DATA NOTE

The genome sequence of the Small Angle Shades, *Euplexia lucipara* (Linnaeus, 1758) [version 1; peer review: 3 approved]

Douglas Boyes[1+], Owen T. Lewis[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]Department of Biology, University of Oxford, Oxford, England, UK

[+] Deceased author

v1 **First published:** 09 Nov 2023, **8**:509
https://doi.org/10.12688/wellcomeopenres.20242.1
Latest published: 09 Nov 2023, **8**:509
https://doi.org/10.12688/wellcomeopenres.20242.1

Abstract

We present a genome assembly from an individual male *Euplexia lucipara* (the Small Angle Shades; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 661.8 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 15.37 kilobases in length. Gene annotation of this assembly on Ensembl identified 20,395 protein coding genes.

Keywords

Euplexia lucipara, Small Angle Shades, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 09 Nov 2023	✓ view	✓ view	✓ view

1. **Savarimuthu Ignacimuthu**, St. Xavier's College, Thirunelveli, India
Sivasankaran Kuppusamy, Loyola College, Chennai, India
2. **Jerome H L Hui**, The Chinese University of Hong Kong, Hong Kong, Hong Kong
3. **Vlad Dincă**, University of Oulu, Oulu, Finland

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; **Lewis OT**: Writing – Original Draft Preparation;

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, Lewis OT, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Small Angle Shades, *Euplexia lucipara* (Linnaeus, 1758) [version 1; peer review: 3 approved]** Wellcome Open Research 2023, **8**:509 https://doi.org/10.12688/wellcomeopenres.20242.1

First published: 09 Nov 2023, **8**:509 https://doi.org/10.12688/wellcomeopenres.20242.1

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; Hadeninae; *Euplexia;* Euplexia lucipara (Linnaeus, 1758) (NCBI:txid987933).

Background

The Small Angle Shades (*Euplexia lucipara*) is a noctuid moth with a wide distributed in a variety of habitats across Britain and Ireland (Randle *et al.*, 2019). Globally, its range extends across temperate Europe and Asia to Japan (GBIF Secretariat, 2023). In appearance, it varies little across its range, and has a distinct resting posture with folder forewings that resemble a dead leaf (Waring *et al.*, 2017).

This is one of a relatively small set of British and Irish insects that feeds frequently on bracken (*Pteridium aquilinum* (L.) Kuhn) and other ferns (Lawton, 1982); its larvae also feed on a wide variety of angiosperms (Henwood *et al.*, 2020). This species has a single generation in the UK, over-wintering as a pupa.

The genome of the Small Angle Shades, *Euplexia lucipara*, 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 *Euplexia lucipara*, based on one male specimen from Wytham Woods, Oxfordshire.

Genome sequence report

The genome was sequenced from one male *Euplexia lucipara* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 26-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 51-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 27 missing joins or mis-joins and removed 4 haplotypic duplications, reducing the assembly length by 0.46% and the scaffold number by 22.83%, and increasing the scaffold N50 by 1.13%.

Figure 1. Photograph of the *Euplexia lucipara* (ilEupLuci1) specimen used for genome sequencing.

The final assembly has a total length of 661.8 Mb in 71 sequence scaffolds with a scaffold N50 of 22.7 Mb (Table 1). Most (99.62%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. A summary of the assembly statistics is shown in Figure 2, 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. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). 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 57.5 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 99.0% (single = 98.5%, duplicated = 0.5%), 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/987933.

Genome annotation report

The *Euplexia lucipara* genome assembly (GCA_921972225.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Euplexia_lucipara_GCA_921972225.1/Info/Index). The resulting annotation includes 20,590 transcribed mRNAs from 20,395 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Euplexia lucipara* (specimen ID Ox000606, ToLID ilEupLuci1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-07-05 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilEupLuci1 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

Table 1. Genome data for *Euplexia lucipara*, ilEupLuci1.1.

Project accession data		
Assembly identifier	ilEupLuci1.1	
Assembly release date	2021-12-17	
Species	*Euplexia lucipara*	
Specimen	ilEupLuci1	
NCBI taxonomy ID	987933	
BioProject	PRJEB46851	
BioSample ID	SAMEA7701470	
Isolate information	ilEupLuci1, male: abdomen (DNA sequencing), head and thorax (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.5	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:99.0%[S:98.5%,D:0.5%], F:0.3%,M:0.7%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.62%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6907906	
10X Genomics Illumina	ERR6688636, ERR6688637, ERR6688638, ERR6688639	
Hi-C Illumina	ERR6688635	
Genome assembly		
Assembly accession	GCA_921972225.1	
Accession of alternate haplotype	GCA_921956305.1	
Span (Mb)	661.8	
Number of contigs	128	
Contig N50 length (Mb)	14.9	
Number of scaffolds	71	
Scaffold N50 length (Mb)	22.7	
Longest scaffold (Mb)	37.2	
Genome annotation		
Number of protein-coding genes	20,395	
Number of gene transcripts	20,590	

* 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_odb10BUSCO 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/Euplexia%20lucipara/dataset/CAKLHH01.1/busco.

