DATA NOTE

The genome sequence of the Grey Poplar Bell, *Epinotia nisella* (Clerck, 1759) [version 1; peer review: 3 approved]

Douglas Boyes[1+], James Hammond[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:** 28 Jun 2023, **8**:281
https://doi.org/10.12688/wellcomeopenres.19621.1
Latest published: 28 Jun 2023, **8**:281
https://doi.org/10.12688/wellcomeopenres.19621.1

Abstract

We present a genome assembly from an individual male *Epinotia nisella* (the Grey Poplar Bell; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence is 585.0 megabases in span. Most of the assembly is scaffolded into 28 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.44 kilobases in length. Gene annotation of this assembly on Ensembl identified 18,952 protein coding genes.

Keywords

Epinotia nisella, Grey Poplar Bell, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 28 Jun 2023	✓ view	✓ view	✓ view

1. **Ann-Marie Waldvogel**, University of Cologne, Cologne, Germany
2. **Sivasankaran Kuppusamy**, Entomology Research Institute, Loyola College, Chennai, India
3. **Jerome H L Hui**, Chinese University of Hong Kong, Hong Kong, China

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; **Hammond J**: 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, Hammond J, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Grey Poplar Bell, *Epinotia nisella* (Clerck, 1759) [version 1; peer review: 3 approved]** Wellcome Open Research 2023, **8**:281 https://doi.org/10.12688/wellcomeopenres.19621.1

First published: 28 Jun 2023, **8**:281 https://doi.org/10.12688/wellcomeopenres.19621.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Apoditrysia; Tortricoidea; Tortricidae; Olethreutinae; Eucosmini; *Epinotia*; *Epinotia nisella* (Clerck, 1759) (NCBI:txid989769).

Background

The Grey Poplar Bell *Epinotia nisella* (Clerck, 1759) is a moth in the Tortricidae family. The species is highly polymorphic and shows extensive variation in the grey, black, and brown colouration of the forewing (Bradley *et al.*, 1979). The species has a Holarctic distribution and is found across northern Eurasia and northern North America (GBIF Secretariat, 2022). The moth is found across the British Isles in woodland, parks, and damp places where the foodplants, Salix and Populus, occur (Bradley *et al.*, 1979).

The species overwinters as an egg or small larva (Elliott *et al.*, 2018). The larva feeds between April and June in catkins or between spun leaves, and pupation occurs in the larval habitation or leaf-litter, and adults occur from July to September (Bradley *et al.*, 1979; Elliott *et al.*, 2018). The adults fly from dusk and come to light, and rest by day on foliage (Bradley *et al.*, 1979).

A genome of *Epinotia nisella* will contribute to our understand of the genomic basis of polymorphism, and its evolution within lepidoptera. The genome of *Epinotia nisella* 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 *Epinotia nisella*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Epinotia nisella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 39-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 9 missing joins or mis-joins and removed 5 haplotypic duplications, reducing the assembly length by 1.49% and the scaffold number by 5.13%.

The final assembly has a total length of 585.0 Mb in 37 sequence scaffolds with a scaffold N50 of 21.3 Mb (Table 1). Most (99.92%) of the assembly sequence was assigned to 28 chromosomal-level scaffolds, representing 27 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.

Figure 1. Photograph of the *Epinotia nisella* (ilEpiNise2) specimen used for genome sequencing.

The estimated Quality Value (QV) of the final assembly is 68.1 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.2% (single = 97.6%, 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/989769.

