DATA NOTE

The genome sequence of the Crescent Plume, *Marasmarcha lunaedactyla* (Haworth, 1811) [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 programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

v1 **First published:** 06 Oct 2023, **8**:431
https://doi.org/10.12688/wellcomeopenres.20107.1
Latest published: 06 Oct 2023, **8**:431
https://doi.org/10.12688/wellcomeopenres.20107.1

Abstract

We present a genome assembly from an individual female *Marasmarcha lunaedactyla* (the Crescent Plume; Arthropoda; Insecta; Lepidoptera; Pterophoridae). The genome sequence is 771.1 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the W and Z sex chromosomes. The mitochondrial genome has also been assembled and is 16.43 kilobases in length. Gene annotation of this assembly on Ensembl identified 21,571 protein coding genes.

Keywords

Marasmarcha lunaedactyla, Crescent Plume, 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) 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, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Crescent Plume, *Marasmarcha lunaedactyla* (Haworth, 1811) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:431 https://doi.org/10.12688/wellcomeopenres.20107.1

First published: 06 Oct 2023, **8**:431 https://doi.org/10.12688/wellcomeopenres.20107.1

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; Pterophoroidea; Pterophoridae; Pterophorinae; Marasmarcha; *Marasmarcha lunaedactyla* (Haworth, 1811) (NCBI:txid1594453).

Background

Marasmarcha lunaedactyla (the Crescent Plume) is a micro-moth in the family Pterophoridae. The species has a southerly distribution in Britain and a scattered distribution in Europe (GBIF Secretariat, 2023). The moth has one generation per year and flies between mid-June to early August. It is disturbed by day but does not usually fly far (Hart, 2011). It flies at night and comes to light. It occurs on downland, quarries and at the coast in locations where its foodplants, rest-harrow (*Ononis* spp), can be found. It lives in small, concentrated colonies, and searching can usually result in finding a number of individuals resting amongst the leaves of its foodplant. Like all the plume moths it rests with its wings outstretched. Its forewing is between 6-8mm in length. There is a distinctive cream crescent at the point that the forewing divides into two lobes. The moth exhibits sexual dimorphism with males having dark brown forewings, and the females lighter brown (Sterling *et al.*, 2012).

Eggs are laid in July, on the underside of the leaves of their foodplant (Langmaid *et al.*, 2018). The larvae hatch in August although they do not grow much before they hibernate, probably in leaf litter at the base of their food plant. The larvae begin feeding again in April, and initially they are difficult to find as they are hidden in the folded terminal leaves. However, the final instar caterpillar is readily found as they are bright green and feed at the top of the rest-harrow stems. As the caterpillars feed at different rates, it is possible to find both small larvae and pupae at the same time (Hart, 2011).

A genome sequence from *M. lunaedactyla* will be useful for comparative studies across the Lepidoptera. The genome of *M. lunaedactyla* 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 *M. lunaedactyla* based on the female specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one female *Marasmarcha lunaedactyla* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.33). A total of 37-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 56-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 8 missing joins or mis-joins and removed one haplotypic

Figure 1. Photograph of the *Marasmarcha lunaedactyla* (ilMarLuna1) specimen used for genome sequencing.

duplications, reducing the scaffold number by 10.42%, and increasing the scaffold N50 by 1.98%.

The final assembly has a total length of 771.1 Mb in 43 sequence scaffolds with a scaffold N50 of 27.1 Mb (Table 1). The snailplot 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 (99.84%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the W and Z sex chromosomes. 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 60.9 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 97.7% (single = 96.8%, duplicated = 0.9%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/1594453.

Genome annotation report

The *Marasmarcha lunaedactyla* genome assembly (GCA_923062675.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Marasmarcha_lunaedactyla_GCA_923062675.1/Info/Index). The resulting annotation includes 21,756 transcribed mRNAs from 21,571 protein-coding genes.

