DATA NOTE

The genome sequence of the Buff Footman, *Eilema depressum* (Esper, 1787) [version 1; peer review: 2 approved]

Douglas Boyes[1+], Chelsea Skojec[2], Akito Y. Kawahara[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]McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, Gainesville, Florida, USA

[+] Deceased author

v1 **First published:** 11 Aug 2023, **8**:342
https://doi.org/10.12688/wellcomeopenres.19808.1
Latest published: 11 Aug 2023, **8**:342
https://doi.org/10.12688/wellcomeopenres.19808.1

Abstract

We present a genome assembly from an individual male *Eilema depressum* (the Buff Footman; Arthropoda; Insecta; Lepidoptera; Erebidae). The genome sequence is 622.0 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.46 kilobases in length. Gene annotation of this assembly on Ensembl identified 20,038 protein coding genes.

Keywords

Eilema depressum, Buff Footman, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 11 Aug 2023	✓ view	✓ view

1. **Kuppusamy Sivasankaran**, Entomology Research Institute, Loyola College, Chennai, India

2. **Komal Kumar Bollepogu Raja**, Baylor College of Medicine, Houston, USA

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; **Skojec C**: Writing – Original Draft Preparation; **Kawahara AY**: 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, Skojec C, Kawahara AY *et al.* **The genome sequence of the Buff Footman, *Eilema depressum* (Esper, 1787) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:342 https://doi.org/10.12688/wellcomeopenres.19808.1

First published: 11 Aug 2023, **8**:342 https://doi.org/10.12688/wellcomeopenres.19808.1

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; Erebidae; Arctiinae; Lithosiini; *Eilema*; *Eilema depressum* (Esper, 1787) (NCBI:txid987419).

Background

The Buff Footman moth, *Eilema depressum* (Esper 1787) is a drab-coloured moth in the family Erebidae. It has a Eurasiatic distribution that ranges throughout Europe to Japan (Grassi & Zilli, 2005). There are over 160 described species in this genus. *Eilema depressum* is sexually dimorphic: females are larger and darker in colour with a yellow border along the edge of their forewings, while males are smaller and paler in colouration (Kobayashi, 1969).

The Buff Footman belongs to a group of lichen and algal feeding moths, feeding on species such as *Parmelia* and *Pleurococcus*. Because of their unique feeding, *Eilema* species have been used to research community dynamics, and an abundant population of lichen feeding *Eilema* species may indicate environmental health or recovery (Seymour *et al.*, 2020). Some species of *Eilema* have been studied for their ability to feed and sequester lichen compounds like parietin, divaricatic acid, and usnic acid (Hesbacher *et al.*, 1995).

A genome of *E. depressum* is needed to better understand community dynamics and the sequestration of secondary compounds in lichen feeding moths. Here we present a chromosomally complete genome sequence for *E. depressum*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Eilema depressum* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 54-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 109-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected four missing joins or misjoins, reducing the scaffold number by 6.25%.

Figure 1. Photograph of the *Eilema depressum* (ilEilDepe1) specimen used for genome sequencing.

The final assembly has a total length of 622.0 Mb in 30 sequence scaffolds with a scaffold N50 of 22.1 Mb (Table 1). The whole assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 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 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 60.7 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.8% (single = 97.9%, duplicated = 0.9%), 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/987419.

Genome annotation report

The *Eilema depressum* genome assembly (GCA_914767945.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Eilema_depressum_GCA_914767945.1/Info/Index). The resulting annotation includes 20,236 transcribed mRNAs from 20,038 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Eilema depressum* (specimen ID Ox000804, individual ilEilDepe1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-08-01 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 ilEilDepe1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Thorax 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 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

Table 1. Genome data for *Eilema depressum*, ilEilDepe1.1.

