DATA NOTE

The genome sequence of the Great Brocade moth, *Eurois occulta* Linnaeus, 1758 [version 1; peer review: awaiting peer review]

Marc Botham[1],
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK

v1 **First published:** 26 Sep 2024, **9**:558
https://doi.org/10.12688/wellcomeopenres.23082.1
Latest published: 26 Sep 2024, **9**:558
https://doi.org/10.12688/wellcomeopenres.23082.1

Abstract

We present a genome assembly from an individual male *Eurois occulta* (the Great Brocade moth; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence has a total length of 603.30 megabases. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.39 kilobases in length. Gene annotation of this assembly on Ensembl identified 19,015 protein-coding genes.

Keywords

Eurois occulta, Great Brocade moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Botham M: Investigation, Resources;

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.

How to cite this article: Botham M, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations *et al.* **The genome sequence of the Great Brocade moth, *Eurois occulta* Linnaeus, 1758 [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:558 https://doi.org/10.12688/wellcomeopenres.23082.1

First published: 26 Sep 2024, **9**:558 https://doi.org/10.12688/wellcomeopenres.23082.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Noctuidae; Noctuinae; Noctuini; *Eurois*; *Eurois occulta* (Linnaeus, 1758) (NCBI:txid987425).

Background

Eurois occulta, commonly referred to as the Great Brocade, is widely distributed across the northern parts of the Palaearctic and Nearctic regions, including North America (GBIF Secretariat, 2024). The species typically inhabits temperate climates and is known to thrive in a variety of habitats including forests, meadows, and grasslands. In the UK, the Great Brocade has a somewhat restricted distribution (nationally scarce B), primarily recorded in Scotland and northern England (Waring *et al.*, 2017). It is recorded as an immigrant in southern parts of England (NBN Atlas, 2024). In Ireland, records are sparse, indicating it is not commonly found or may be under-recorded. The larva feeds on bog myrtle (*Myrica gale*), but other plants and bushes, such as sallow (*Salix*) and birch (*Betula*) are also eaten (UK Moths, 2024).

The Great Brocade is a medium-sized moth with a wingspan ranging between 45 and 55 millimetres. Its forewings are generally dark brown or grey with intricate patterns, including lighter, wavy lines and distinct, dark spots. The hindwings are typically a lighter grey, often with a dark discal spot and a paler fringe. Immigrant moths from Europe are typically paler and grey (Waring *et al.*, 2017).

The genome of the great brocade, *Eurois occulta*, 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 (Blaxter *et al.*, 2022). Here we present a chromosomally complete genome sequence for *Eurois occulta*, based on one male specimen from Glen Strathfarrar, Scottish Highlands, UK.

Genome sequence report

The genome of an adult male *Eurois occulta* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 23.62 Gb (gigabases) from 2.21 million reads, providing approximately 37-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 132.61 Gb from 878.23 million reads, yielding an approximate coverage of 220-fold. Specimen and sequencing information is summarised in Table 1.

Manual assembly curation corrected 18 missing joins or mis-joins and 10 haplotypic duplications, reducing the assembly length by 2.81% and the scaffold number by 16.28%. The final assembly has a total length of 603.30 Mb in 35 sequence scaffolds with a scaffold N50 of 20.5 Mb (Table 2). The total count of gaps in the scaffolds is 58. The snail plot

Figure 1. Photograph of the *Eurois occulta* (not the specimen used for genome sequencing).

in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.99%) 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 5; Table 3). The Z chromosome was identified based on synteny with *Xestia c-nigrum* (GCA_916618015.1) (Broad *et al.*, 2022). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 68.7 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.5% (single = 98.0%, duplicated = 0.5%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/987425.

Genome annotation report

The *Eurois occulta* genome assembly (GCA_950022335.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 19,196 transcribed mRNAs from 19,015 protein-coding genes (Table 2; https://rapid.ensembl.org/Eurois_occulta_GCA_950022335.1/Info/Index). The average transcript length is 6,503.83. There are 1.01 coding transcripts per gene and 5.25 exons per transcript.