Table 1. **Genome data for *Marasmarcha lunaedactyla*, ilMarLuna1.1.**

Project accession data		
Assembly identifier	ilMarLuna1.1	
Species	*Marasmarcha lunaedactyla*	
Specimen	ilMarLuna1	
NCBI taxonomy ID	1594453	
BioProject	PRJEB47467	
BioSample ID	SAMEA7701293	
Isolate information	ilMarLuna1, female: whole organism (DNA sequencing) ilMarLuna2, whole organism (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	60.9	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:97.7%[S:96.8%,D:0.9%], F:0.6%,M:1.7%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.94%	*≥ 95%*
Sex chromosomes	Z and W	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6808072, ERR6939285	
10X Genomics Illumina	ERR6747941, ERR6747942, ERR6747943, ERR6747944, ERR7732708, ERR7732707	
Hi-C Illumina	ERR6747945	
Genome assembly		
Assembly accession	GCA_923062675.1	
Accession of alternate haplotype	GCA_923062595.1	
Span (Mb)	771.1	
Number of contigs	56	
Contig N50 length (Mb)	25.5	
Number of scaffolds	43	
Scaffold N50 length (Mb)	27.1	
Longest scaffold (Mb)	47.1	
Genome annotation		
Number of protein-coding genes	21,571	
Number of gene transcripts	21,756	

* 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/ilMarLuna1.1/dataset/CAKLPU01.1/busco.

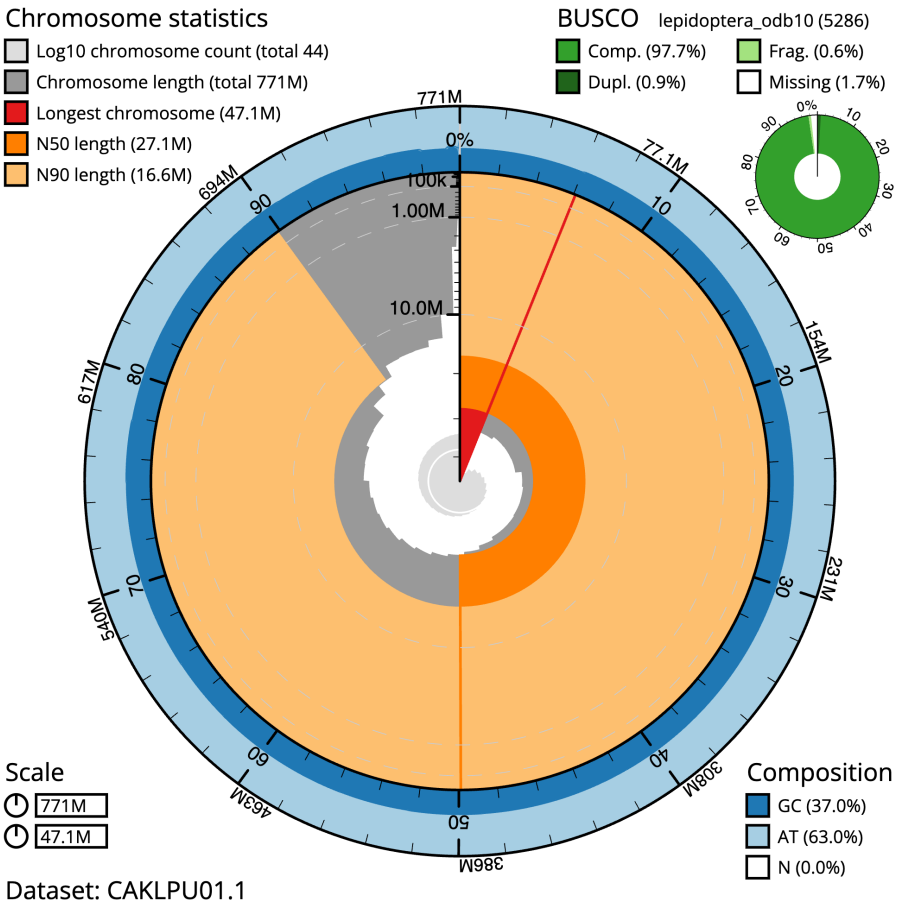

Figure 2. Genome assembly of *Marasmarcha lunaedactyla*, ilMarLuna1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 771,099,907 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (47,084,863 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (27,090,851 and 16,582,527 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/ilMarLuna1.1/dataset/CAKLPU01.1/snail.