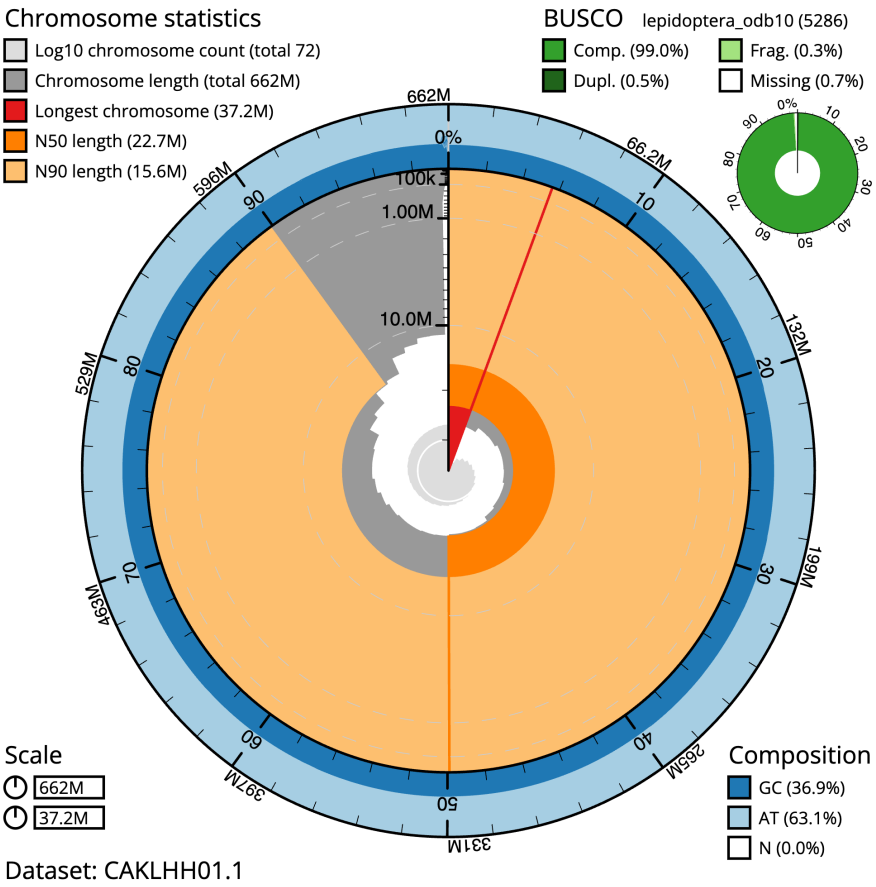

Figure 2. Genome assembly of *Euplexia lucipara*, ilEupLuci1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 661,825,765 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (37,151,478 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (22,669,833 and 15,555,094 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/Euplexia%20lucipara/dataset/CAKLHH01.1/snail.

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 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 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 ilEupLuci1 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