Methods

Sample acquisition

An adult male *Eurois occulta* (specimen ID SAN00002613, ToLID ilEurOccu1) was collected from Glen Strathfarrar,

Table 1. Specimen and sequencing data for *Eurois occulta*.

Project information			
Study title	Eurois occulta (great brocade)		
Umbrella BioProject	PRJEB60716		
Species	*Eurois occulta*		
BioSample	SAMEA112198534		
NCBI taxonomy ID	987425		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	ilEurOccu1	SAMEA112198572	thorax
Hi-C sequencing	ilEurOccu1	SAMEA112198571	head
RNA sequencing	ilEurOccu1	SAMEA112198572	thorax
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR11040193	8.78e+08	132.61
PacBio Sequel IIe	ERR11029702	2.21e+06	23.62
RNA Illumina NovaSeq 6000	ERR12245547	9.70e+07	14.65

Scotland, UK (latitude 57.41, longitude –4.73) on 2022-06-27, using a moth trap. The specimen was collected and identified by Marc Botham (UK Centre for Ecology & Hydrology) and preserved by flash freezing.

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The ilEurOccu1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the thorax was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from thorax tissue of ilEurOccu1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus 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 IIe (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments.

Hi-C data were generated from the head tissue of ilEurOccu1, using the Arima-HiC v2 kit. In brief, frozen tissue (–80 °C) was fixed, and the DNA crosslinked using a TC buffer containing formaldehyde. The crosslinked DNA was then digested using

Table 2. Genome assembly data for *Eurois occulta*, ilEurOccu1.1.

Genome assembly		
Assembly name	ilEurOccu1.1	
Assembly accession	GCA_950022335.1	
Accession of alternate haplotype	*GCA_950023035.1*	
Span (Mb)	603.30	
Number of contigs	94	
Contig N50 length (Mb)	11.6	
Number of scaffolds	35	
Scaffold N50 length (Mb)	20.5	
Longest scaffold (Mb)	30.6	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	68.7	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:98.5%[S:98.0%,D:0.5%], F:0.3%,M:1.2%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.99%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.39 kb	*complete single alleles*
Genome annotation of assembly GCA_950022335.1 at Ensembl		
Number of protein-coding genes	19,015	
Number of gene transcripts	19,196	

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

** BUSCO scores based on the lepidoptera_odb10 BUSCO set using version 5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/ilEurOccu1_1/dataset/ilEurOccu1_1/busco.

a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with a biotinylated nucleotide and proximally ligated. The biotinylated DNA construct was fragmented to a fragment size of 400 to 600 bp using a Covaris E220 sonicator. The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit, following manufacturers' instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline (article in