Methods

Sample acquisition and nucleic acid extraction

Marasmarcha lunaedactyla specimens were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.33) on 2020-06-25 using a light trap. The specimens were collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice. One specimen (specimen ID Ox000524, individual ilMarLuna1) was used for DNA sequencing, and a second (specimen ID Ox000525, individual ilMarLuna2) was used for Hi-C scaffolding.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilMarLuna1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Tissue from the whole organism was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. 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 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.

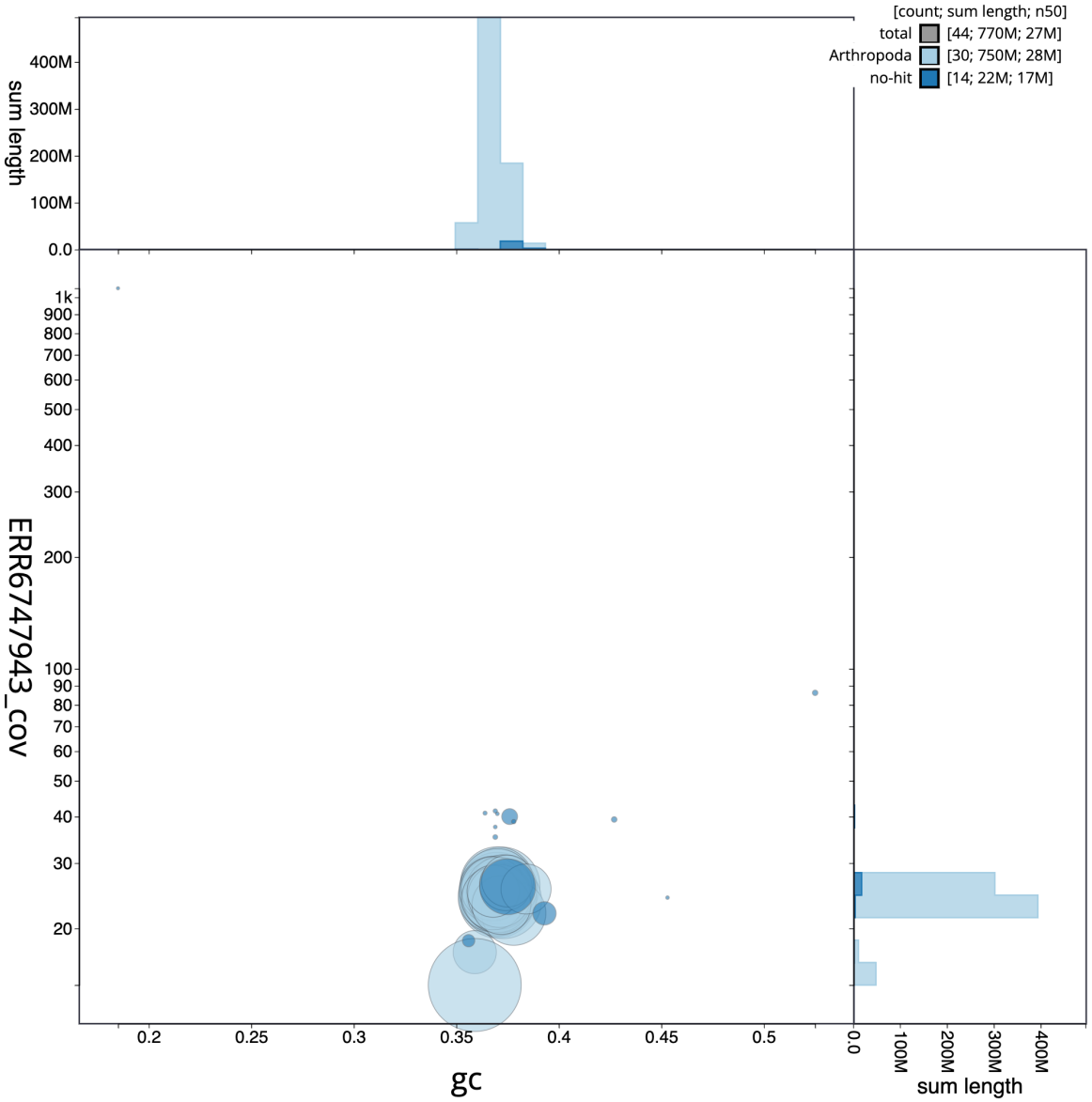

Figure 3. Genome assembly of *Marasmarcha lunaedactyla*, ilMarLuna1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilMarLuna1.1/dataset/CAKLPU01.1/blob.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000, Illumina HiSeq 2500 (10X) instruments. Hi-C data were also generated from whole organism tissue of ilMarLuna2 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). 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 as described previously

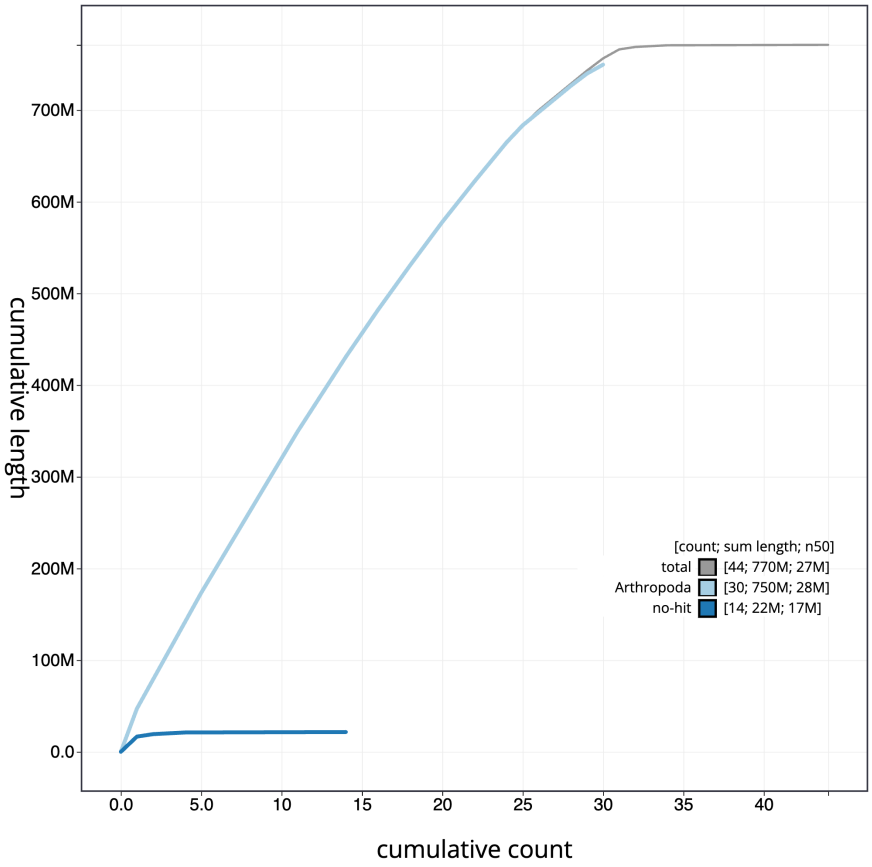

Figure 4. Genome assembly of *Marasmarcha lunaedactyla*, ilMarLuna1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilMarLuna1.1/dataset/CAKLPU01.1/cumulative.