Genome annotation report

The *Epinotia nisella* genome assembly (GCA_932294315.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Epinotia_nisella_GCA_932294315.1/Info/Index). The resulting annotation includes 19,249 transcribed mRNAs from 18,952 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Two *Epinotia nisella* specimens (ilEpiNise2 and ilEpiNise1) were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-09-08. The specimens were collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice. The specimen used for genome sequencing was ilEpiNise2 (specimen ID Ox000953), while ilEpiNise1 (specimen ID Ox000952) was used for Hi-C scaffolding.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilEpiNise2 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

Table 1. **Genome data for *Epinotia nisella*, ilEpiNise2.1.**

Project accession data		
Assembly identifier	ilEpiNise2.1	
Species	*Epinotia nisella*	
Specimen	ilEpiNise2	
NCBI taxonomy ID	989769	
BioProject	PRJEB50738	
BioSample ID	SAMEA8603184	
Isolate information	ilEpiNise2, male: whole organism (DNA sequencing) ilEpiNise1: whole organism (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	68.1	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.2%[S:97.6%,D:0.6%], F:0.6%,M:1.2%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.92%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR8575373, ERR8575374	
Hi-C Illumina	ERR8571655	
Genome assembly		
Assembly accession	GCA_932294315.1	
Accession of alternate haplotype	GCA_932294385.1	
Span (Mb)	585.0	
Number of contigs	46	
Contig N50 length (Mb)	21.2	
Number of scaffolds	37	
Scaffold N50 length (Mb)	21.3	
Longest scaffold (Mb)	45.6	
Genome annotation		
Number of protein-coding genes	18,952	
Number of gene transcripts	19,249	

* 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/ilEpiNise2.1/dataset/CAKOAI01.1/busco.

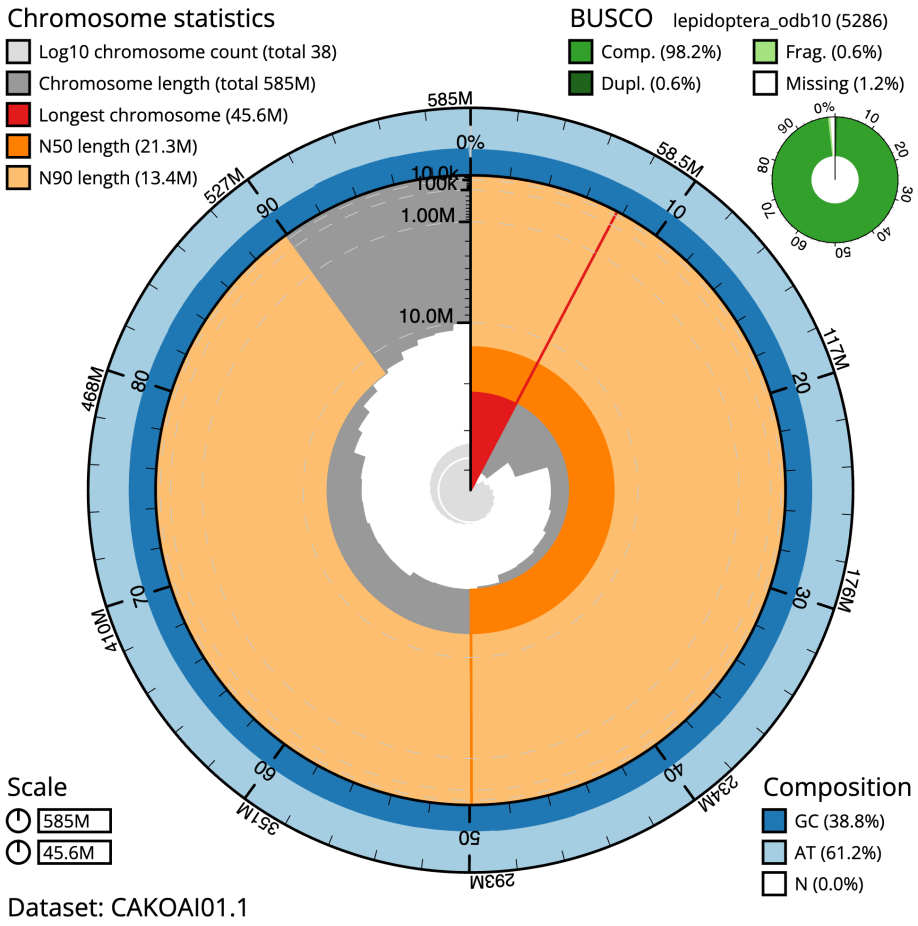

Figure 2. Genome assembly of *Epinotia nisella*, ilEpiNise2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 585,001,234 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (45,621,359 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (21,276,924 and 13,438,000 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/ilEpiNise2.1/dataset/CAKOAI01.1/snail.

