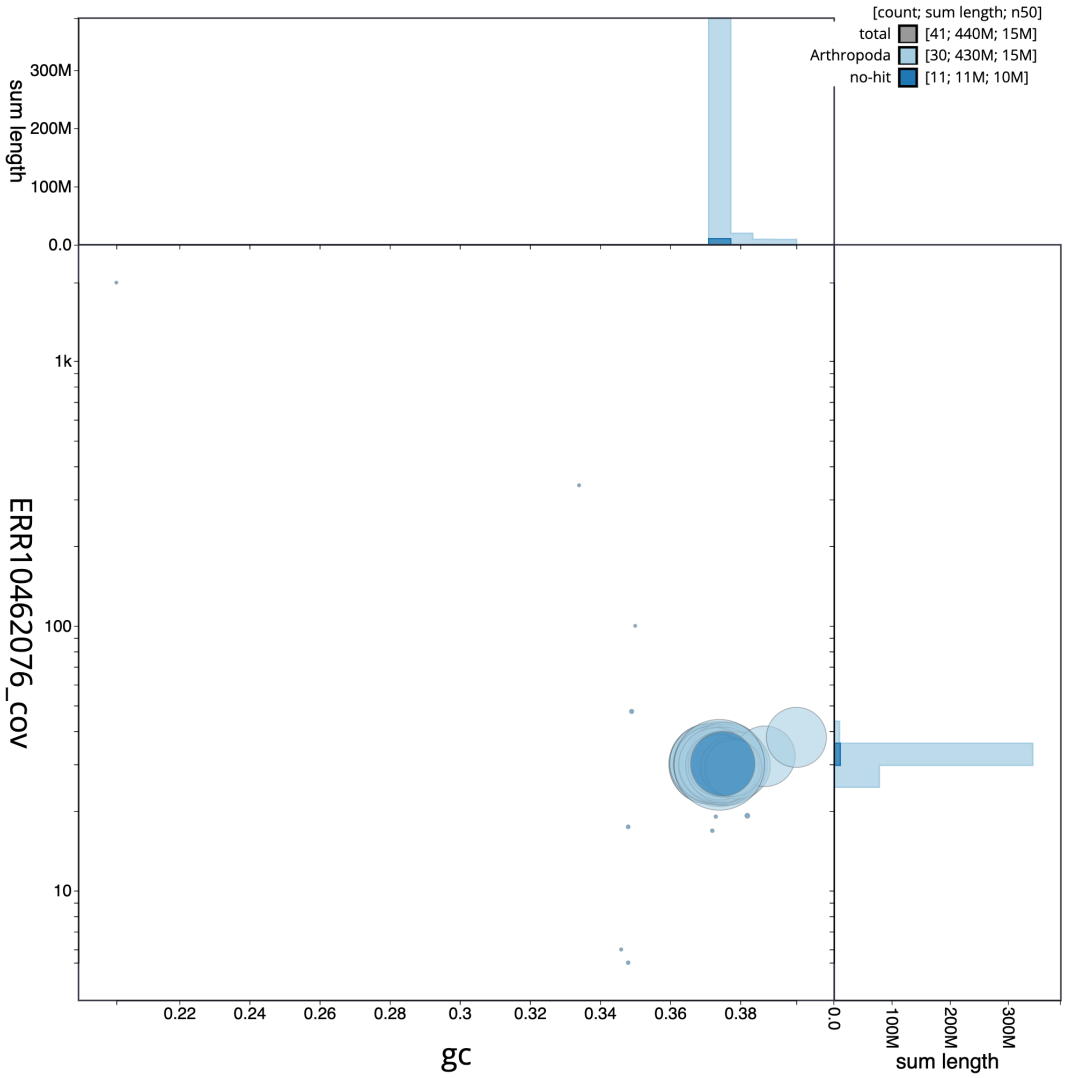


Figure 3. Genome assembly of *Eupithecia insigniata*, ilEupInsi1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupInsi1.1/dataset/CANUEN01/blob.

