DATA NOTE

The genome sequence of the Coxcomb Prominent, *Ptilodon capucinus* (Linnaeus, 1758) [version 1; peer review: 2 approved]

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:** 14 Jun 2023, **8**:242
https://doi.org/10.12688/wellcomeopenres.19539.1
Latest published: 14 Jun 2023, **8**:242
https://doi.org/10.12688/wellcomeopenres.19539.1

Abstract

We present a genome assembly from an individual male *Ptilodon capucinus* (the Coxcomb Prominent; Arthropoda; Insecta; Lepidoptera; Notodontidae). The genome sequence is 348.7 megabases in span. The assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.38 kilobases in length. Gene annotation of this assembly on Ensembl identified 16,968 protein coding genes.

Keywords

Ptilodon capucinus, Coxcomb Prominent, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 14 Jun 2023	✓ view	✓ view

1. **Vlad Dincă**, University of Oulu, Oulu, Finland

2. **James Mallet**, Harvard University, Cambridge, USA

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 Coxcomb Prominent, *Ptilodon capucinus* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:242 https://doi.org/10.12688/wellcomeopenres.19539.1

First published: 14 Jun 2023, **8**:242 https://doi.org/10.12688/wellcomeopenres.19539.1

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; Ptilodontinae; *Ptilodon*; *Ptilodon capucinus* (Linnaeus, 1758) (NCBI:txid987449).

Background

Ptilodon capucinus (Coxcomb Prominent) is a notodontid moth, which is common throughout the Palearctic from Ireland to Japan. It is found across Britain and Ireland but has undergone a significant decline in both distribution and abundance in the last 50 years (Randle *et al.*, 2019) (Randle *et al.*, 2019). It is found in woodlands, scrub, and gardens.

This medium sized moth (forewing length 17–22 mm) varies in colour between light and dark brown (Waring *et al.*, 2017). Like all notodontid moths, it has a small tuft of scales on its back which gives rise to its common name of prominent. It is believed that this tuft breaks up the outline of the moth, affording some protection from predators. Coxcomb refers to the white quiff of scales which decorates its head. This is thought to resemble a form of jester's hat (Marren, 2019), or perhaps the crest of a cockerel. The result of its colouration is that when in the resting position, the wings are folded down, and the moth resembles a dead leaf (Heath & Emmet, 1983).

The moth has two generations a year in the southern part of Britain, flying from April to June, and August to September. In good years there can be two generations in the northern part of its range. The caterpillar is green, with two red projections towards the end of the body. It has an interesting threat response, whereby it curls its head back over its body when alarmed. The larva is polyphagous, eating a wide range of leaves of trees and shrubs. It overwinters as a pupa, often under tree roots (Heath & Emmet, 1983).

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

Genome sequence report

The genome was sequenced from one male *Ptilodon capucinus* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 55-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 121-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected

Figure 1. Photograph of the *Ptilodon capucinus* (ilPtiCapc1) specimen used for genome sequencing.

2 missing joins or mis-joins and removed 1 haplotypic duplication, reducing the scaffold number by 5.88%.

The final assembly has a total length of 348.7 Mb in 31 sequence scaffolds with a scaffold N50 of 12.6 Mb (Table 1). All 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 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 57.9 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.7% (single = 98.4%, duplicated = 0.3%), 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/987449.

Genome annotation report

The *Ptilodon capucinus* genome assembly (GCA_914767695.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Ptilodon_capucinus_GCA_914767695.1/Info/Index). The resulting annotation includes 17,172 transcribed mRNAs from 16,968 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

The specimen selected for genome sequencing was a male *Ptilodon capucinus* (specimen number Ox000813; individual