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 from whole organism tissue of ilEpiNise1 using the Arimav2 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

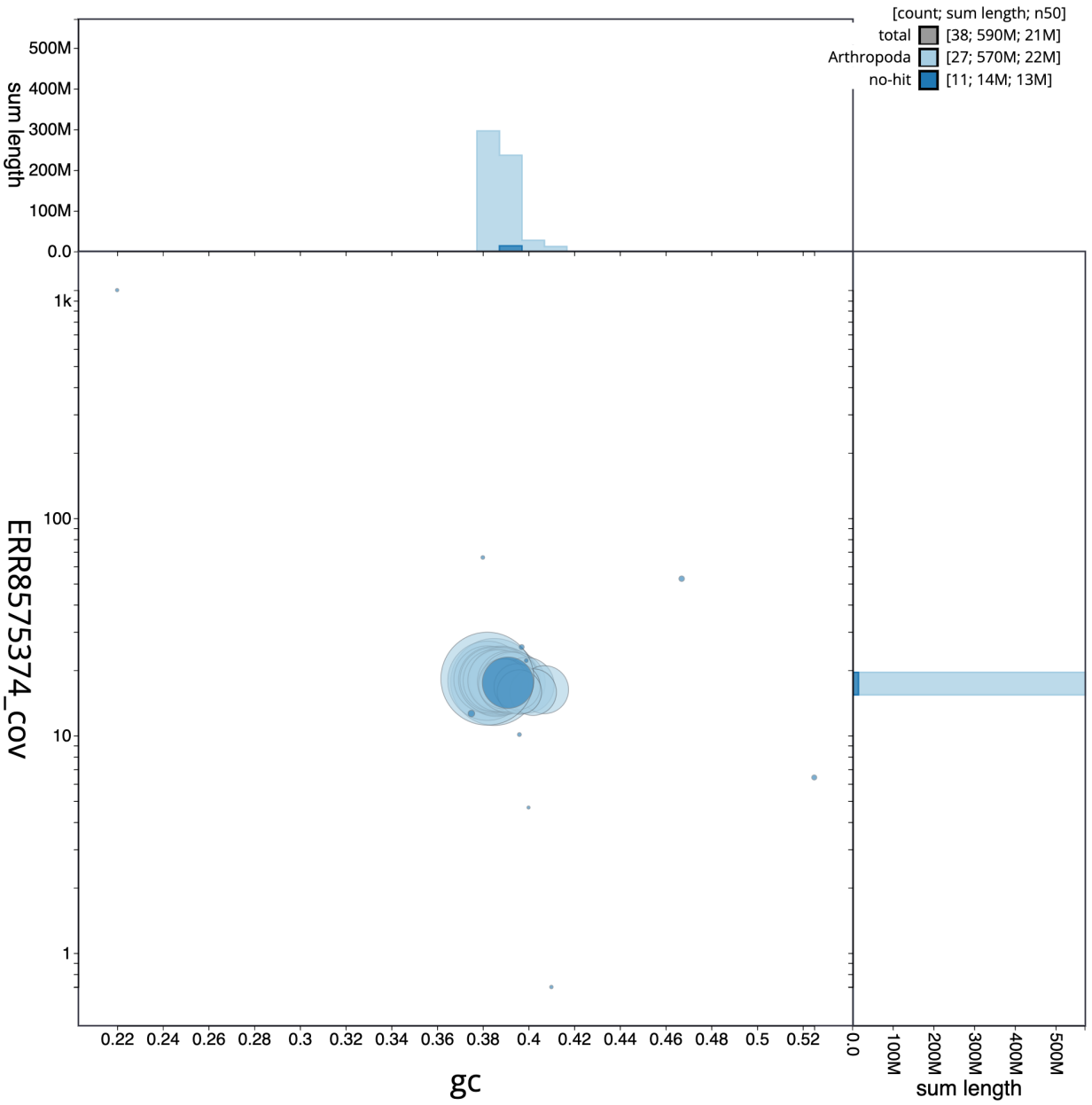

Figure 3. Genome assembly of *Epinotia nisella*, ilEpiNise2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEpiNise2.1/dataset/CAKOAI01.1/blob.

