DATA NOTE

The genome sequence of the Hoary Bell moth, *Eucosma cana* (Haworth, 1811) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Denise C. Wawman[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]Edward Grey Institute, Department of Biology, University of Oxford, Oxford, England, UK

[+] Deceased author

v1 **First published:** 12 Apr 2024, **9**:185
https://doi.org/10.12688/wellcomeopenres.21212.1
Latest published: 12 Apr 2024, **9**:185
https://doi.org/10.12688/wellcomeopenres.21212.1

Abstract

We present a genome assembly from an individual female *Eucosma cana* (the Hoary Bell; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence is 580.3 megabases in span. Most of the assembly is scaffolded into 28 chromosomal pseudomolecules, including the Z and W sex chromosomes. The mitochondrial genome has also been assembled and is 16.76 kilobases in length.

Keywords

Eucosma cana, Hoary Bell 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: Boyes D: Investigation, Resources; **Wawman DC**: 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, Wawman DC, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Hoary Bell moth, *Eucosma cana* (Haworth, 1811) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:185 https://doi.org/10.12688/wellcomeopenres.21212.1

First published: 12 Apr 2024, **9**:185 https://doi.org/10.12688/wellcomeopenres.21212.1

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; Apoditrysia; Tortricoidea; Tortricidae; Tortricinae; Olethreutinae; Eucosmini; *Eucosma*; *Eucosma cana* (Haworth, 1811) (NCBI:txid1100988).

Background

Eucosma cana (homotypic synonym *Cenopis cana*), the Hoary Bell or Hoary Tortrix, is a micro-moth in the family Tortricidae. Its forewing is 7–11 mm in length and dark or greyish brown with white or pale yellowish-brown streaks. It is paler and has a streaky appearance not seen in the similar species in the *Eucosma hohenwartiana* group: *E. hohenwartiana*, *E. fulvana* and *E. parulana* (Sterling *et al.*, 2023).

Eucosma cana is found throughout the United Kingdom and Ireland, with adults usually on the wing from late-May to August. The larvae feed from August to May in the flowerheads and seedheads of Greater Knapweed *Centaurea scabiosa*, Common or Black Knapweed *Centaurea nigra*, Spear Thistle *Cirsium vulgare*, and thistles in the genus *Carduus* (Bullock *et al.*, 1994; Ehlers & Olesen, 2003; Sterling *et al.*, 2023). *E. cana* overwinters as pupae in a cocoon hidden in leaf litter or soil (Sterling *et al.*, 2023).

Experiments have shown that plants of *Crisium vulgare* with more seedheads suffer a higher proportion of attacks by *E. cana* than those with fewer seedheads (Bullock *et al.*, 1994) and that as the distance between plants of *Centaurea scabiosa* increases the damage caused by *E. cana* decreases (Ehlers & Olesen, 2003).

We present a chromosomally complete genome sequence for a female *Eucosma cana*, based on one specimen collected at Wytham Woods, Oxfordshire, as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from one female *Eucosma cana* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 47-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 8 missing joins or mis-joins and removed 4 haplotypic duplications, reducing the scaffold number by 0.81%, and decreasing the scaffold N50 by 3.29%.

The final assembly has a total length of 580.3 Mb in 121 sequence scaffolds with a scaffold N50 of 19.9 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

Figure 1. Photograph of the *Eucosma cana* (ilEucCana1) specimen used for genome sequencing.

curves for subsets of scaffolds assigned to different phyla. Most (99.39%) of the assembly sequence was assigned to 28 chromosomal-level scaffolds, representing 26 autosomes and the Z and W sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). The W chromosome was not scaffolded as Hi-C data are from a male (ZZ) sample (ilEucCana2). 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 67.5 with *k*-mer completeness of 100.0%, 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, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/1100988.

Methods

Sample acquisition and nucleic acid extraction

The specimen used for DNA sequencing was a female *Eucosma cana* (specimen ID Ox000691, ToLID ilEucCana1), collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-07-20 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford), and snap-frozen on dry ice.

The specimen used for Hi-C and RNA sequencing (specimen ID SAN00002574, ToLID ilEucCana2) was collected

Table 1. Genome data for *Eucosma cana*, ilEucCana1.1.