Project accession data		
Assembly identifier	ilEilDepe1.1	
Species	*Eilema depressum*	
Specimen	ilEilDepe1	
NCBI taxonomy ID	987419	
BioProject	PRJEB46309	
BioSample ID	SAMEA7746611	
Isolate information	ilEilDepe1, male: thorax (DNA sequencing), head (Hi-C scaffolding), abdomen (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	60.9	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.8%[S:97.9%,D:0.9%], F:0.3%,M:0.9%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	100%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6939233, ERR6807996	
10X Genomics Illumina	ERR6688466, ERR6688468, ERR6688469, ERR6688467	
Hi-C Illumina	ERR6688465	
PolyA RNA-Seq Illumina	ERR9434997	
Genome assembly		
Assembly accession	GCA_914767945.1	
Accession of alternate haplotype	GCA_914767765.1	
Span (Mb)	622.0	
Number of contigs	38	
Contig N50 length (Mb)	21.7	
Number of scaffolds	30	
Scaffold N50 length (Mb)	22.1	
Longest scaffold (Mb)	46.0	
Genome annotation		
Number of protein-coding genes	20,038	
Number of gene transcripts	20,236	

* 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/Eilema depressum/dataset/ilEilDepe1_1.1/busco.

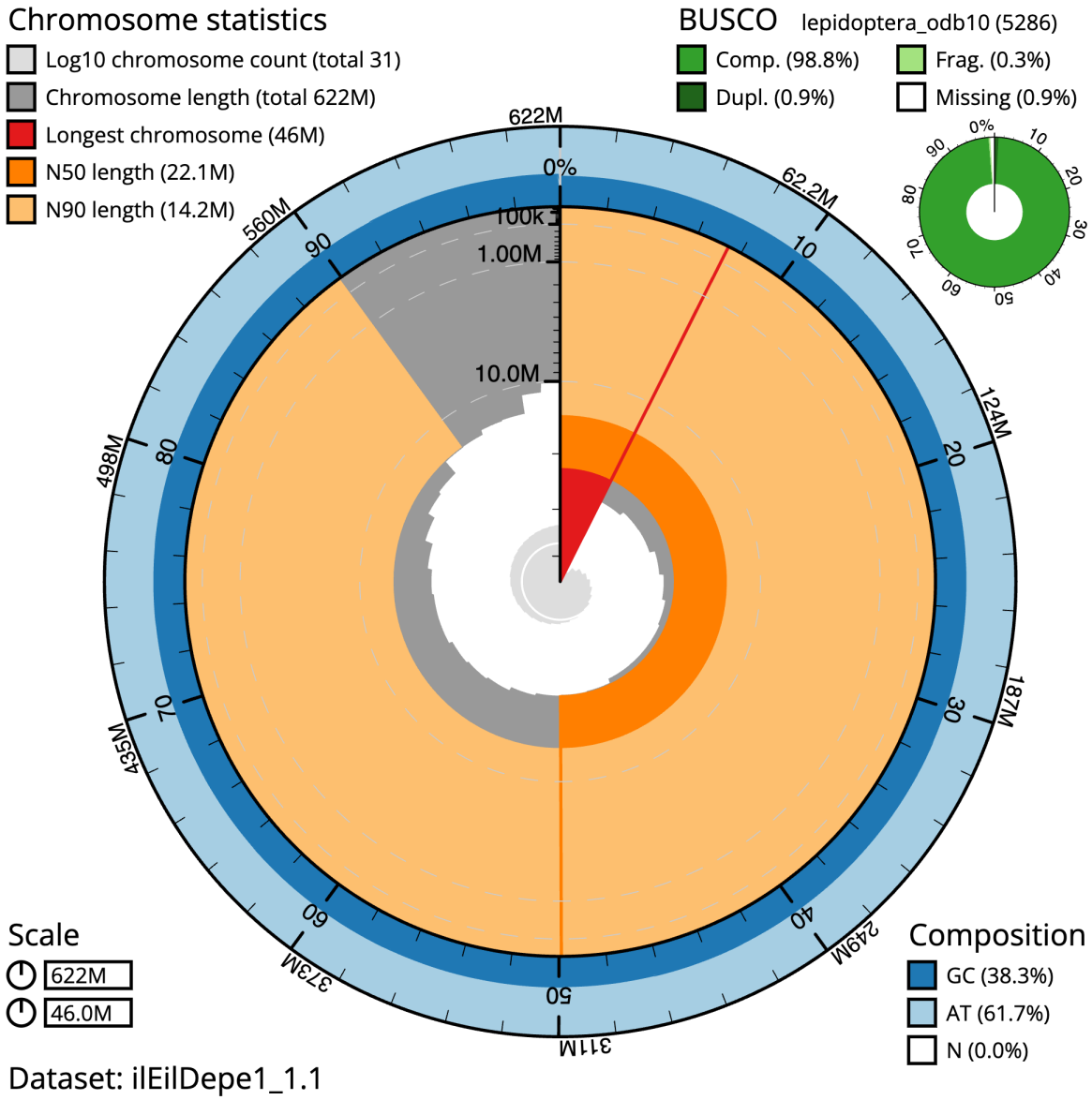

Figure 2. Genome assembly of *Eilema depressum*, ilEilDepe1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 622,022,279 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (46,029,750 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (22,085,244 and 14,221,244 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/Eilema depressum/dataset/ilEilDepe1_1.1/snail.

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.

RNA was extracted from abdomen tissue of ilEilDepe1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using