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

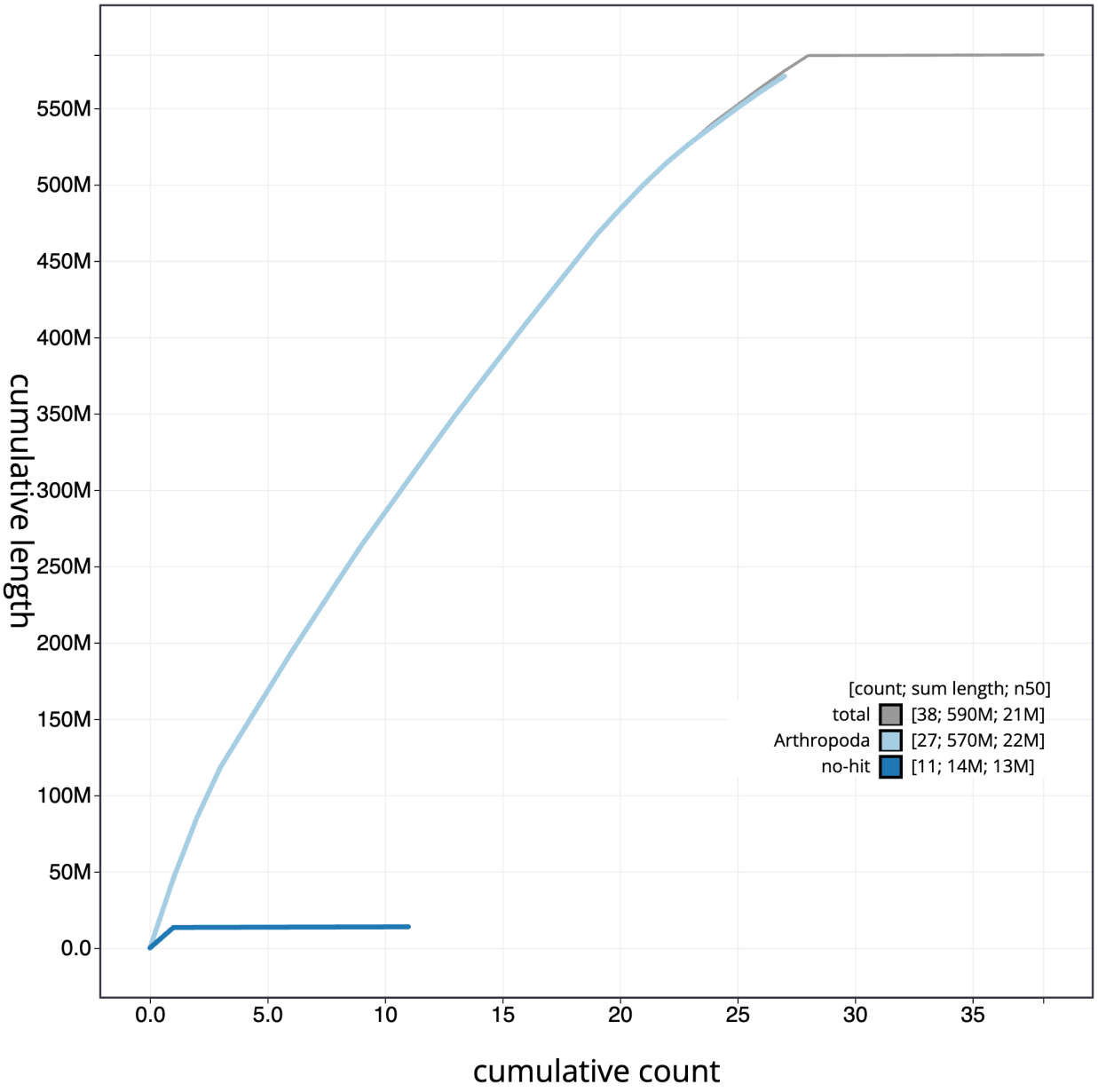

Figure 4. Genome assembly of *Epinotia nisella*, ilEpiNise2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEpiNise2.1/dataset/CAKOAI01.1/cumulative.

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 *Epinotia nisella* assembly (GCA_932294315.1) 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

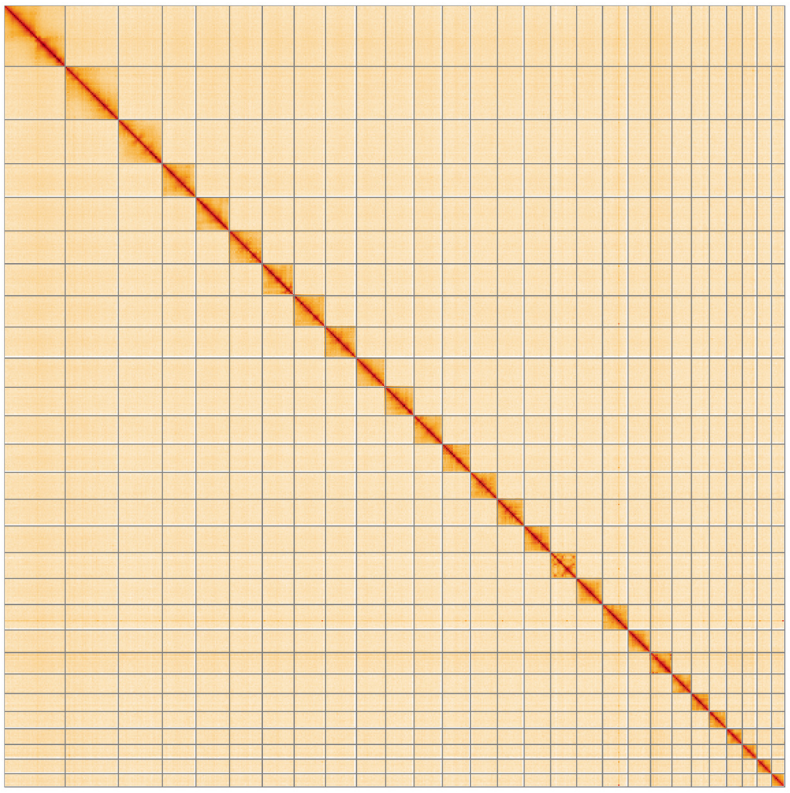

Figure 5. Genome assembly of *Epinotia nisella*, ilEpiNise2.1: Hi-C contact map of the ilEpiNise2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=PFvoduDaQYKqY9WV75PHww.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Epinotia nisella*, ilEpiNise2.

INSDC accession	Chromosome	Length (Mb)	GC%
OW028674.1	1	39.8	38.5
OW028675.1	2	32.97	38.0
OW028676.1	3	25.15	38.5
OW028677.1	4	25.12	39.0
OW028678.1	5	24.57	38.5
OW028679.1	6	23.85	39.0
OW028680.1	7	23.45	38.0
OW028681.1	8	23.27	38.5
OW028682.1	9	21.78	39.0
OW028683.1	10	21.28	38.5
OW028684.1	11	21.24	39.0
OW028685.1	12	21.1	39.0
OW028686.1	13	20.08	38.5
OW028687.1	14	20.02	38.5
OW028688.1	15	19.87	38.5
OW028689.1	16	19.48	39.0
OW028690.1	17	19.35	39.0
OW028691.1	18	19.12	39.5
OW028692.1	19	16.88	39.0
OW028693.1	20	16.0	40.0
OW028694.1	21	14.6	39.0
OW028695.1	22	13.44	39.0
OW028696.1	23	12.88	39.0
OW028697.1	24	11.77	40.5
OW028698.1	25	11.05	40.0
OW028699.1	26	10.85	39.5
OW028700.1	27	9.99	39.5
OW028673.1	Z	45.62	38.0
OW028701.1	MT	0.02	22.0

