DATA NOTE

The genome sequence of the Chocolate-tip, *Clostera curtula* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Tista Mallick[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 Consor

[1]University of Oxford, Oxford, England, UK
[2]Department of Biotechnology, Rama Devi Women's University, Bhubaneswar, Odisha, India

[+] Deceased author

v1 **First published:** 18 Sep 2023, **8**:405
https://doi.org/10.12688/wellcomeopenres.19984.1
Latest published: 18 Sep 2023, **8**:405
https://doi.org/10.12688/wellcomeopenres.19984.1

Abstract

We present a genome assembly from an individual female *Clostera curtula* (the Chocolate-tip; Arthropoda; Insecta; Lepidoptera; Notodontidae). The genome sequence is 512.7 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the W and Z sex chromosomes. The mitochondrial genome has also been assembled and is 15.37 kilobases in length. Gene annotation of this assembly on Ensembl identified 12,251 protein coding genes.

Keywords

Clostera curtula, Chocolate-tip, 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 Consor (Mark.Blaxter@sanger.ac.uk)

Author roles: Boyes D: Investigation, Resources; **Mallick T**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Boyes D, Mallick T, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Chocolate-tip, *Clostera curtula* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:405 https://doi.org/10.12688/wellcomeopenres.19984.1

First published: 18 Sep 2023, **8**:405 https://doi.org/10.12688/wellcomeopenres.19984.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Notodontidae; Pygaerinae; *Clostera; Clostera curtula*, (Linnaeus, 1758) (NCBI:txid987902).

Background

Clostera curtula (Notodontidae: Pygaerinae), also known as the Chocolate-tip, is a moth species first characterised by Carl Linnaeus in the 1758 tenth edition of 'Systema Naturae' (Linnaeus, 1758). The adult moth is medium sized, with a wingspan of 27 to 35 mm. This is a distinctive looking moth, with a buff-hued body, three distinct white crosslines adorning the breadth of its forewings, and a rich, chocolate-brown blotch at the forewing tip, from which it gets its common name. A notable characteristic of this species is the presence of pheromone-producing ring glands, found in the distal half of the eighth/ninth intersegmental membrane (Percy-Cunningham & MacDonald, 1987).

Globally, *C. curtula* is primarily found throughout Europe, spanning from the western regions to as far east as Russia, and extending northwards up to the southern parts of Scandinavia and southward to the Mediterranean (GBIF Secretariat, 2022). This moth has an uneven distribution in Britain, inhabiting predominantly the southern regions of England and eastern Wales, with isolated populations found in specific areas of Scotland, and it is not recorded from Ireland (Waring *et al.*, 2017).

The moth's favoured habitat encompasses woodland areas. Larvae of *C. curtula* feed primarily on poplar species (*Populus*), with a preference for aspen (*P. tremula*), and it also consumes sallow (*Salix*). In Scotland, aspen is the main food plant, and it is only found in aspen woodlands of significant size and is associated with mature trees (Young, 2001).

The English population has two generations, with adults flying in April and May, and then reappearing in August and September. Conversely, populations in Scotland are single-brooded, taking flight from June through July (Kimber, 2023). The species overwinters as a pupa in a cocoon spun on the larval foodplant.

The generation of a reference genome for *Clostera curtula* as part of the Darwin Tree of Life project will serve as a valuable resource for the molecular surveillance of the species. Additionally, it will aid in phylogenetic studies, providing insights into its evolutionary relationships with related species.

Genome sequence report

The genome was sequenced from one female *Clostera curtula* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 24-fold coverage in Pacific

Figure 1. Photograph of the *Clostera curtula* (ilCloCurt1) specimen used for genome sequencing.

Biosciences single-molecule HiFi long reads and 77-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 94 missing joins or mis-joins and removed 14 haplotypic duplications, reducing the assembly length by 0.54% and the scaffold number by 67.39%, and increasing the scaffold N50 by 25.04%.

The final assembly has a total length of 512.7 Mb in 30 sequence scaffolds with a scaffold N50 of 18.3 Mb (Table 1). The whole assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 28 autosomes and the W and Z sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 57.4 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.5% (single = 98.1%, duplicated = 0.4%), 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/987902.

Genome annotation report

The *Clostera curtula* genome assembly (GCA_905475355.2) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Clostera_curtula_GCA_905475355.1/Info/Index). The resulting annotation includes 20,574 transcribed mRNAs from 12,251 protein-coding and 1,878 non-coding genes.

Table 1. Genome data for *Clostera curtula*, ilCloCurt1.2.