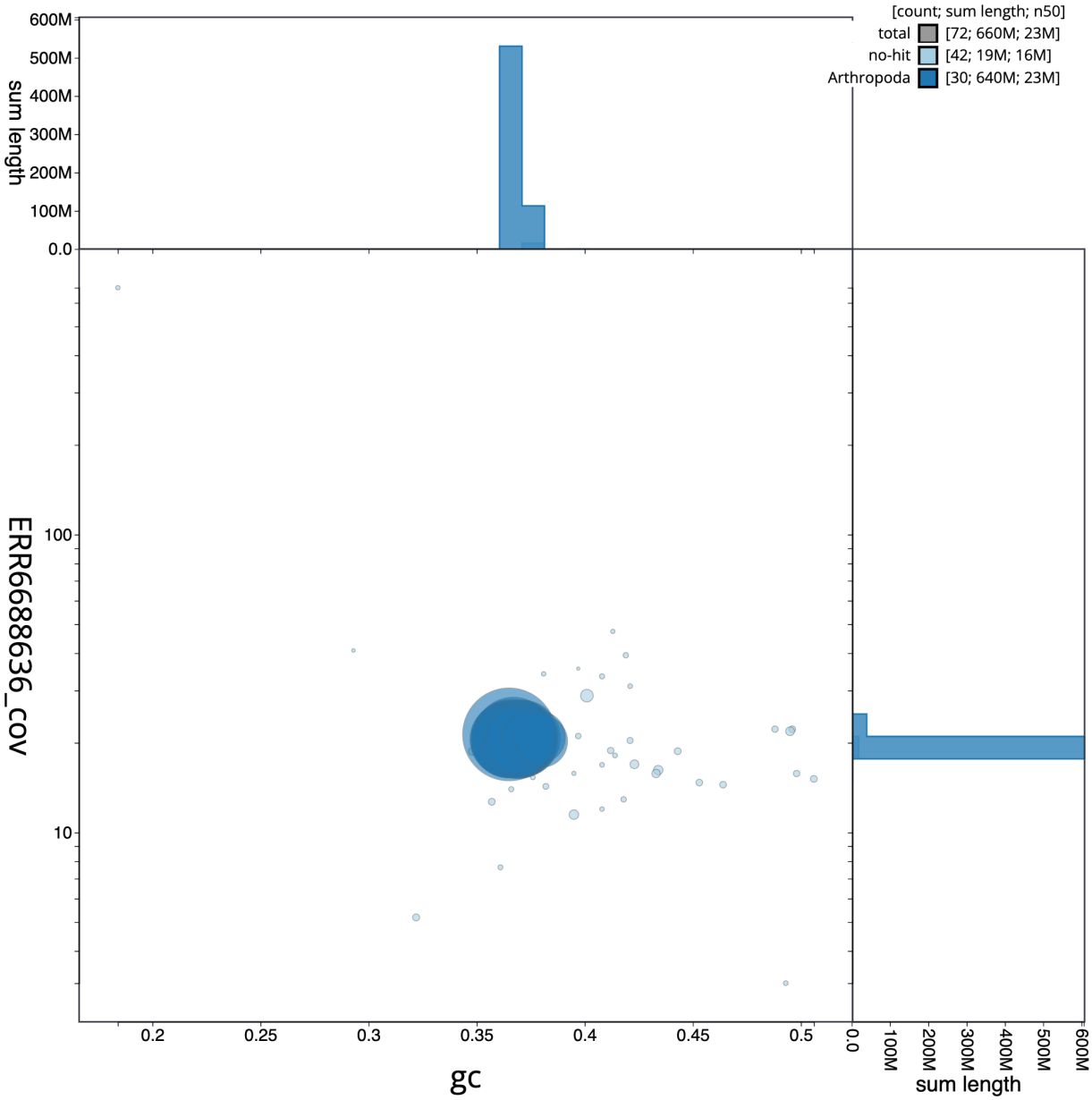

Figure 3. Genome assembly of *Euplexia lucipara*, ilEupLuci1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Euplexia%20lucipara/dataset/CAKLHH01.1/blob.

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 SALSA2 (Ghurye *et al.*, 2019). 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) 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