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.0.7	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	yahs-1.1.91eebc2	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: *Epinotia nisella* (grey poplar bell). Accession number PRJEB50738; https://identifiers.org/ena.embl/PRJEB50738. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Epinotia nisella* 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): 313–9.
PubMed Abstract | Publisher Full Text

Bradley JD, Tremewan WG, Smith A: **British Tortricoid Moths - Tortricidae: Olethreutinae.** The Ray Society, 1979.

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

Elliott B, Langmaid JR, Palmer RM, *et al.*: **Tortricidae.** In: Langmaid, J. R., Palmer, S. M., and Young, M. R. (eds.) *A Field Guide to the Smaller Moths of Great Britain and Ireland.* The British Entomological and Natural History Society: 2018; 279.
Reference Source

GBIF Secretariat: ***Epinotia nisella* (Clerck, 1759).** *GBIF Backbone Taxonomy.* 2022; [Accessed 3 May 2023].
Reference Source

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

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 Grey Poplar Bell, *Epinotia nisella* (Clerck, 1759).** European Nucleotide Archive. [dataset], accession number PRJEB50738, 2022.

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 13 May 2024

https://doi.org/10.21956/wellcomeopenres.21737.r71154

✔ **Jerome H L Hui**
Chinese University of Hong Kong, Hong Kong, China

Boyes and colleagues report the genome sequence of a male grey poplar bell *Epinotia nisella* (Clerck 1759). This species is widely distributed in Britain. Molecular data of this species are scarce prior to this report, and are all COI sequences deposited to the NCBI database. This new genome resource will be very useful for further studies, such as understanding their geographical distribution, population structure, evolutionary relationships with other lepidopterans, and the underlying mechanism contributing to their (morphological) polymorphism.

This genome resource is excellent from the summary statistics, with high BUSCO numbers and sequence continuity (scaffold N50), and majority of sequences contained on the 28 pseudochromosomes (plus mitochondrion). To sum up, this is a valuable contribution.

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 Prof. Peter Holland more than 3 years ago. I 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 11 May 2024

https://doi.org/10.21956/wellcomeopenres.21737.r77044

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

The authors have done the whole genome sequence of *Epinotia nisella* (Clerck, 1759). The authors sequenced over 585.0 million base pairs for the genome assembly using 37 scaffolds. The authors observed 19, 249 gene transcripts and 18,952 protein-coding genes during genome assembly. The standard methods were used for DNA isolation and followed appropriate software for the sequence assembly and annotations.

Overall, there are corrections on this manuscript. The manuscript is accepted scientific standards.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Phylogenetic analysis of moths

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 04 March 2024

https://doi.org/10.21956/wellcomeopenres.21737.r74980

Ann-Marie Waldvogel
University of Cologne, Cologne, Germany

This report presents a high-quality genome assembly of the lepidoptera genome of *Epinotia nisella* , a species that is of scientific interest for its variation in wing pattern coloration. With this, the rational for the generation of high-quality genomic resources is obvious and the genome assembly with furthermore contribute to comparative studies.
The assembly has been generated and curated according to standard state-of-the art methodology and fulfills all expected criteria by its summary statistics.

Here are some minor comments:
*Lepidopterans are known to be very beautiful, and I highly recommend adding a nicer picture that documents the coloration more clearly.

*Last paragraph of the description of background:
"A genome ... will contribute to our **understanding** of the genomic basis of **morphotype** polymorphism."

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: ecological genomics, biodiversity genomics

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.