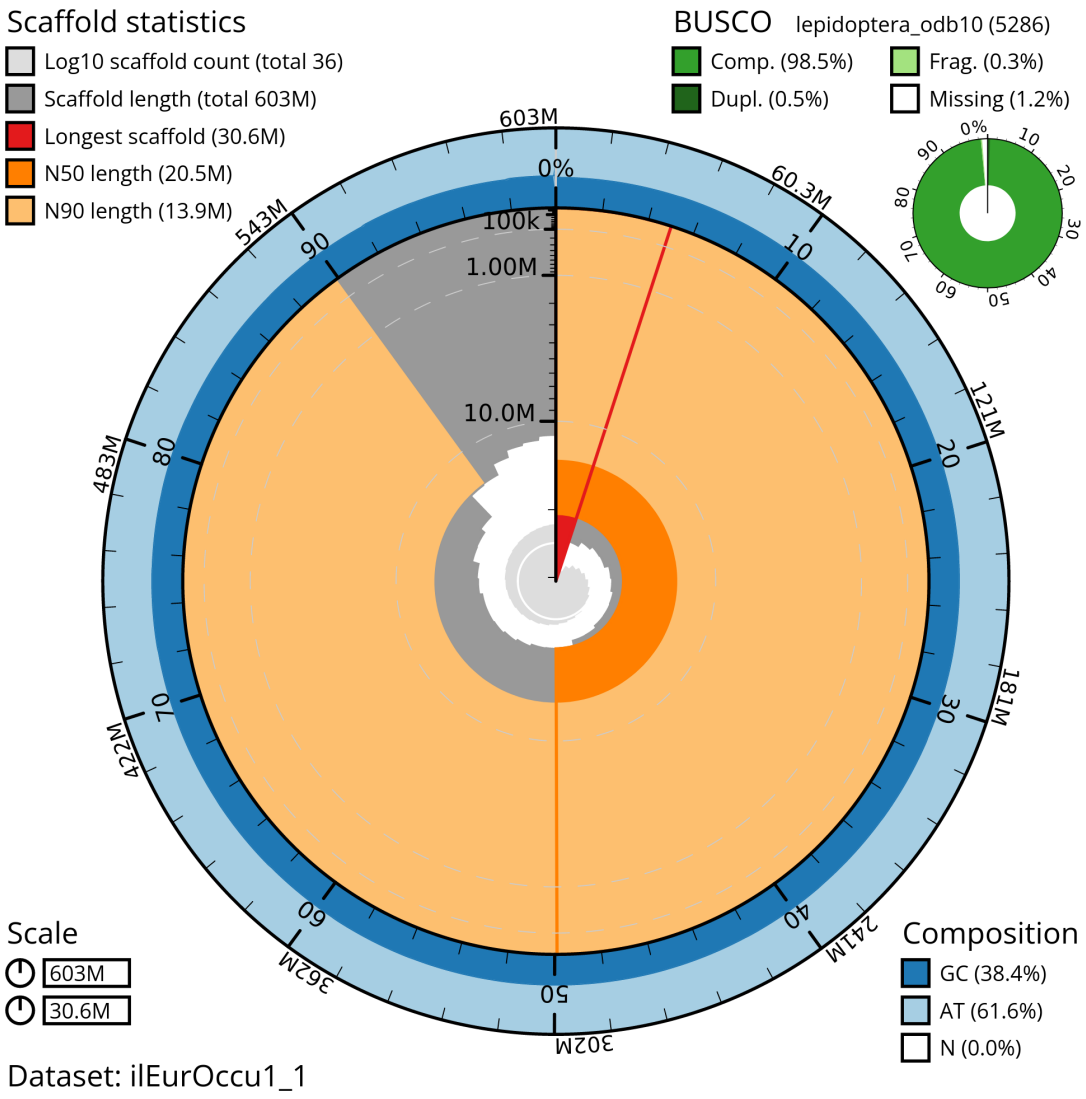

Figure 2. Genome assembly of *Eurois occulta*, ilEurOccu1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 603,364,441 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (30,601,589 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (20,500,800 and 13,928,380 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/ilEurOccu1_1/dataset/ilEurOccu1_1/snail.

preparation). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The sex chromosome was identified by synteny analysis. The entire process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