Project accession data		
Assembly identifier	ilCloCurt1.2	
Species	*Clostera curtula*	
Specimen	ilCloCurt1	
NCBI taxonomy ID	987902	
BioProject	PRJEB43795	
BioSample ID	SAMEA7520526	
Isolate information	ilCloCurt1, female: abdomen (DNA sequencing); head and thorax (Hi-C scaffolding and RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.4	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:98.5%[S:98.1%,D:0.4%],F:0.4%, M:1.1%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	100%	*≥ 95%*
Sex chromosomes	Z and W chromosomes	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6412033	
10X Genomics Illumina	ERR6054599, ERR6054600, ERR6054597, ERR6054598	
Hi-C Illumina	ERR6054596	
PolyA RNA-Seq Illumina	ERR9434970	
Genome assembly		
Assembly accession	GCA_905475355.2	
Accession of alternate haplotype	GCA_905475325.1	
Span (Mb)	512.7	
Number of contigs	141	
Contig N50 length (Mb)	6.8	
Number of scaffolds	30	
Scaffold N50 length (Mb)	18.3	
Longest scaffold (Mb)	27.4	
Genome annotation		
Number of protein-coding genes	12,251	
Number of non-coding genes	1,878	
Number of gene transcripts	20,574	

* 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/Clostera%20curtula/dataset/ilCloCurt1_2/busco.