UK (latitude 51.6, longitude –1.14) on 2022-05-01. Peter Holland (University of Oxford) collected and identified the specimen. The specimen was snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilEupInsi1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Tissue from the whole organism 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 a Pacific Biosciences SEQUEL II (HiFi) instrument. Hi-C data were also generated

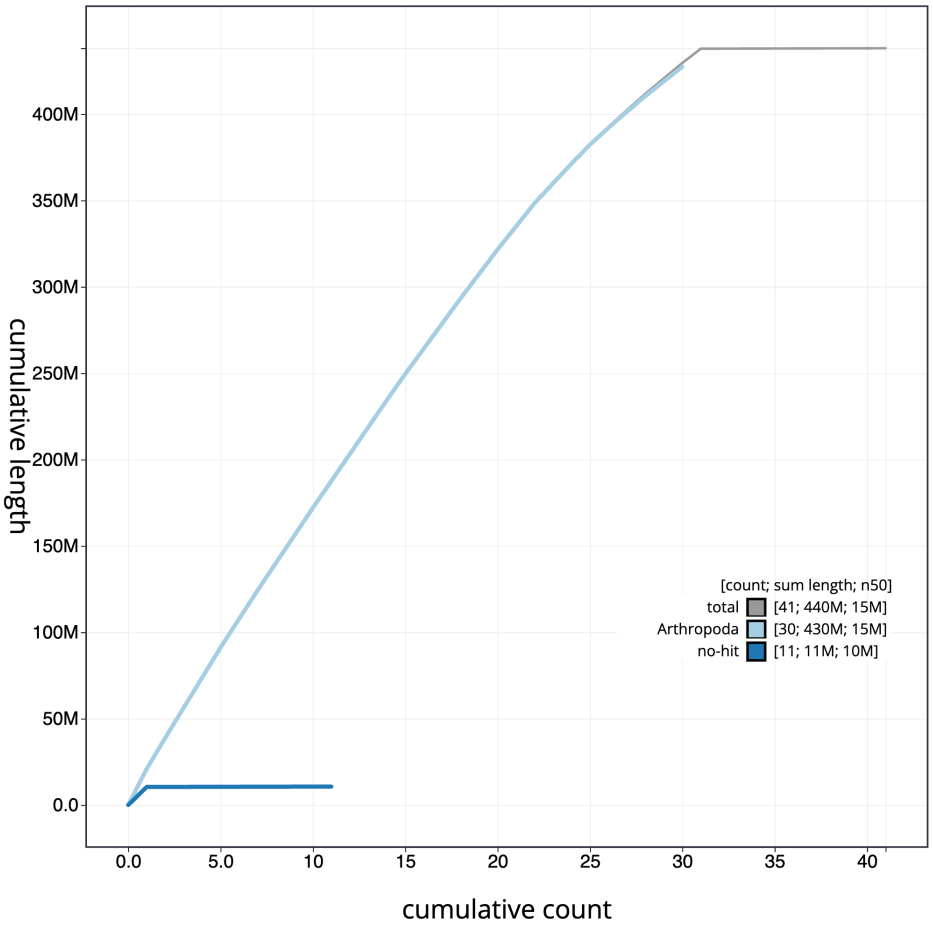


Figure 4. Genome assembly of *Eupithecia insigniata*, ilEupInsi1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEupInsi1.1/dataset/CANUEN01/cumulative.

from tissue of ilEupInsi1 that had been set aside, 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.*, 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, 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