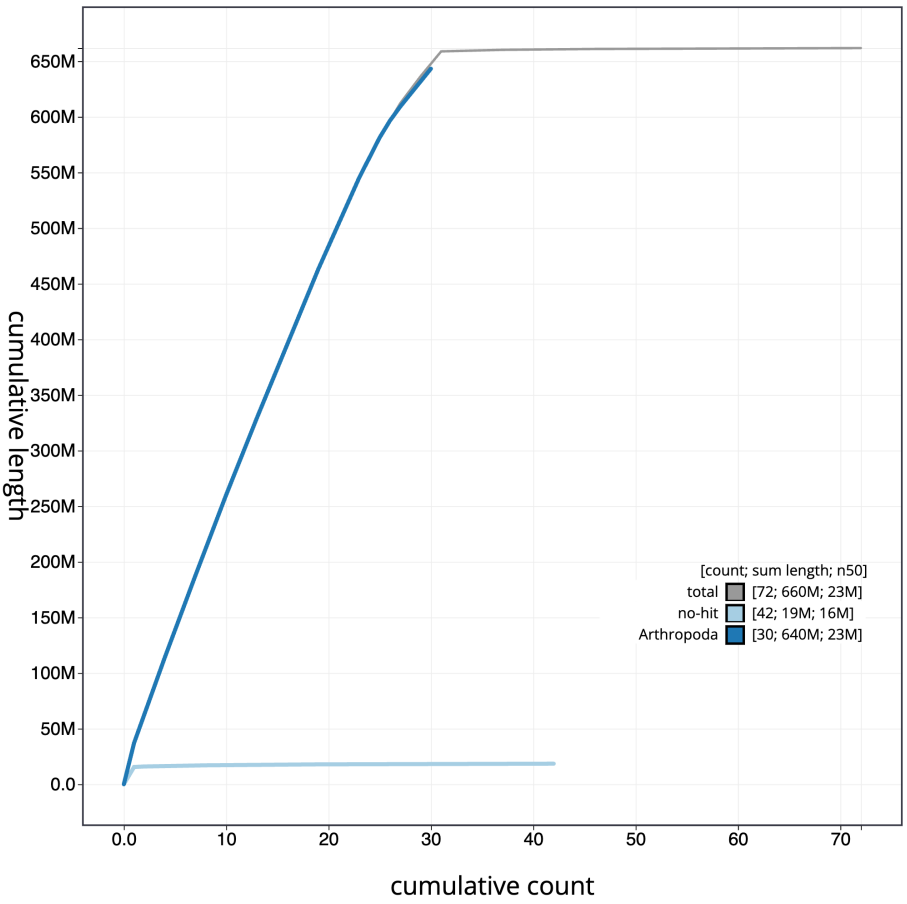

Figure 4. Genome assembly of *Euplexia lucipara*, ilEupLuci1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Euplexia%20lucipara/dataset/CAKLHH01.1/cumulative.

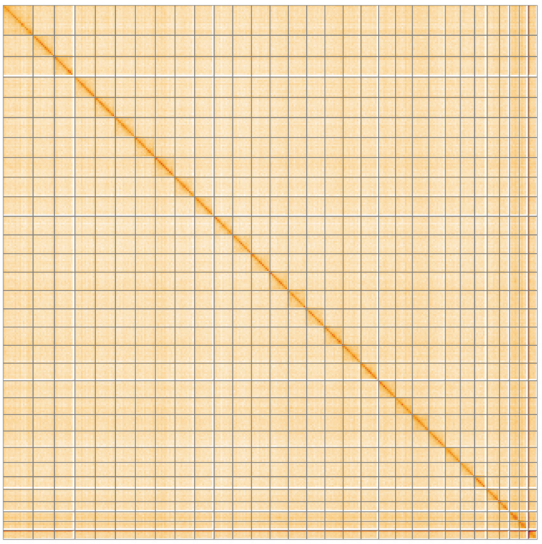

Figure 5. Genome assembly of *Euplexia lucipara*, ilEupLuci1.1: Hi-C contact map of the ilEupLuci1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=e6psxBA4RJqdiQYZcvsoiw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Euplexia lucipara*, ilEupLuci1.

INSDC accession	Chromosome	Length (Mb)	GC%
OV179143.1	1	26.32	36.5
OV179144.1	2	25.29	36.5
OV179145.1	3	25.21	37.0
OV179146.1	4	24.67	37.0
OV179147.1	5	24.66	36.5
OV179148.1	6	24.57	36.5
OV179149.1	7	24.25	37.0
OV179150.1	8	24.24	36.5
OV179151.1	9	23.91	37.0
OV179152.1	10	23.1	36.5
OV179153.1	11	22.99	36.5
OV179154.1	12	22.98	36.5
OV179155.1	13	22.67	37.0
OV179156.1	14	22.58	37.0
OV179157.1	15	22.44	36.5
OV179158.1	16	22.42	36.5
OV179159.1	17	22.13	37.0
OV179160.1	18	21.56	37.0
OV179161.1	19	20.97	37.0
OV179162.1	20	20.56	37.0
OV179163.1	21	20.53	37.0
OV179164.1	22	20.26	36.5
OV179165.1	23	18.4	37.0
OV179166.1	24	17.64	37.0
OV179167.1	25	15.56	37.0
OV179168.1	26	15.09	37.0
OV179169.1	27	12.58	37.5
OV179170.1	28	11.85	37.5
OV179171.1	29	11.15	38.0
OV179172.1	30	11.11	38.0
OV179142.1	Z	37.15	36.5
OV179173.1	MT	0.02	18.5

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 *Euplexia lucipara* assembly (GCA_921972225.1) 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	4.1.7	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	0.15.3	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
SALSA	2.2	https://github.com/salsa-rs/salsa
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

Data availability

European Nucleotide Archive: *Euplexia lucipara* (small angle shades). Accession number PRJEB46851; https://identifiers.org/ena.embl/PRJEB46851 (Wellcome Sanger Institute, 2021). The genome sequence is released openly for reuse. The *Euplexia lucipara* 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

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

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

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; [Accessed 26 July 2023].
Reference Source

GBIF Secretariat: ***Euplexia lucipara* (Linnaeus, 1758).** *GBIF Backbone Taxonomy.* 2023; [Accessed 11 October 2023].
Reference Source

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | Publisher Full Text | Free 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

Henwood P, Sterling P, Lewington R: **Field Guide to the Caterpillars of Great Britain and Ireland.** London: Bloomsbury, 2020.
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

Lawton JH: **Vacant niches and unsaturated communities: a comparison of bracken herbivores at sites on two continents.** *J Anim Ecol.* 1982; **51**(2): 573–595.
Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths.** Newbury: NatureBureau, 2019.

