DATA NOTE

The genome sequence of the Dusky Sallow, *Eremobia ochroleuca* (Denis & Schiffermüller) 1775 [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Clare Boyes[2],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
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
[2]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

v1 **First published:** 17 Apr 2024, **9**:206
https://doi.org/10.12688/wellcomeopenres.21226.1
Latest published: 17 Apr 2024, **9**:206
https://doi.org/10.12688/wellcomeopenres.21226.1

Abstract
We present a genome assembly from an individual male *Eremobia ochroleuca* (the Dusky Sallow; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 625.4 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 15.36 kilobases in length. Gene annotation of this assembly on Ensembl identified 18,530 protein coding genes.

Keywords
Eremobia ochroleuca, dusky sallow, 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: Boyes D: Investigation, Resources; **Boyes C**: 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, 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: Boyes D, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Dusky Sallow, *Eremobia ochroleuca* (Denis & Schiffermüller) 1775 [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:206 https://doi.org/10.12688/wellcomeopenres.21226.1

First published: 17 Apr 2024, **9**:206 https://doi.org/10.12688/wellcomeopenres.21226.1

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; Xyleninae; *Eremobia*; *Eremobia ochroleuca* (Denis & Schiffermüller) 1775 (NCBI:txid1870628).

Background

Dusky Sallow (*Eremobia ochroleuca*) is a moth in the family Noctuidae which is common and widespread in eastern England but much more local in the west of England (Randle *et al.*, 2019). It is found throughout Europe, although it is often local (GBIF Secretariat, 2024). The adult moth has a forewing length of 14–16 mm and is fairly distinctive. The forewings are variegated light orangey-brown, dark brown and white and have chequered fringes. The adult has one generation a year, and flies in July and August. It readily comes to light and sugar and can be found nectaring on flowers. During daytime, adults can also be found resting on knapweeds (Waring *et al.*, 2017). There are several sub-species in Europe (Leraut, 2019).

Dusky Sallow lays its barrel-shaped eggs within the sheath of tall grasses and so favours a range of grassy habitats including grassland, vegetated shingle and woodland rides. Eggs overwinter before hatching the following spring. The larvae are active during the night and day, feeding on the seeds of grasses such as cock's-foot (*Dactylis glomerata*), and can be readily swept from vegetation in suitable habitats (Heath & Emmet, 1983). Daytime feeding makes them vulnerable to the sand wasp, *Ammophila sabulosa*, which commonly parasitises this species (Field, 1989). The larvae pupate underground in a cocoon (Heath & Emmet, 1983) before emerging during July. Since the 1970s, the flight period has changed and it now peaks about 2 weeks earlier (Randle *et al.*, 2019).

The genome of *Eremobia ochroleuca* 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 *Eremobia ochroleuca* based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from a male *Eremobia ochroleuca* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.31). A total of 34-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 404 missing joins or mis-joins and removed 54 haplotypic duplications, reducing the assembly length by 0.71% and the scaffold number by 36.67%, and increasing the scaffold N50 by 2.52%.

Figure 1. Photograph of the *Eremobia ochroleuca* (ilEreOchr1) specimen used for genome sequencing.

The final assembly has a total length of 625.4 Mb in 277 sequence scaffolds with a scaffold N50 of 21.5 Mb (Table 1). The snail plot 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 (99.21%) 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 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 60.1 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 94.9% (single = 94.2%, duplicated = 0.7%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/1870628.

Genome annotation report

The *Eremobia ochroleuca* genome assembly (GCA_949629135.1) was annotated using the Ensembl rapid annotation pipeline at the European Bioinformatics Institute (EBI). The resulting annotation includes 18,733 transcribed mRNAs from 18,530 protein-coding genes (Table 1; https://rapid.ensembl.org/Eremobia_ochroleuca_GCA_949629135.1/Info/Index).

Table 1. Genome data for *Eremobia ochroleuca*, ilEreOchr1.1.