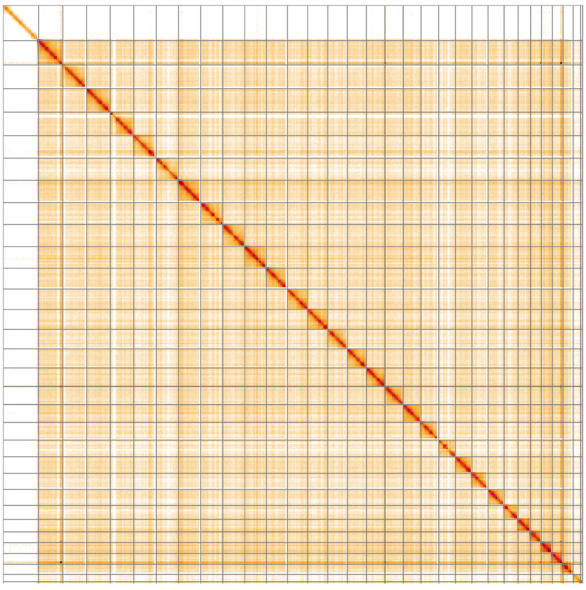

Figure 5. Genome assembly of *Marasmarcha lunaedactyla*, ilMarLuna1.1: Hi-C contact map of the ilMarLuna1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=d_5nq_DvSkylVT7dlPABSQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Marasmarcha lunaedactyla*, ilMarLuna1.

INSDC accession	Chromosome	Length (Mb)	GC%
OV281340.1	1	32.15	37.0
OV281341.1	2	31.45	37.0
OV281342.1	3	31.4	37.0
OV281343.1	4	31.33	37.5
OV281344.1	5	29.79	37.0
OV281345.1	6	29.53	37.0
OV281346.1	7	29.48	37.0
OV281347.1	8	29.3	37.0
OV281348.1	9	29.3	37.0
OV281349.1	10	28.62	37.0
OV281350.1	11	27.76	37.0
OV281351.1	12	27.09	37.0
OV281352.1	13	26.57	37.0
OV281353.1	14	25.88	37.0
OV281354.1	15	25.33	37.0
OV281355.1	16	24.61	37.0
OV281356.1	17	24.11	37.5
OV281357.1	18	23.82	37.0
OV281358.1	19	23.33	37.0
OV281359.1	20	22.16	38.0
OV281360.1	21	22.05	37.0
OV281361.1	22	21.39	37.0
OV281362.1	23	21.27	37.0
OV281363.1	24	18.54	37.0
OV281364.1	25	16.58	37.5
OV281365.1	26	14.47	37.0
OV281366.1	27	14.37	37.5
OV281367.1	28	14.31	37.5
OV281369.1	29	13.27	38.5
OV281368.1	W	9.79	36.0
OV281339.1	Z	47.08	36.0
OV281370.1	MT	0.02	19.0

(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 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 *Marasmarcha lunaedactyla* assembly (GCA_923062675.1) in Ensembl Rapid Release.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission 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)

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	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
Hifiasm	0.15.3	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

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: *Marasmarcha lunaedactyla* (crescent plume). Accession number PRJEB47467; https://identifiers.org/ena.embl/PRJEB47467. (Wellcome Sanger Institute, 2021) The genome sequence is released openly for reuse. The *Marasmarcha lunaedactyla* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789928.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.4783585.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References

Abdennur N, Mirny LA: **Cooler: Scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
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

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.
Publisher Full Text

GBIF Secretariat: ***Marasmarcha lunaedactyla* (Haworth, 1811).** *GBIF Backbone Taxonomy.* 2023; [Accessed 25 June 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

Hart C: **British Plume Moths.** Reading: British Entomological and Natural History Society, 2011.
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

Langmaid JR, Palmer S, Young MR: **A Field Guide to the Smaller Moths of Great Britain and Ireland. 3rd ed.** British Entomological and Natural History Society, 2018.
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

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

Sterling P, Parsons M, Lewington R: **Field Guide to the Micro Moths of Great Britain and Ireland.** Gillingham, Dorset: British Wildlife Publishing, 2012.
Reference Source

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.
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 Crescent Plume, *Marasmarcha lunaedactyla* (Haworth, 1811).** European Nucleotide Archive, [dataset], accession number PRJEB47467, 2021.