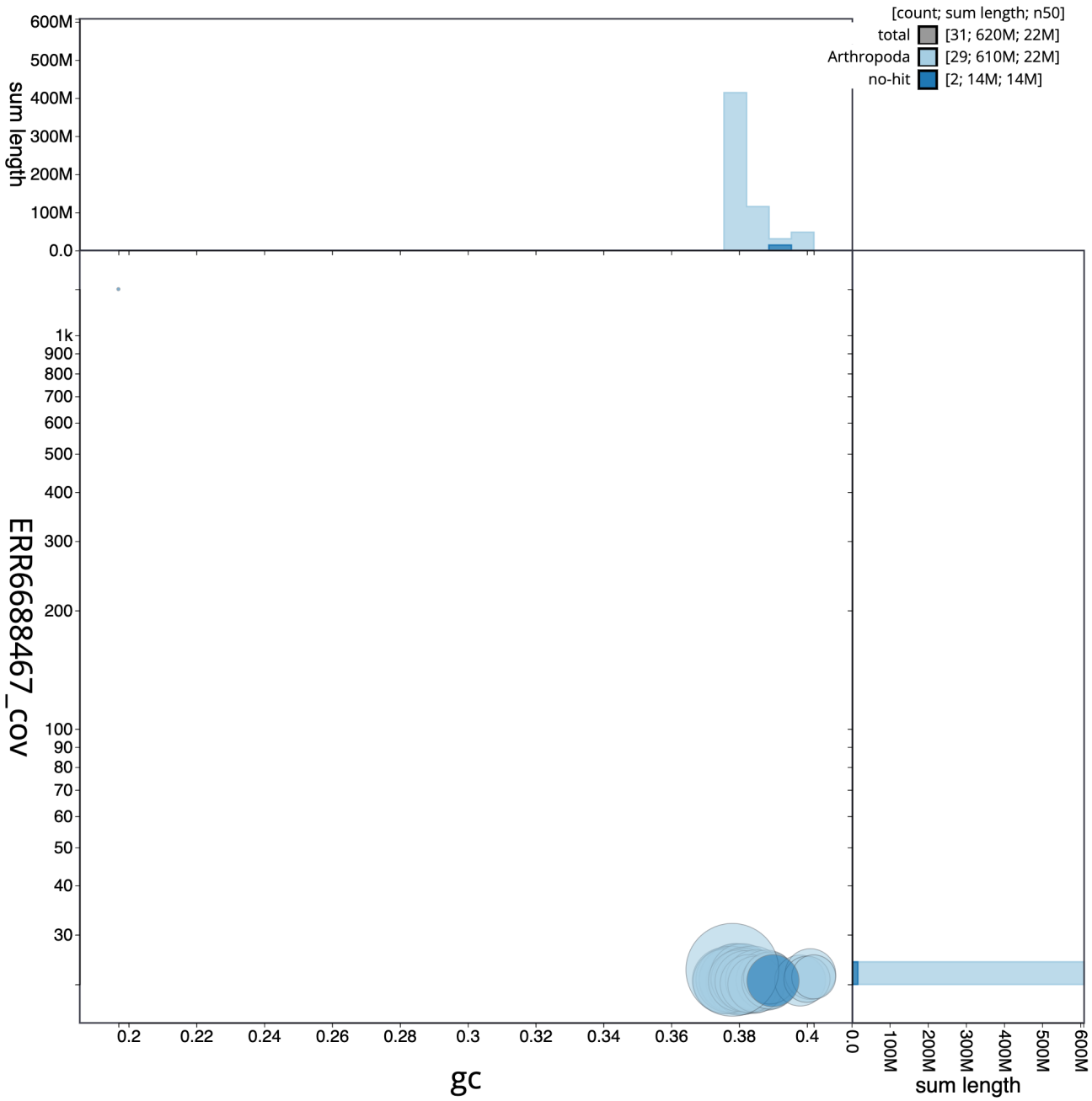

Figure 3. Genome assembly of *Eilema depressum*, ilEilDepe1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Eilema depressum/dataset/ilEilDepe1_1.1/blob.

the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and NovaSeq 6000 (10X) instruments. Hi-C data were also generated from head tissue of ilEilDepe1 using the