Project accession data		
Assembly identifier	ilEreOchr1.1	
Species	*Eremobia ochroleuca*	
Specimen	ilEreOchr1	
NCBI taxonomy ID	1870628	
BioProject	PRJEB58950	
BioSample ID	SAMEA10978936	
Isolate information	ilEreOchr1, male: thorax (DNA sequencing), head (Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	60.1	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:94.9%[S:94.2%,D:0.7%], F:1.1%,M:4.1%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.21%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.36 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10798425	
Hi-C Illumina	ERR10786026	
Genome assembly		
Assembly accession	GCA_949629135.1	
Accession of alternate haplotype	GCA_949629145.1	
Span (Mb)	625.4	
Number of contigs	2,138	
Contig N50 length (Mb)	0.7	
Number of scaffolds	277	
Scaffold N50 length (Mb)	21.5	
Longest scaffold (Mb)	32.47	
Genome annotation		
Number of protein-coding genes	18,530	
Number of gene transcripts	18,733	

* 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/ilEreOchr1_1/dataset/ilEreOchr1_1/busco.

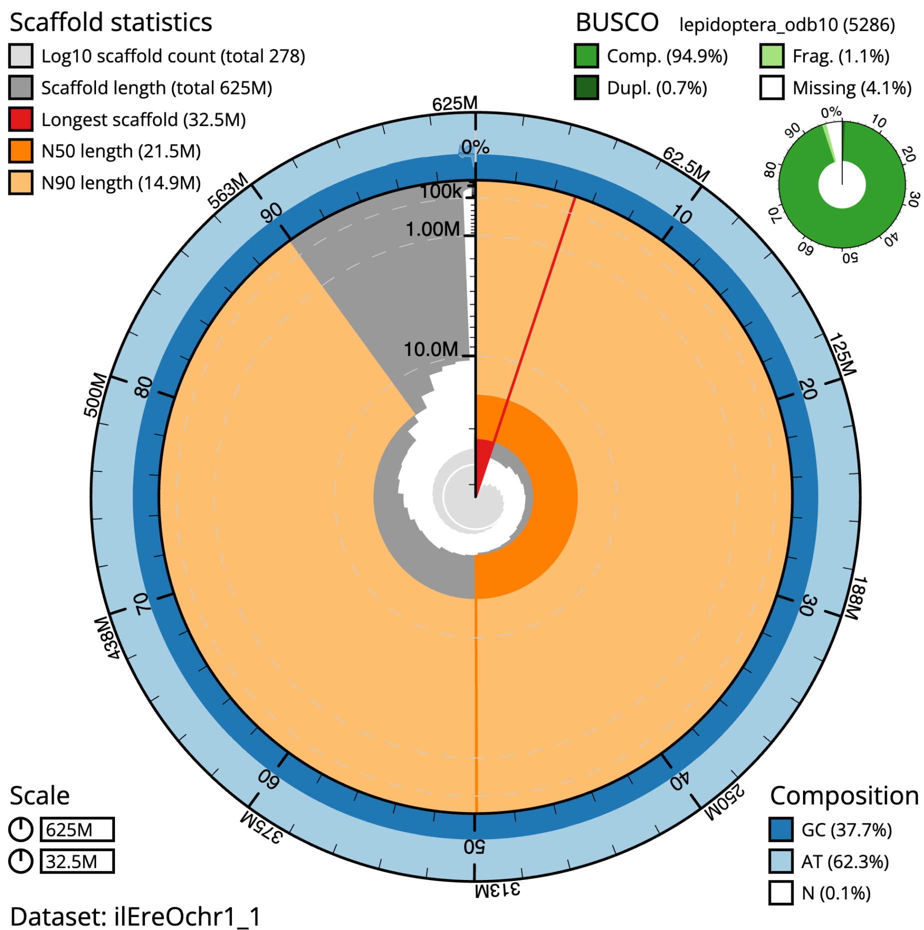

Figure 2. Genome assembly of *Eremobia ochroleuca*, ilEreOchr1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 625,465,034 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (32,468,123 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (21,456,284 and 14,898,524 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/ilEreOchr1_1/dataset/ilEreOchr1_1/snail.