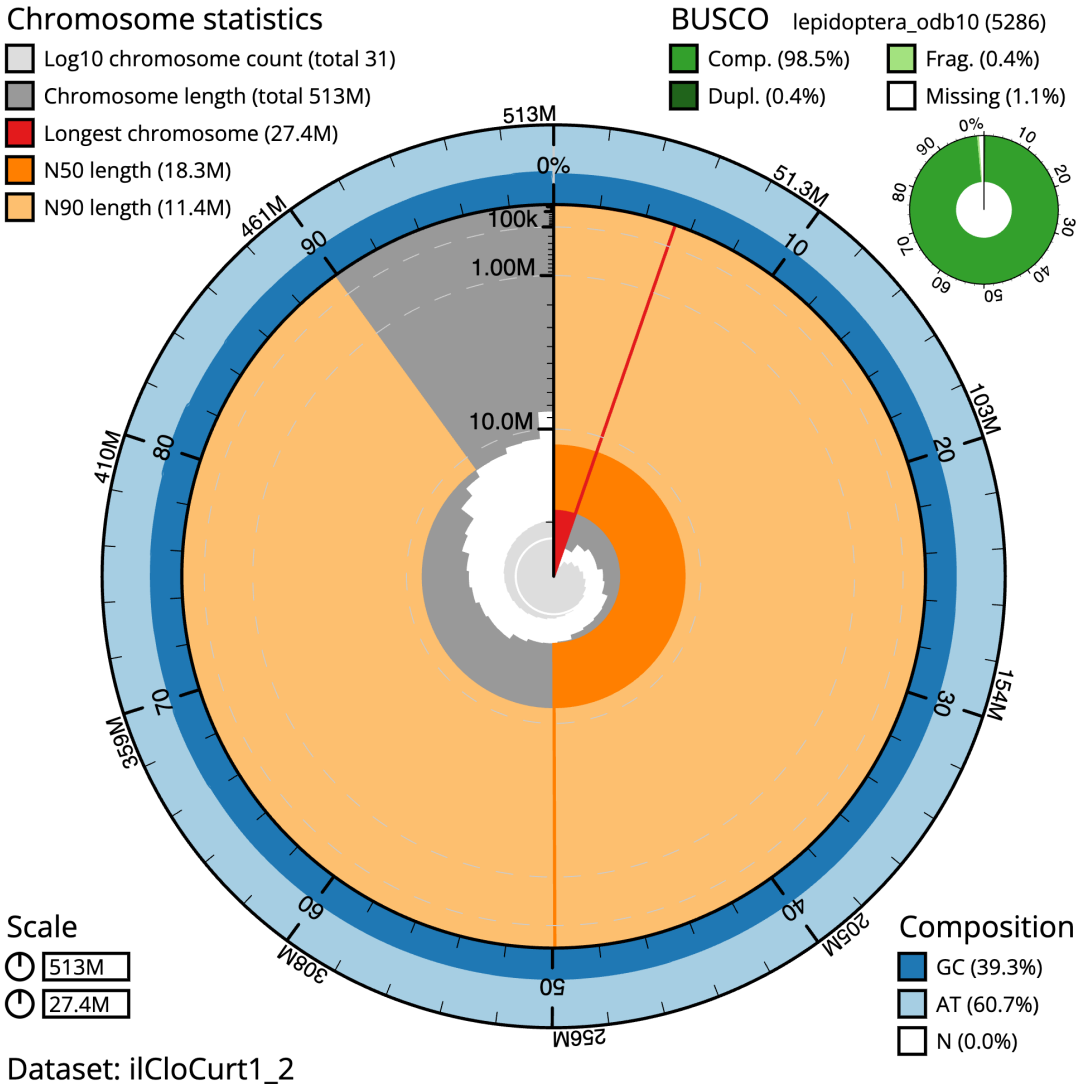

Figure 2. Genome assembly of *Clostera curtula*, ilCloCurt1.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 512,681,271 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (27,404,194 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (18,278,162 and 11,407,441 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/Clostera%20curtula/dataset/ilCloCurt1_2/snail.

Methods

Sample acquisition and nucleic acid extraction

A female *Clostera curtula* (specimen ID Ox000404, individual ilCloCurt1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-05-22 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilCloCurt1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Tissue from the abdomen was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using the 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA

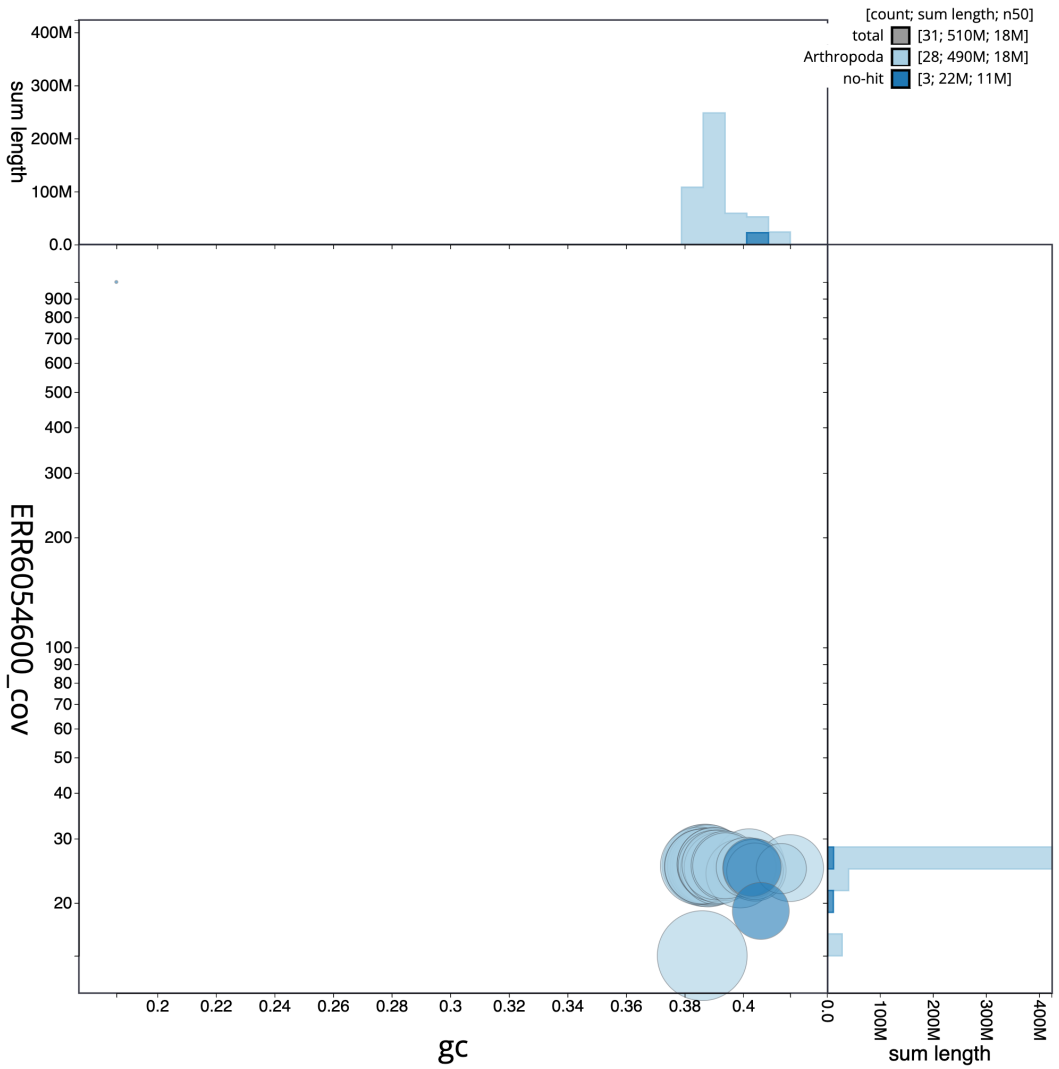

Figure 3. Genome assembly of *Clostera curtula*, ilCloCurt1.2: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Clostera%20curtula/dataset/ilCloCurt1_2/blob.