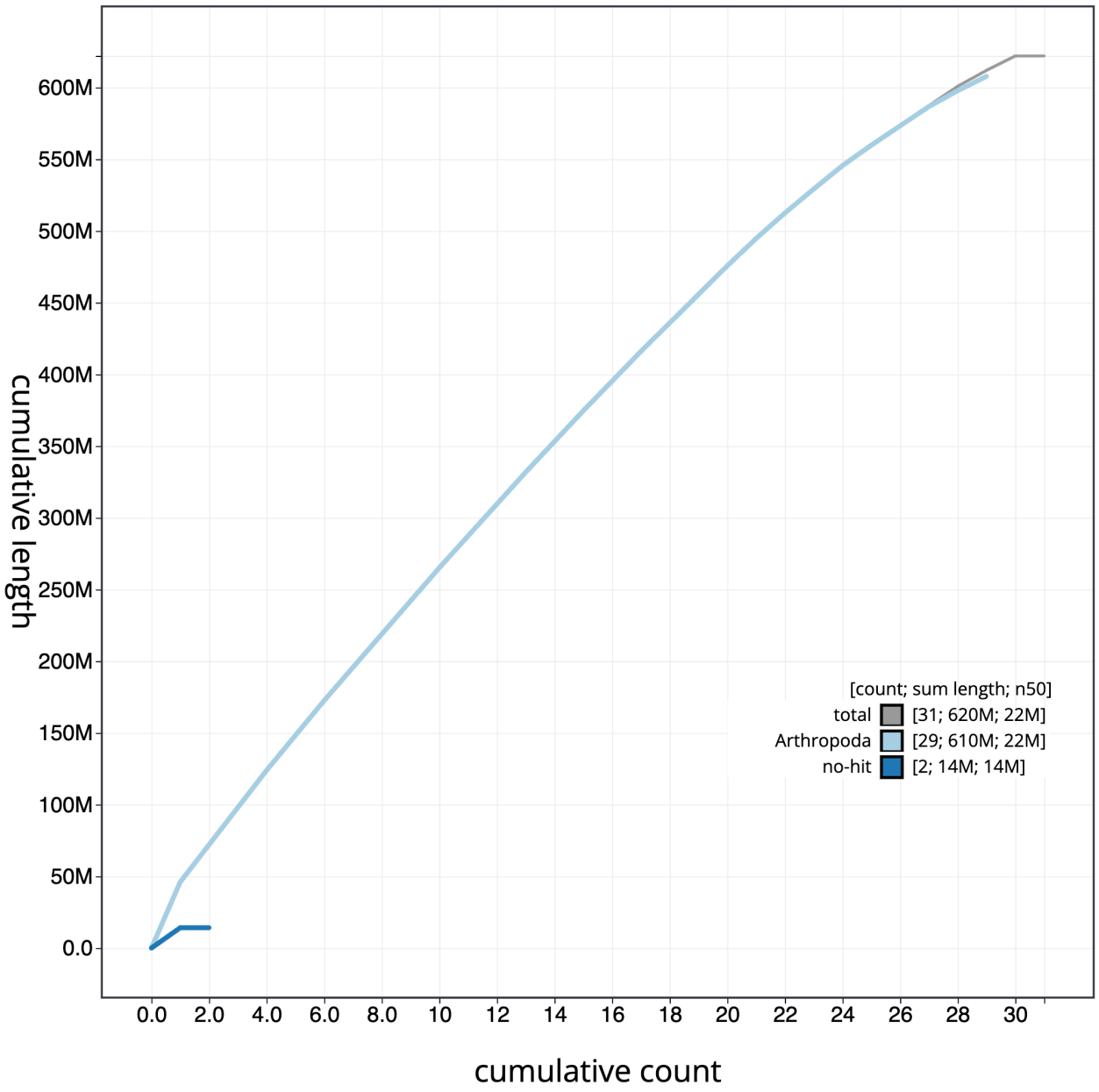

Figure 4. Genome assembly of *Eilema depressum*, ilEilDepe1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Eilema depressum/dataset/ilEilDepe1_1.1/cumulative.

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 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),

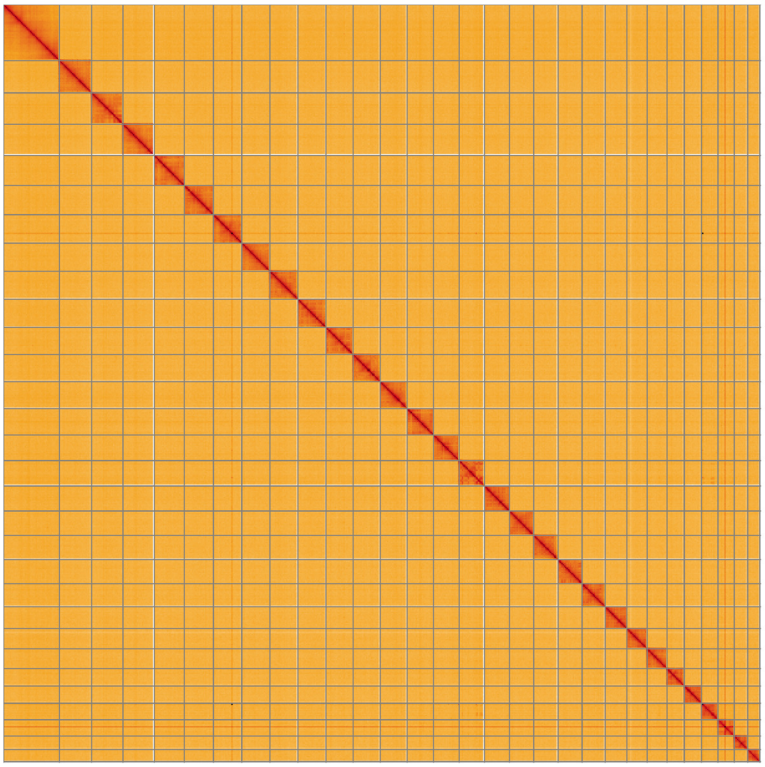

Figure 5. Genome assembly of *Eilema depressum*, ilEilDepe1.1: Hi-C contact map of the ilEilDepe1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=dMhsDho0SFypSA4I_HI3eQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eilema depressum*, ilEilDepe1.

INSDC accession	Chromosome	Length (Mb)	GC%
OU612013.1	1	26.5	37.9
OU612014.1	2	25.73	38
OU612015.1	3	25.67	38.1
OU612016.1	4	24.7	37.9
OU612017.1	5	23.96	37.8
OU612018.1	6	23.37	37.6
OU612019.1	7	23.1	38
OU612020.1	8	23.09	37.8
OU612021.1	9	23.06	37.7
OU612022.1	10	22.33	38.1
OU612023.1	11	22.23	37.6
OU612024.1	12	22.09	38.1
OU612025.1	13	21.72	38.2
OU612026.1	14	21.11	38.4
OU612027.1	15	20.81	38.4
OU612028.1	16	20.54	38.1
OU612029.1	17	20	38
OU612030.1	18	19.99	38.2
OU612031.1	19	19.7	38.5
OU612032.1	20	19.03	38.8
OU612033.1	21	17.79	38.3
OU612034.1	22	16.7	38.5
OU612035.1	23	16.34	38.9
OU612036.1	24	14.22	38.9
OU612037.1	25	14.15	39
OU612038.1	26	13.53	39.8
OU612039.1	27	13.41	40.1
OU612040.1	28	11.06	40
OU612041.1	29	10.03	40.2
OU612012.1	Z	46.03	37.8
OU612042.1	MT	0.02	19.8

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.