Methods

Sample acquisition and nucleic acid extraction

A male *Eremobia ochroleuca* (specimen ID Ox001667, ToLID ilEreOchr1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.31) on 2021-07-17 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilEreOchr1 sample was weighed and dissected 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 using the Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

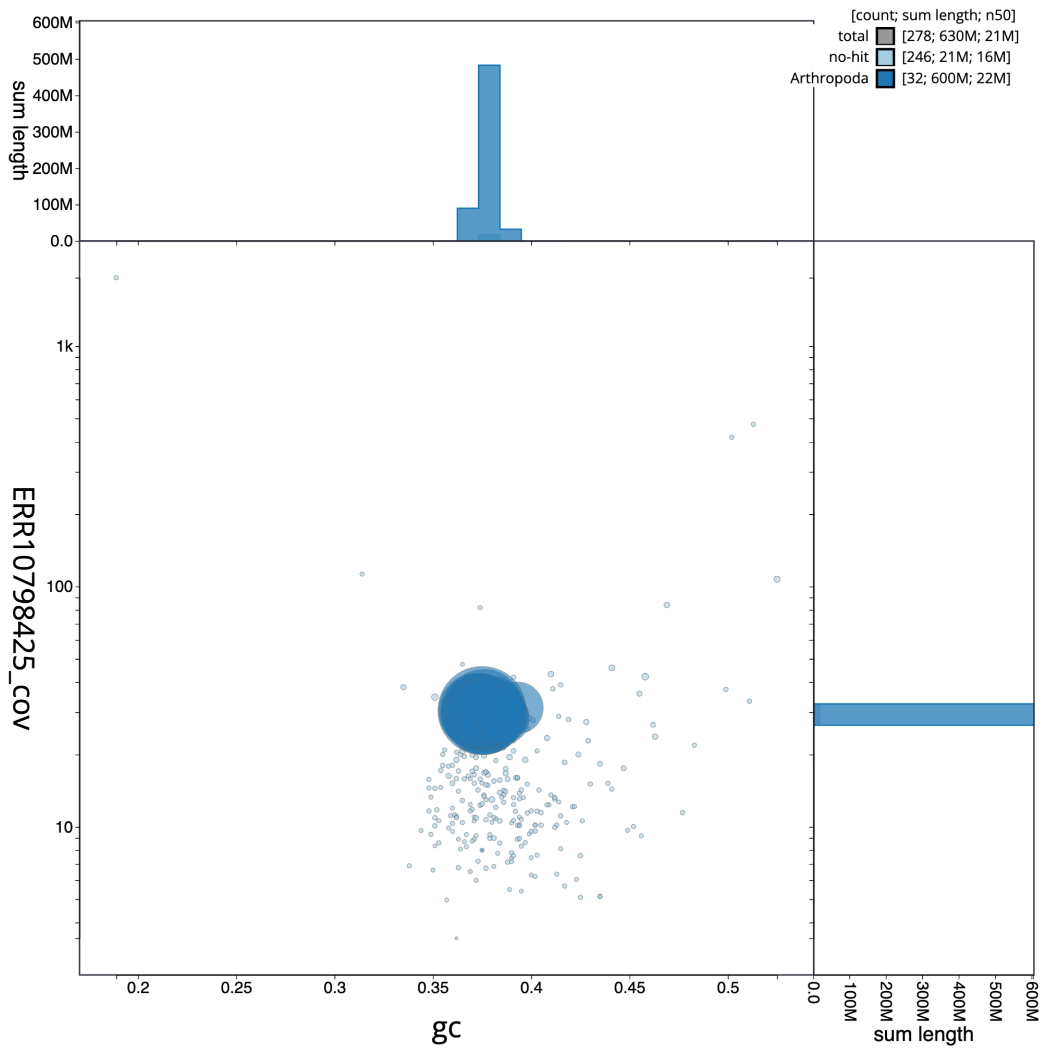

Figure 3. Genome assembly of *Eremobia ochroleuca*, ilEreOchr1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEreOchr1_1/dataset/ilEreOchr1_1/blob.