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

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

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].
Reference Source

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 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

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 Small Angle Shades, *Euplexia lucipara* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB46851, 2021.

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 1

Reviewer Report 29 August 2024

https://doi.org/10.21956/wellcomeopenres.22417.r93368

✔ **Vlad Dincă**
University of Oulu, Oulu, Finland

In this study, the authors have successfully generated a chromosome-level genome assembly of a male *Euplexia lucipara* (Lepidoptera, Noctuidae). The mitochondrial genome has been assembled as well.

Although I am not an expert in several of the methodological approaches used, as far as I can tell, the manuscript is technically sound and uses methodologies and pipelines that are fairly well-established through use by the Sanger Institute.

The genome appears to be of high quality and has a BUSCO v5.3.2 completeness of 99.0%.
It is a pity that the W sex chromosome is lacking (a male was sequenced).

Reference genomes such as this one represent a valuable resource for the scientific community and each new addition provides new opportunities for research.

Other comments:

I wonder how much is known in terms of genetic structure for *E. lucipara*. Even if limited data is available (e.g. mtDNA, a few nDNA markers etc), if any notable aspect is known (e.g. diverged lineages), it could be worth briefly mentioning it in the Background.
It may also be worth mentioning whether the species can have more than one generation in other parts of its range.

Species taxonomy, third line: *Euplexia lucipara* should be written in italics

Background, first line: “with a wide distribution” or “widely distributed”

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: Lepidoptera speciation, taxonomy and phylogeography, with a focus on Western Palearctic butterflies.

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 26 August 2024

https://doi.org/10.21956/wellcomeopenres.22417.r83573

Jerome H L Hui
The Chinese University of Hong Kong, Hong Kong, Hong Kong

Boyes, Lewis, and colleagues report the genome sequence of a male moth *Euplexia lucipara* (Linnaeus, 1758). According to the UKmoths (www.ukmoths.org.uk), this species is common and widespread over most of the British Isles. Prior to this report, molecular data of this species was rather limited (mainly mitochondrion COI gene sequences deposited to the NCBI database).

This new genome resource will be useful for further studies such as understanding their population genetic structure, identify cryptic/sub- species, as well as understanding their evolutionary relationships with other moths.

This genome resource is excellent from the summary statistics, with high BUSCO number, high sequence continuity, and majority of sequences contained on the 30 pseudochromosomes (plus Z chromosome and mitochondrion). To sum up, this is another valuable contribution by the Darwin Tree of Life.

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: I have published with Peter Holland more than three years ago, and confirm that this potential conflict of interest did not affect my ability to write an objective and unbiased review of the article.

Reviewer Expertise: Genomics, evolution, invertebrates

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 18 July 2024

https://doi.org/10.21956/wellcomeopenres.22417.r87644

Savarimuthu Ignacimuthu
St. Xavier's College, Thirunelveli, TN, India
Sivasankaran Kuppusamy
Entomology Research Institute, Loyola College, Chennai, Tamil Nadu, India

I appreciate the authors for assembling the genome of Small Angle Shades, *Euplexia lucipara* (Linnaeus, 1758). Authors sequenced approximately 661.6 mega bases using 71 scaffolds. They have used appropriate methods for DNA isolation and used standard software for the sequence assembly and annotations. A total of 20,395 protein-coding genes and 20,590 gene transcripts were observed through the annotations.

Minor comments on the manuscript
The species (*Euplexia lucipara*) is not in italics in the species taxonomy hierarchy. It can be changed.
In the background, the second paragraph can be changed as “This is one of a relatively small set of British and Irish insects that feeds frequently on bracken (*Pteridium aquilinum* (L.) Kuhn) and other ferns (Lawton, 1982). Its larva also feeds on a wide variety of angiosperms (Henwood et al., 2020). This species has an univoltine in the UK, hibernating in the pupal stage.

The authors have given the genus name *Euplexia lucipara* full form throughout the article. For the first time the genus name can be given in full form then subsequently can be given in short form

like *E. lucipara*

The total length of the mitochondrial genome sequence must be provided in the text.

Above all, I confirm that the manuscript meets the necessary scientific standards and is suitable for indexing.

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: Molecular biology

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