Evaluation of the final assembly

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

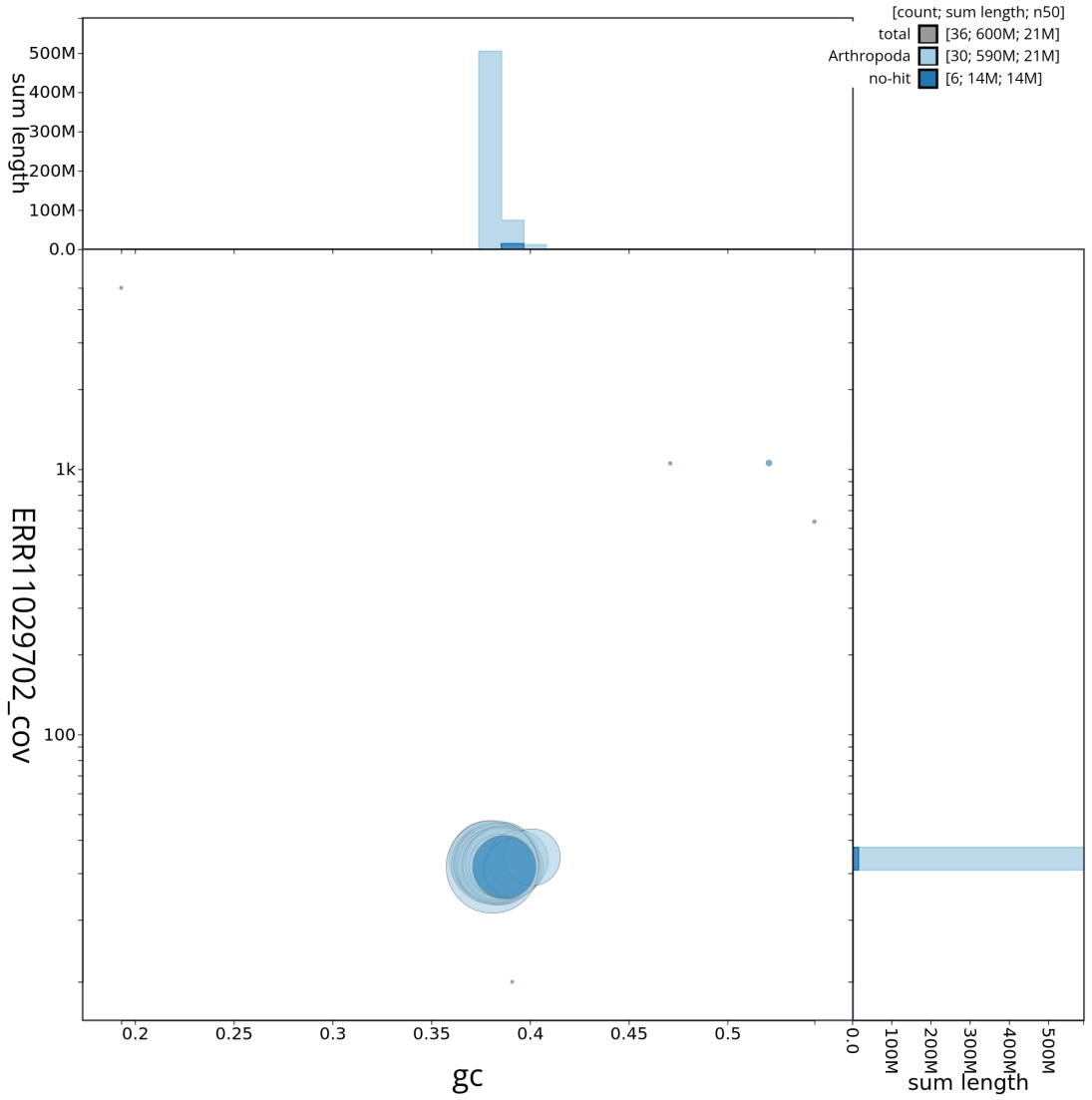

Figure 3. Genome assembly of *Eurois occulta*, ilEurOccu1.1: Blob plot of base coverage against GC proportion for sequences in the assembly. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEurOccu1_1/dataset/ilEurOccu1_1/blob.

were calculated in Merqury (Rhie *et al.*, 2020). This work was done using the "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b) pipelines. The genome readmapping pipelines were developed using the nf-core tooling (Ewels *et al.*, 2020), use MultiQC (Ewels *et al.*, 2016), and make extensive use of the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), and the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions. The genome was also analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021) were calculated.

Table 4 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 *Eurois occulta* assembly (GCA_950022335.1) in Ensembl Rapid Release at the EBI.