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Eilema depressum* assembly (GCA_914767945.1) 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	4.1.5	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	0.15.3	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: *Eilema depressum* (buff footman). Accession number PRJEB46309; https://identifiers.org/ena.embl/PRJEB46309. (Wellcome Sanger Institute, 2021) The genome sequence is released openly for reuse. The *Eilema depressum* 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].
Publisher Full Text

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

Grassi A, Zilli A: **New data on the distribution and ecology of some Italian species of *Eilema* and reappraisal of *Eilema marcida* new rank (Insecta, Lepidoptera: Arctiidae).** *Aldrovandia.* 2005; **1**: 5–15.
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

Hesbacher S, Giez I, Embacher G, *et al.*: **Sequestration of lichen compounds by lichen-feeding members of the Arctiidae (Lepidoptera).** *J Chem Ecol.* 1995; **21**(12): 2079–2089.
PubMed Abstract | 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

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

Kobayashi H: **On the genus *Eilema* from Japan (Lepidoptera, Arctiidae).** *Lepidoptera Science.* 1969; **20**(1–2): 41–56.
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

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

Seymour M, Brown N, Carvalho GR, *et al.*: **Ecological community dynamics: 20 years of moth sampling reveals the importance of generalists for community stability.** *Basic Appl Ecol.* 2020; **49**: 34–44.
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

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

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

Wellcome Sanger Institute: **The genome sequence of the Buff Footman, *Eilema depressum* (Esper, 1787).** European Nucleotide Archive, [dataset], accession number PRJEB46309, 2021.

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 13 August 2024

https://doi.org/10.21956/wellcomeopenres.21940.r93175

✔ **Komal Kumar Bollepogu Raja**
Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, USA

The article by Boyes et al., presents genome sequence for *Eilema depressum*. The authors provide comprehensive and detailed data on the assembly, annotation, methods and quality of the genome for this moth species. The final assembly has a high Quality Value (QV) of 60.7, with 100% k-mer completeness and a BUSCO completeness score of 98.8%. These metrics indicate a well-assembled genome with a low level of fragmentation. Furthermore, the scaffolds are well-documented with a total length of 622.0 Mb and 30 chromosome-scale scaffolds. This detailed work supports further research into community dynamics and secondary compound sequestration in lichen-feeding moths. Overall, the data is of high quality and the manuscript is well written.

Minor comment:

The authors mention that contigs of second haplotype were deposited separately. It would benefit if the authors add text noting differences between the haplotypes and if they affect analyses.

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: Genomics, Bioinformatics, Computational biology, single cell omics

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 06 August 2024

https://doi.org/10.21956/wellcomeopenres.21940.r88932

Kuppusamy Sivasankaran
Entomology Research Institute, Loyola College, Chennai, India

Authors have sequenced and assembled the whole genome of Buff Footman moth *Eilema depressum* (Esper, 1787). A total length of 622 mb sequence was assembled. A total of 30 chromosomes were confirmed through the assembly. Totally 20, 038 protein coding genes and 20, 236 gene transcripts were observed through the assembly and annotation.

Minor comments on the manuscripts:

- Authors have given the genus name of the species *Eilema depressum* full form throughout the article. And given a short form somewhere. First time the genus name can be given in full form then subsequently can be given in short form like *E. depressum*.
- In the background the first sentence can be modified as “The Buff Footman moth, *Eilema depressum* (Esper, 1787) is a drab-coloured moth belonging to the family *Erebidae*”.
- In the genome sequence report (subheading), the first line can be modified as “The genome was sequenced from a male *Eilema depressum* specimen (Figure 1)”. The collection details need not be given here. Because the collection details were given in the methods.
- The table 3 legend was given inside text. The legend can be changed as a text “A list of relevant software tool versions and sources are given in the Table 3”.
- Above all, the manuscript was well prepared, and I confirm that the manuscript meets the necessary scientific standard 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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: phylogenetic analysis of moths using mitogenome

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.