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 instrument. Hi-C data were also generated from head tissue of ilEreOchr1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination 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

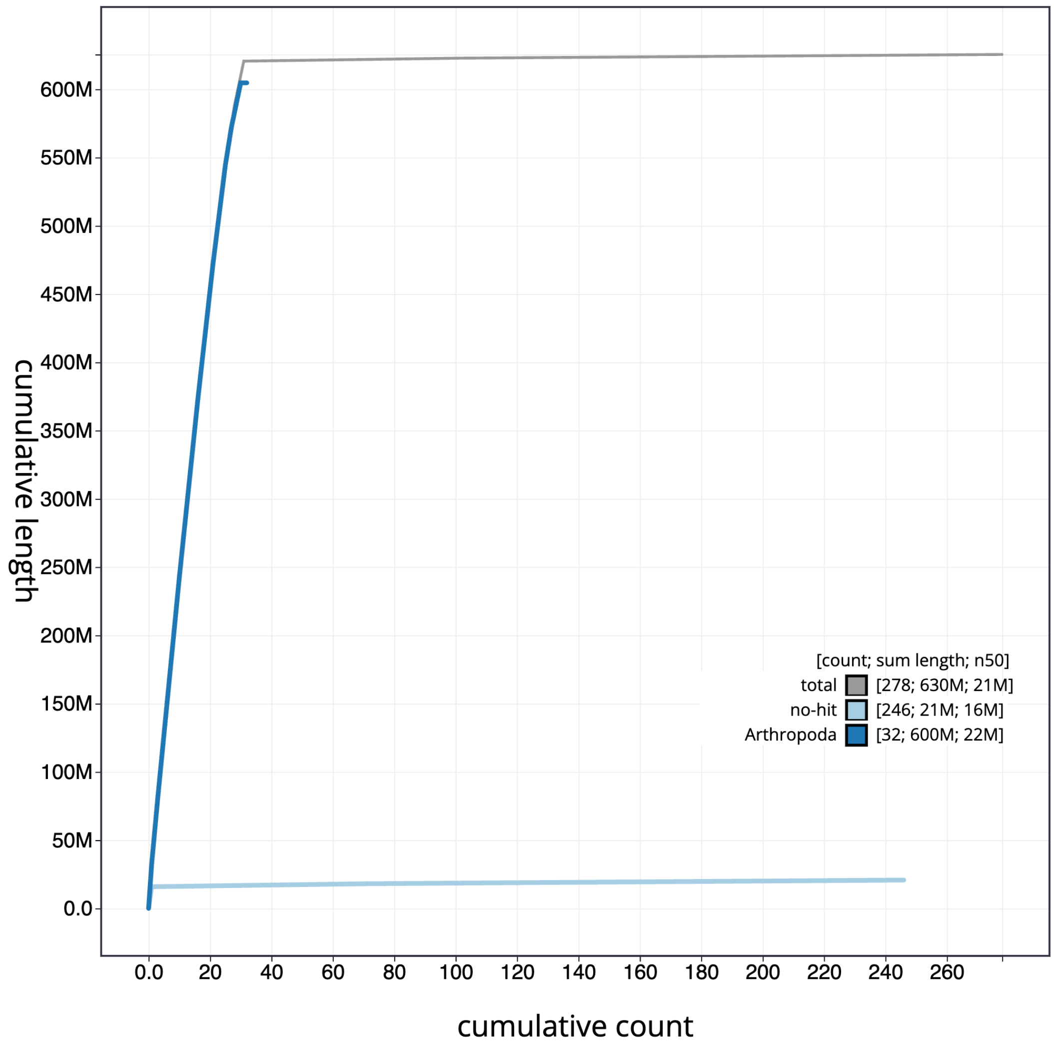

Figure 4. Genome assembly of *Eremobia ochroleuca*, ilEreOchr1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEreOchr1_1/dataset/ilEreOchr1_1/cumulative.

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 *Eremobia ochroleuca* assembly (GCA_949629135.1) 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 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