Project accession data		
Assembly identifier	ilEucCana1.1	
Species	*Eucosma cana*	
Specimen	ilEucCana1	
NCBI taxonomy ID	1100988	
BioProject	PRJEB55956	
BioSample ID	SAMEA7701552	
Isolate information	ilEucCana1, female: whole organism (DNA sequencing) ilEucCana2, male: thorax (Hi-C and RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	67.5	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:98.2%[S:97.6%,D:0.6%], F:0.4%,M:1.4%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.39%	*≥ 95%*
Sex chromosomes	ZW	*localised homologous pairs*
Organelles	Mitochondrial genome: 16.76 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10224903	
Hi-C Illumina	ERR11271507	
PolyA RNA-Seq Illumina	ERR11242508	
Genome assembly		
Assembly accession	GCA_951800055.1	
Accession of alternate haplotype	GCA_951800005.1	
Span (Mb)	580.3	
Number of contigs	153	
Contig N50 length (Mb)	16.7	
Number of scaffolds	121	
Scaffold N50 length (Mb)	19.9	
Longest scaffold (Mb)	48.21	

* 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/ilEucCana1_1/dataset/ilEucCana1_1/busco.

from Pentland Hills, South Lanarkshire, Scotland, UK (latitude 55.72, longitude –3.51) on 2022-06-17 using a moth trap. The specimen was collected and identified by Jo Davis (independent researcher) and then snap-frozen 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.

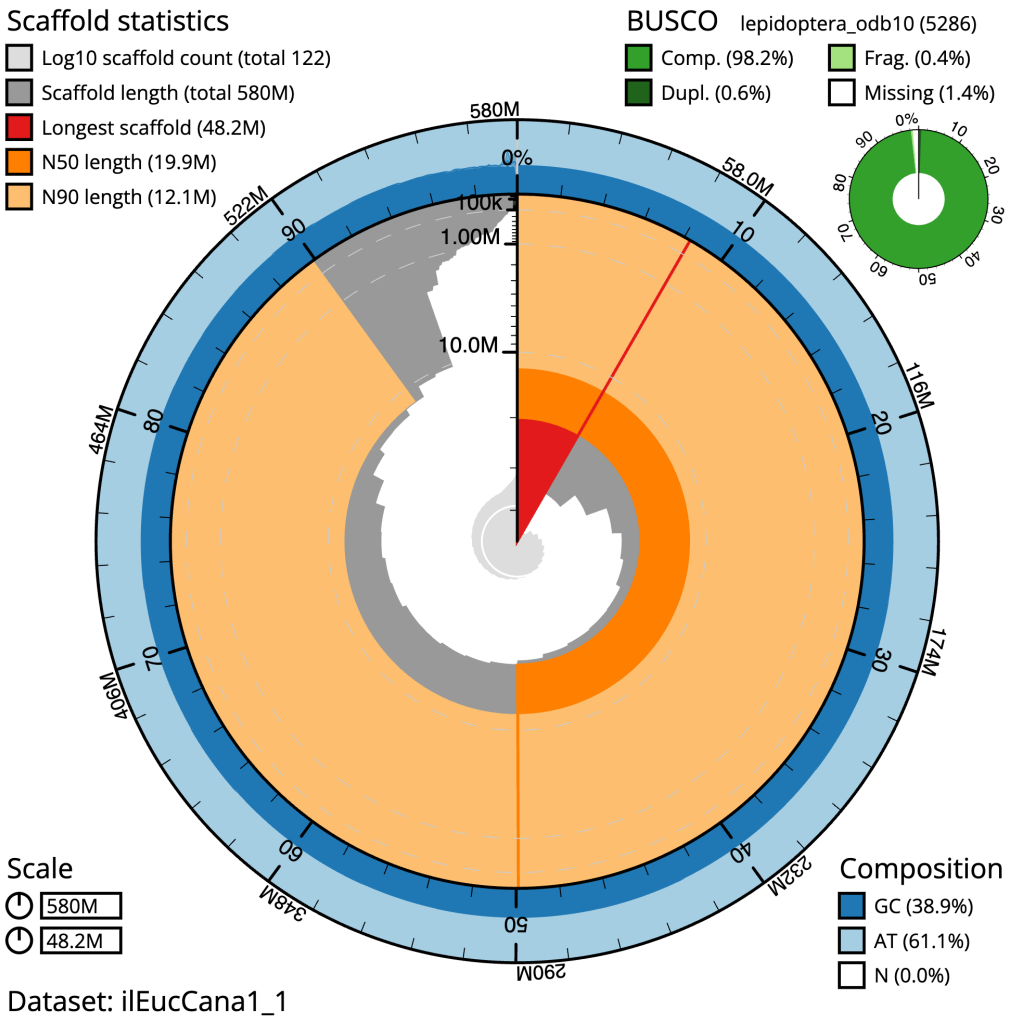

Figure 2. Genome assembly of *Eucosma cana*, ilEucCana1.1: metrics. The BlobToolKit Snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 580,353,609 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (48,205,600 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (19,939,996 and 12,140,148 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/ilEucCana1_1/dataset/ilEucCana1_1/snail.