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

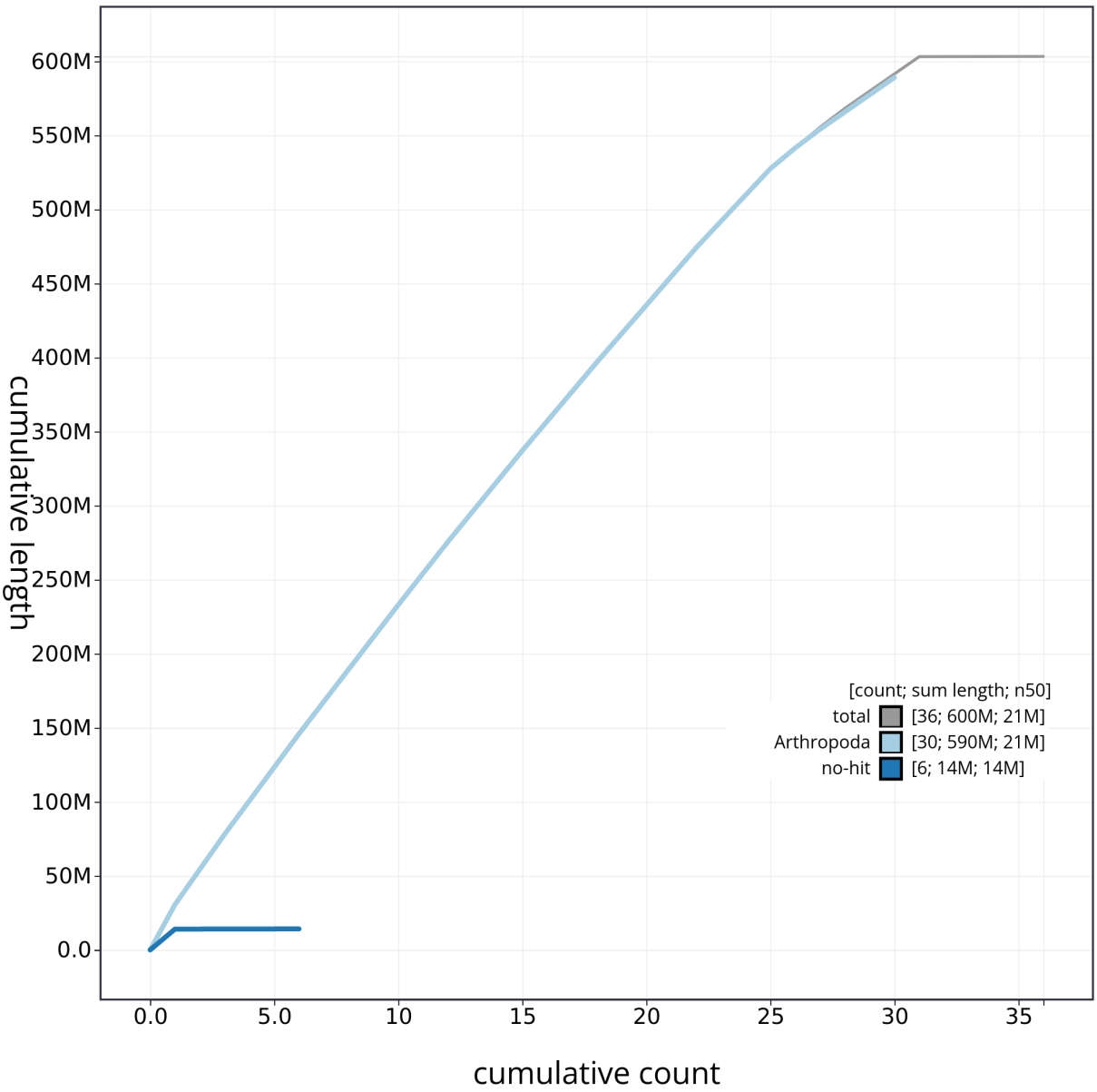

Figure 4. Genome assembly of *Eurois occulta* ilEurOccu1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEurOccu1_1/dataset/ilEurOccu1_1/cumulative.