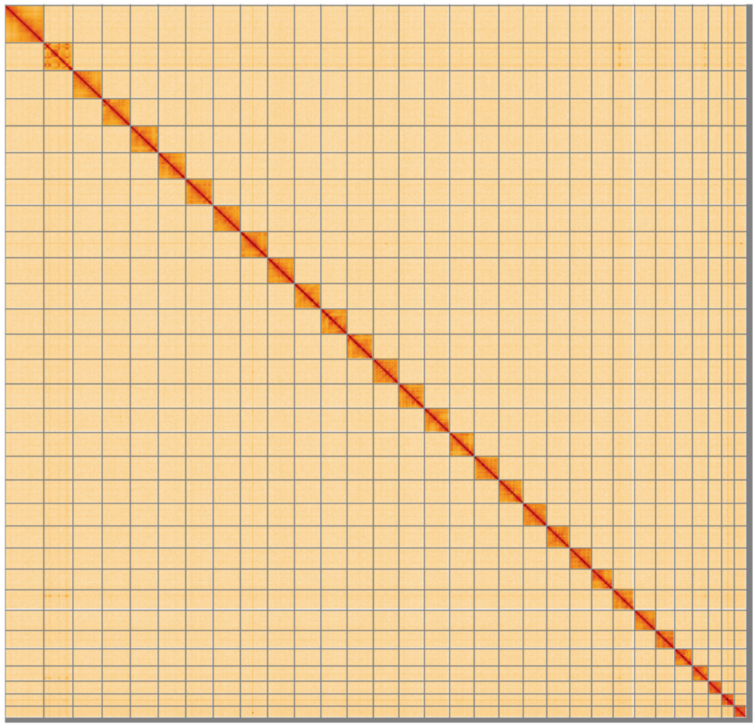

Figure 5. Genome assembly of *Eremobia ochroleuca*, ilEreOchr1.1: Hi-C contact map of the ilEreOchr1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=e-kWqAlBQ-yNSy22Ch5sXQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eremobia ochroleuca*, ilEreOchr1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX451385.1	1	24.36	37.5
OX451386.1	2	24.35	37.5
OX451387.1	3	23.66	37.5
OX451388.1	4	23.32	37.5
OX451389.1	5	23.05	37.0
OX451390.1	6	22.9	37.5
OX451391.1	7	22.83	37.5
OX451392.1	8	22.72	37.5
OX451393.1	9	22.5	37.5
OX451394.1	10	22.1	37.5
OX451395.1	11	21.99	37.5
OX451396.1	12	21.88	37.5
OX451397.1	13	21.46	37.5
OX451398.1	14	21.26	37.5
OX451399.1	15	21.1	37.5
OX451400.1	16	20.64	37.5
OX451401.1	17	20.6	37.5
OX451402.1	18	20.36	37.5
OX451403.1	19	19.71	38.0
OX451404.1	20	19.22	38.0
OX451405.1	21	18.36	37.5
OX451406.1	22	18.02	37.5
OX451407.1	23	17.82	38.0
OX451408.1	24	17.69	38.0
OX451409.1	25	15.91	38.0
OX451410.1	26	14.9	38.0
OX451411.1	27	13.18	38.5
OX451412.1	28	11.18	38.5
OX451413.1	29	10.66	39.5
OX451414.1	30	10.41	38.5
OX451384.1	Z	32.47	37.5
OX451415.1	MT	0.02	19.0

Table 3. 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
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	1.2a	https://github.com/c-zhou/yahs

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: *Eremobia ochroleuca*. Accession number PRJEB58950; https://identifiers.org/ena.embl/PRJEB58950 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Eremobia ochroleuca* 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.7125292.

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 Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

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

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

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

Field J: **Intraspecific parasitism and nesting success in the solitary wasp *Ammophila sabulosa*.** *Behaviour.* 1989; **110**(1– 4): 23–46.
Reference Source

GBIF Secretariat: ***Eremobia ochroleuca* (Denis & Schiffermüller) 1775**. *GBIF Backbone Taxonomy.* 2024; [Accessed 29 February 2024].
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

Heath J, Emmet AM: **The moths and butterflies of Great Britain and Ireland Noctuidae (Part II) and Agaristidae**. Colchester: Harley Books, 1983.

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

Leraut P: **Moths of Europe – Noctuids 1.** Verrières-le-Buisson: N.A.P. Editions, 2019.

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

Sheerin E, Sampaio F, Oatley G, *et al.*: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.1**. *protocols.io*. 2023.
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

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

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: Diagenode Megaruptor®3 for PacBio HiFi**. *protocols.io*. 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

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 Dusky Sallow, *Eremobia ochroleuca* (Denis & Schiffermüller) 1775.** European Nucleotide Archive, [dataset], accession number PRJEB58950, 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