In sample preparation, the ilEucCana1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023), and then tissue from the whole organism 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.

RNA was extracted from thorax tissue of ilEucCana2 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

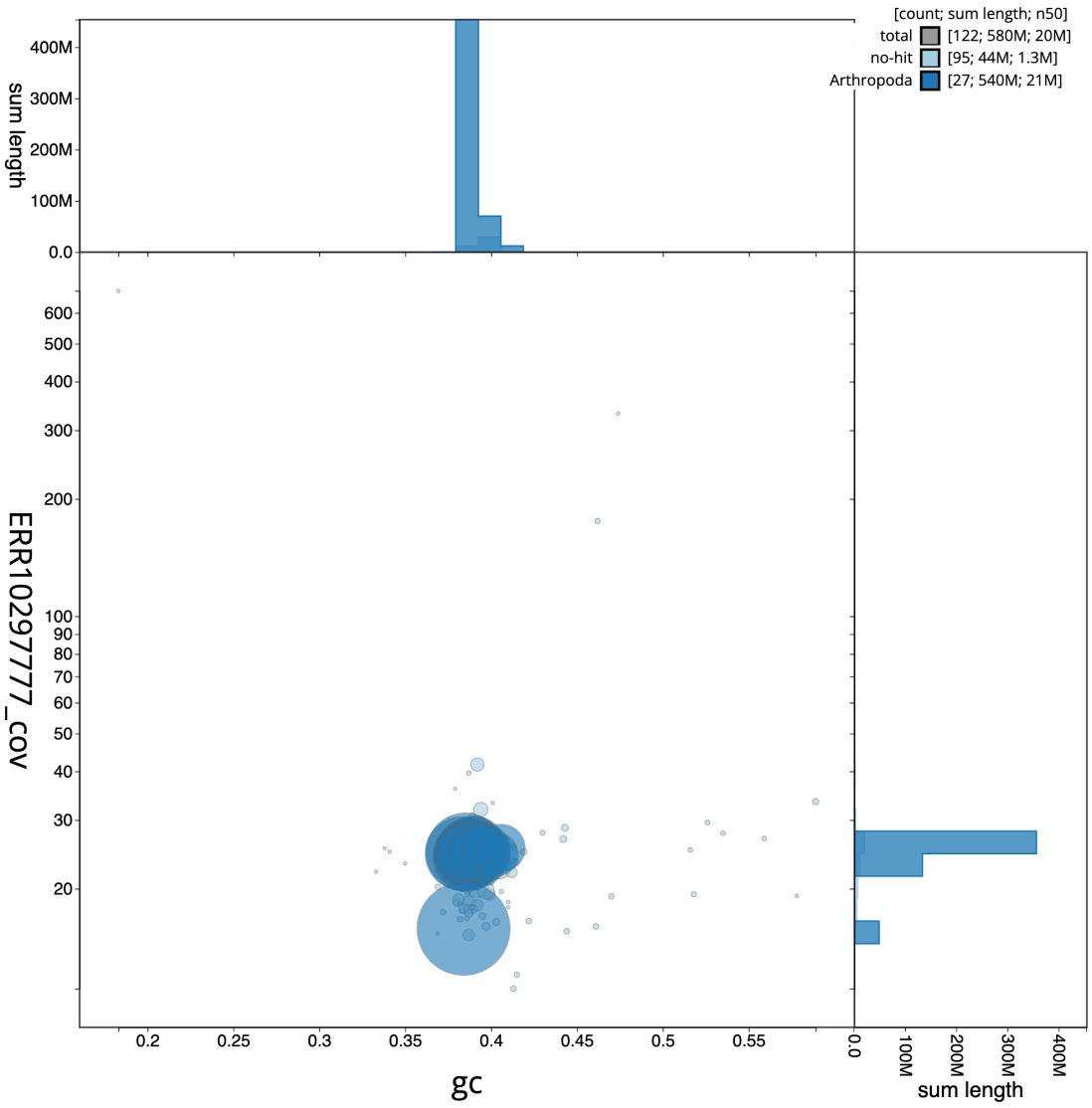

Figure 3. Genome assembly of *Eucosma cana*, ilEucCana1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEucCana1_1/dataset/ilEucCana1_1/blob.