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.

RNA was extracted from head and thorax tissue of ilCloCurt1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from head and thorax tissue of ilCloCurt1 using the Arima2 kit and sequenced on the HiSeq X Ten instrument.

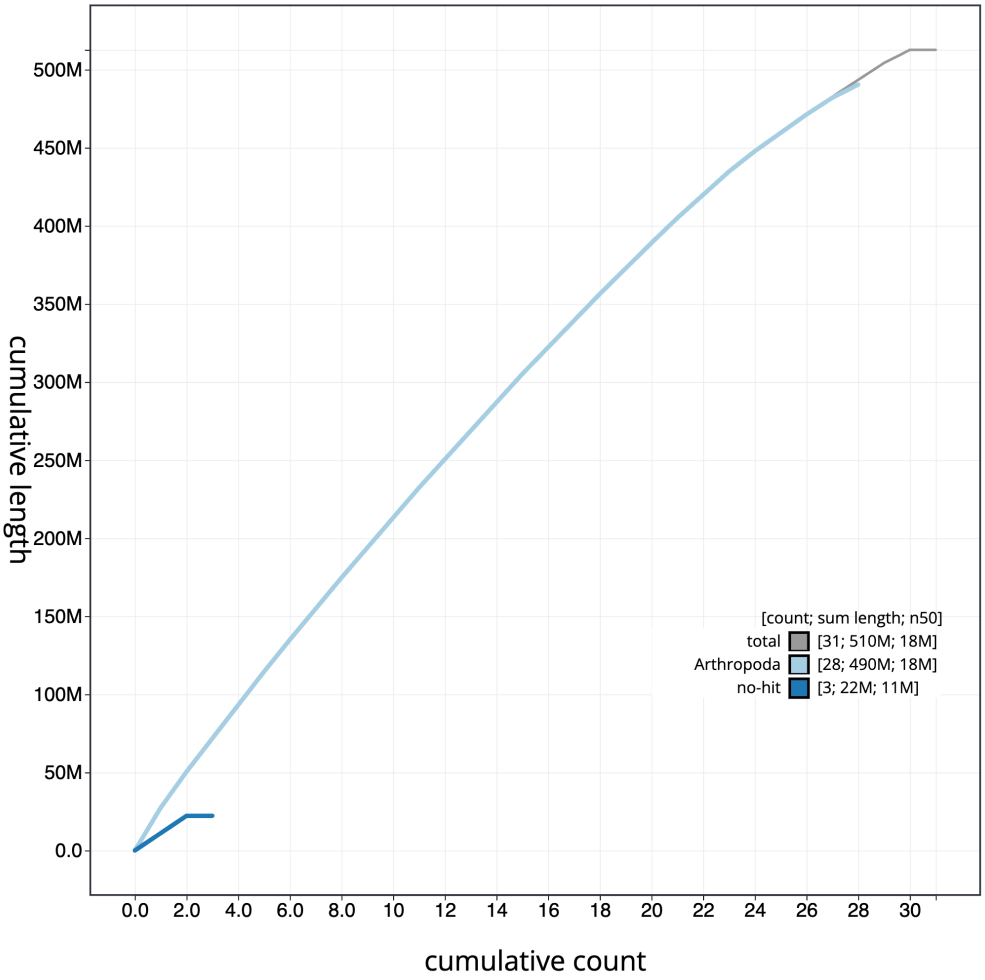

Figure 4. Genome assembly of *Clostera curtula*, ilCloCurt1.2: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Clostera%20curtula/dataset/ilCloCurt1_2/cumulative.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, 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 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 Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Clostera curtula*