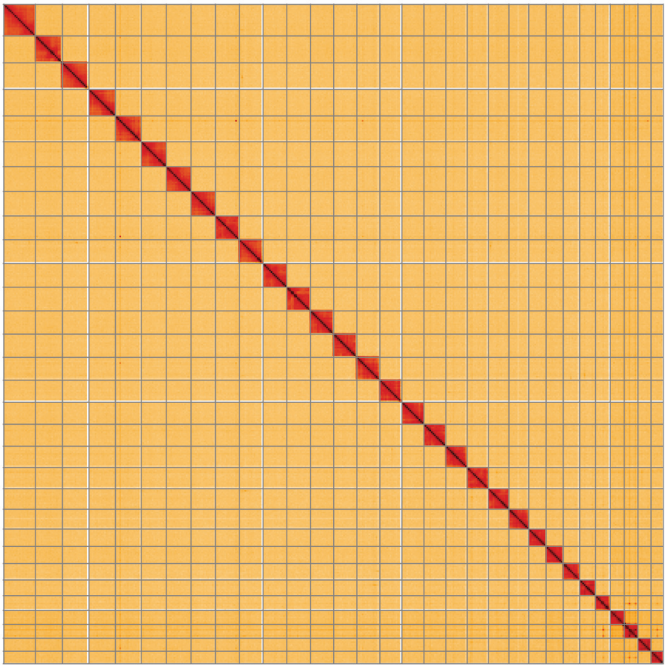


Figure 5. Genome assembly of *Eupithecia insigniata*, ilEupInsi1.1: Hi-C contact map of the ilEupInsi1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=HF_z_x2eQyWPR7MqIKNNfg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eupithecia insigniata*, ilEupInsi1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX401994.1	1	17.87	37.5
OX401995.1	2	17.67	37.5
OX401996.1	3	17.58	37.5
OX401997.1	4	17.15	37.5
OX401998.1	5	16.58	37.5
OX401999.1	6	16.42	37.5
OX402000.1	7	16.26	37.0
OX402001.1	8	15.87	37.5
OX402002.1	9	15.75	37.5
OX402003.1	10	15.73	37.0
OX402004.1	11	15.73	37.0
OX402005.1	12	15.44	37.5
OX402006.1	13	15.42	37.0
OX402007.1	14	15.19	37.0
OX402008.1	15	14.67	37.0
OX402009.1	16	14.61	37.5
OX402010.1	17	14.55	37.5
OX402011.1	18	14.15	37.5
OX402012.1	19	14.05	37.5
OX402013.1	20	13.71	37.5
OX402014.1	21	13.12	37.5
OX402015.1	22	11.52	37.5
OX402016.1	23	11.38	38.0
OX402017.1	24	10.97	37.5
OX402018.1	25	10.44	37.5
OX402019.1	26	9.78	37.5
OX402020.1	27	9.29	38.5
OX402021.1	28	9.11	39.5
OX402022.1	29	8.7	37.5
OX402023.1	30	8.26	38.0
OX401993.1	Z	20.93	37.5
OX402024.1	MT	0.02	20.5

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
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	1.1a.2	https://github.com/c-zhou/yahs

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.

Data availability

European Nucleotide Archive: *Eupithecia insigniata* (pinion-spotted pug). Accession number PRJEB57272; https://identifiers.org/ena.embl/PRJEB57272. (Wellcome Sanger Institute, 2023)

The genome sequence is released openly for reuse. The *Eupithecia insigniata* 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

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

GBIF Secretariat: ***Eupithecia insigniata* (Hübner, 1790).** *GBIF Backbone Taxonomy.* 2022; [Accessed 8 March 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

Haggett GM: **The occurrence and foodplant in Britain of *Eupithecia insigniata* (Huebner) (Lepidoptera: Geometridae).** *Entomol Gaz.* 1980; **31**: 221–226.

Harpur Crewe H: **Description of the larvae of *Eupithecia consigniata,* Bork.** *Entomologist's Monthly Magazine.* 1868; **5**: 72.

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

Holland P: **The 'Queen of Entomologists' Part 1: Mrs Hutchinson and the Pinion-spotted Pug.** *Bulletin of the Amateur Entomologists' Society.* 2009; **68**: 113–116.

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

Hutchinson T: **Herefordshire Lepidoptera, *Transactions of the Woolhope Naturalists' Field Club*.** 1892; **1887**: 104–119.

Jackson BC: **A note on breeding the Pinion-spotted Pug, *Eupithecia insigniata* (Hubner). Lep., Geometridae.** *Proceedings of the British Entomological and Natural History Society.* 1979; **12**: 76–77.

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

Mironov V, Galsworthy A: **The Eupithecia of China: a revision.** Leiden, The Netherlands: Brill NV, 2014.
Reference Source

NBN Atlas Partnership: ***Eupithecia insigniata* (Hübner, 1790) Pinion-spotted Pug.** *NBN Atlas.* 2021; [Accessed 8 March 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): 1665–80.
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

Riley AM, Prior G: **The British and Irish Pug Species.** In: *British and Irish Pug Moths-a Guide to their Identification and Biology.* Leiden, The Netherlands: Brill, 2003; 33–145.
Reference Source

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

Sterling DH: ***Eupithecia* - delayed emergence.** *Entomologists' Record.* 1985; **97**: 93.

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

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
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.* 2022.
Publisher Full Text

Vasimuddin M, 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 Pinion-spotted Pug, *Eupithecia insigniata* (Hübner, 1790).** European Nucleotide Archive. [dataset], accession number PRJEB57272, 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

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 15 July 2024

https://doi.org/10.21956/wellcomeopenres.21769.r86822

© **2024 Fuehrer H.** 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.

✔ **Hans-Peter Fuehrer**
University of Veterinary Medicine Vienna, Vienna, Austria

The author describes a genome assembly from an individual pinion-spotted pug (*Eupithecia insigniata;* 438.1 megabases; 31 chromosomal pseudomolecules; Z sex chromosome). Moreover, the mt genome was assembled (15.52 kb). This study was performed within the Darwin Tree of life project.

The quality of the manuscript is excellent. A nice overview on the biology and distribution of this moth species is given. Methodology (from sampling to genome analysis) is clear.

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: Medical Entomology, Parasitology

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.