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.

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

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 II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from thorax tissue of ilEucCana2 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 and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView

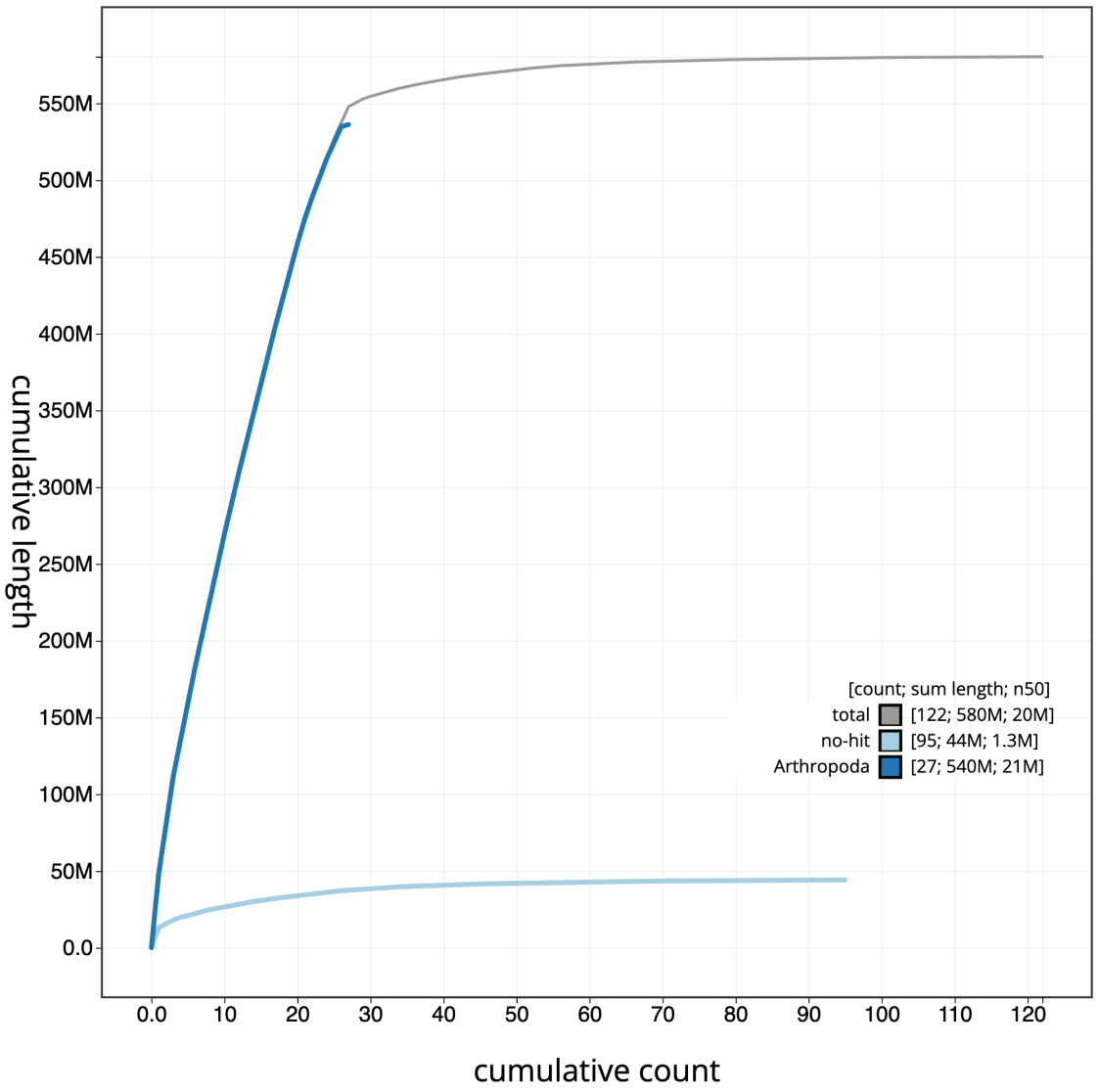

Figure 4. Genome assembly of *Eucosma cana*, ilEucCana1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilEucCana1_1/dataset/ilEucCana1_1/cumulative.

(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 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.