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)

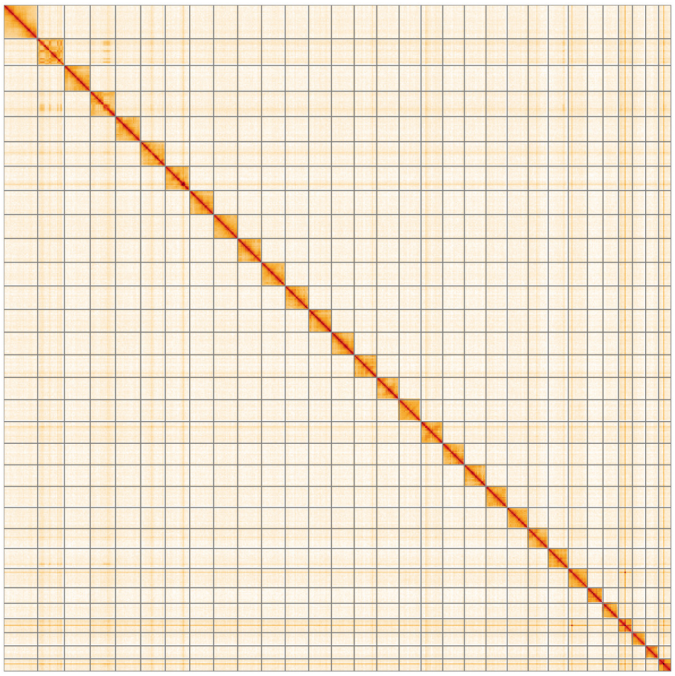

Figure 5. Genome assembly of *Eurois occulta* ilEurOccu1.1: Hi-C contact map of the ilEurOccu1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=O_MlenrBStqUgwMZ4acdsA.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Eurois occulta*, ilEurOccu1.

INSDC accession	Name	Length (Mb)	GC%
OX465478.1	1	24.12	38.5
OX465479.1	2	23.36	38.5
OX465480.1	3	22.88	38.5
OX465481.1	4	22.79	38.5
OX465482.1	5	22.19	38.0
OX465483.1	6	22.04	38.0
OX465484.1	7	21.73	38.0
OX465485.1	8	21.66	38.5
OX465486.1	9	21.63	38.0
OX465487.1	10	21.3	38.5
OX465488.1	11	21.28	38.0
OX465489.1	12	20.59	38.0
OX465490.1	13	20.5	38.5
OX465491.1	14	20.45	38.0
OX465492.1	15	20.16	38.5
OX465493.1	16	19.88	38.0
OX465494.1	17	19.65	38.5
OX465495.1	18	19.55	38.5
OX465496.1	19	19.54	38.0
OX465497.1	20	19.15	38.5
OX465498.1	21	18.98	38.5
OX465499.1	22	18.11	38.5
OX465500.1	23	18.05	38.5
OX465501.1	24	17.41	38.5
OX465502.1	25	14.05	38.5
OX465503.1	26	13.93	38.5
OX465504.1	27	12.68	39.5
OX465505.1	28	11.93	39.0
OX465506.1	29	11.7	39.0
OX465507.1	30	11.31	40.0
OX465477.1	Z	30.6	38.0
OX465508.1	MT	0.02	19.5

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury.FK	d00d98157618f4e8d1a9 190026b19b471055b22e	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
Singularity	3.9.0	https://github.com/sylabs/singularity
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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: *Eurois occulta* (great brocade). Accession number PRJEB60716; https://identifiers.org/ena.embl/PRJEB60716 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Eurois occulta* 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 and Table 2.

Author information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.12162482.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.12165051.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.12160324.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.12205391.

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

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **Sequence locally, think globally: the Darwin Tree of Life project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Broad GR, Boyes D, Natural History Museum Genome Acquisition Lab, *et al.*: **The genome sequence of the setaceous Hebrew character, *Xestia c-nigrum*, (Linnaeus, 1758) [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2022; **7**: 295.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Brůna T, Hoff KJ, Lomsadze A, *et al.*: **BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom Bioinform.* 2021; **3**(1): 1–11, 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**

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | **Publisher Full Text**

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

GBIF Secretariat: ***Eurois occulta* (Linnaeus, 1758)**. 2024.
Reference Source

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
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.
Reference Source

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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

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

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
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**

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
Reference Source

NBN Atlas: ***Eurois occulta* (Linnaeus, 1758)**. 2024.
Reference Source

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.
Publisher 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**

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

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

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

UK Moths: ***Eurois occulta* – Great Brocade**. 2024.
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 Great Brocade, *Eurois occulta* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB60716, 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**