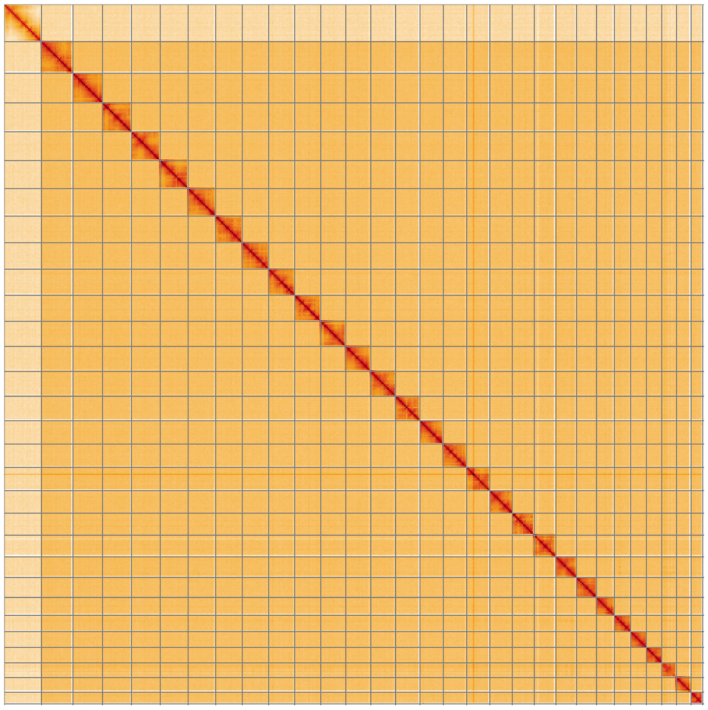

Figure 5. Genome assembly of *Clostera curtula*, ilCloCurt1.2: Hi-C contact map of the ilCloCurt1.2 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=TkYvdbn6TRCOaQlzXkVLxg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Clostera curtula*, ilCloCurt1.

INSDC accession	Chromosome	Length (Mb)	GC%
FR997795.1	1	23.05	38.5
FR997796.1	2	21.74	38.5
FR997797.1	3	21.15	39.0
FR997798.1	4	21.15	39.0
FR997799.1	5	20.48	38.5
FR997800.1	6	20.25	39.0
FR997801.1	7	19.48	39.0
FR997802.1	8	19.37	38.5
FR997803.1	9	19.22	38.5
FR997804.1	10	19.03	38.5
FR997805.1	11	18.48	39.0
FR997806.1	12	18.28	39.0
FR997807.1	13	18.22	39.0
FR997808.1	14	17.87	39.0
FR997809.1	15	17.18	39.0
FR997810.1	16	17.15	39.0
FR997811.1	17	16.91	40.0
FR997812.1	18	16.48	39.0
FR997813.1	19	16.21	39.5
FR997814.1	20	15.93	40.0
FR997815.1	21	15.07	41.5
FR997816.1	22	14.69	39.5
FR997817.1	23	13.02	40.5
FR997818.1	24	12.08	40.0
FR997819.1	25	11.41	40.5
FR997820.1	26	11.38	40.5
FR997822.1	27	10.74	40.5
FR997823.1	28	8.41	41.5
FR997821.1	W	10.84	40.5
FR997794.1	Z	27.4	38.5
FR997824.2	MT	0.02	19.0

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
FreeBayes	1.3.1-17-gaa2ace8	https://github.com/freebayes/freebayes
gEVAL	N/A	https://geval.org.uk/
Hifiasm	0.12	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
SALSA	2.2	https://github.com/salsa-rs/salsa
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0

assembly (GCA_905475355.2). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website here. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Clostera curtula* (chocolate-tip). Accession number PRJEB43795; https://identifiers.org/ena.embl/PRJEB43795. (Wellcome Sanger Institute, 2021)

The genome sequence is released openly for reuse. The *Clostera curtula* 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

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

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

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–10.
PubMed Abstract | Publisher Full Text | Free Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012; [Accessed 26 July 2023].
Publisher Full Text

GBIF Secretariat: ***Clostera curtula* (Linnaeus, 1758).** *GBIF Backbone Taxonomy. Checklist dataset.* 2022; [Accessed 10 July 2023].
Reference Source

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Kimber I: **Chocolate-tip *Clostera curtula* (Linnaeus, 1758).** ukmoths.co.uk. 2023; [Accessed 10 July 2023].
Reference Source

Linnaeus C: **Systema naturae per regna tria naturae, secundum classes, ordines, genera, species cum characteribus, differentiis, synonymis locis.** 10th revised edn., Stockholm: Lars Salvius. 1758; **1**.
Reference Source

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

Percy-Cunningham JE, MacDonald JA: **Biology and Ultrastructure of Sex Pheromone–Producing Glands.** In: *Pheromone Biochemistry.* Elsevier, 1987; 27-75.
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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a; [Accessed 21 July 2023].
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023].
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.
Publisher Full Text

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
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 Chocolate-tip, *Clostera curtula* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB43795, 2021.

Young M: **The importance of Aspen for Lepidoptera.** In: *The Biodiversity and Management of Aspen Woodlands: Proceedings of a one-day conference held in Kingussie, Scotland, on 25th May 2001.* 2001.