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

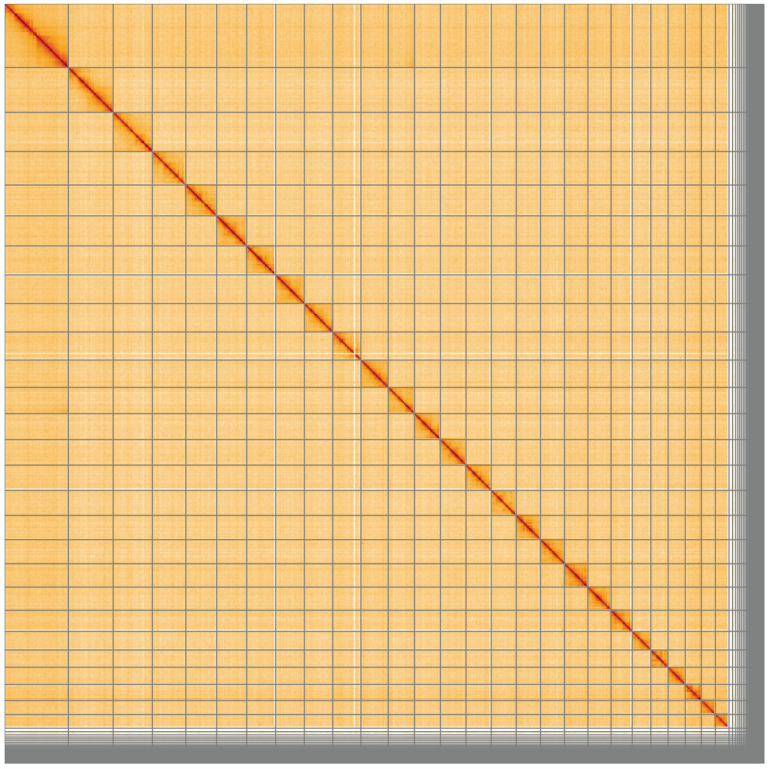

Figure 5. Genome assembly of *Eucosma cana*, ilEucCana1.1: Hi-C contact map of the ilEucCana1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=XdT7LQ6WRWuQKyUo_G8tSg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Eucosma cana*, ilEucCana1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX637472.1	1	33.9	38.5
OX637473.1	2	29.7	38.5
OX637475.1	3	25.22	38.5
OX637476.1	4	23.36	38.5
OX637477.1	5	22.74	39.0
OX637478.1	6	21.81	39.0
OX637479.1	7	21.73	38.5
OX637480.1	8	21.54	38.5
OX637481.1	9	21.23	39.0
OX637482.1	10	20.62	39.0
OX637483.1	11	19.94	39.0
OX637484.1	12	19.59	38.5
OX637485.1	13	19.36	39.0
OX637486.1	14	19.02	38.5
OX637487.1	15	18.99	39.5
OX637488.1	16	18.34	39.0
OX637489.1	17	18.15	38.5
OX637490.1	18	17.67	39.0
OX637491.1	19	17.54	39.0
OX637492.1	20	15.97	39.5
OX637493.1	21	14.06	39.0
OX637494.1	22	13.1	39.5
OX637495.1	23	13.05	39.5
OX637496.1	24	12.14	40.5
OX637497.1	25	10.59	39.5
OX637498.1	26	10.27	40.5
OX637474.1	W	2.66	40.5
OX637471.1	Z	48.21	38.5
OX637499.1	MT	0.02	18.5

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	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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: *Eucosma cana* (hoary bell). Accession number PRJEB55956; https://identifiers.org/ena.embl/PRJEB55956 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Eucosma cana* 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. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. 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

Bullock JM, Hill BC, Silvertown J: **Demography of *Cirsium vulgare* in a Grazing Experiment.** *J Ecol.* 1994; **82**(1): 101–111.
Publisher 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

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

Ehlers BK, Olesen JM: **Flower and fruit herbivory in a population of *Centaurea scabiosa* (Asteraceae): Importance of population size and isolation.** *Écoscience.* 2003; **10**(1): 45–48.
Publisher Full Text

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

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

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

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

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

Sterling P, Parsons M, Lewington R: **Field Guide to the Micro-moths of Great Britain and Ireland.** London: Bloomsbury Publishing, 2023.

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

Wellcome Sanger Institute: **The genome sequence of the Hoary Bell moth, *Eucosma cana* (Haworh, 1811).** European Nucleotide Archive, [dataset], accession number PRJEB55956, 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