Table 1. **Genome data for *Ptilodon capucinus*, ilPtiCapc1.1.**

Project accession data		
Assembly identifier	ilPtiCapc1.1	
Species	*Ptilodon capucinus*	
Specimen	ilPtiCapc1	
NCBI taxonomy ID	987449	
BioProject	PRJEB46308	
BioSample ID	SAMEA7746620	
Isolate information	ilPtiCapc1, male: thorax (DNA sequencing); head (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.9	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.7%[S:98.4%,D:0.3%], F:0.3%,M:1.1%,n:5286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	100%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6807995	
10X Genomics Illumina	ERR6464932–ERR6464935	
Hi-C Illumina	ERR6464931	
PolyA RNA-Seq Illumina	ERR9434996	
Genome assembly		
Assembly accession	GCA_914767695.1	
Accession of alternate haplotype	GCA_914767775.1	
Span (Mb)	348.7	
Number of contigs	33	
Contig N50 length (Mb)	12.6	
Number of scaffolds	31	
Scaffold N50 length (Mb)	12.6	
Longest scaffold (Mb)	15.6	
Genome annotation		
Number of protein-coding genes	16,968	
Number of gene transcripts	17,172	

* 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/ilPtiCapc1.1/dataset/ilPtiCapc1_1.1/busco.

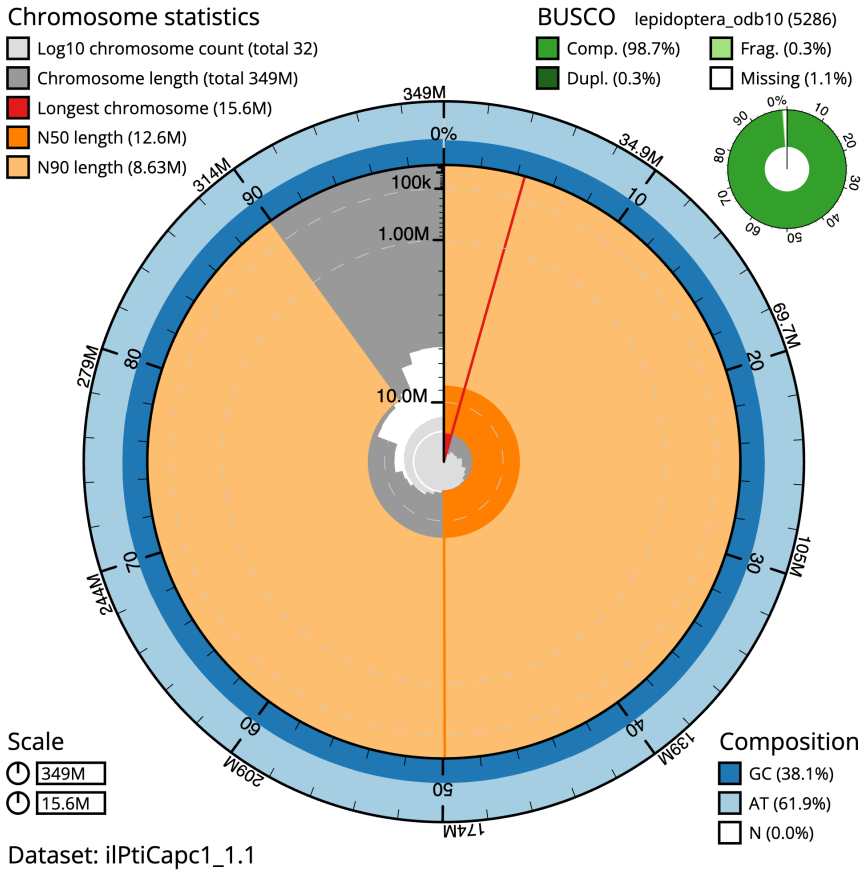

Figure 2. Genome assembly of *Ptilodon capucinus*, ilPtiCapc1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 348,711,871 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (15,614,753 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,643,213 and 8,629,352 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/ilPtiCapc1.1/dataset/ilPtiCapc1_1.1/snail.

ilPtiCapc1) collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-08-01. The specimen was taken from a woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The specimen was identified by the collector and then preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilPtiCapc1 specimen was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Thorax tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. 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.

RNA was extracted from abdomen tissue of ilPtiCapc1 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

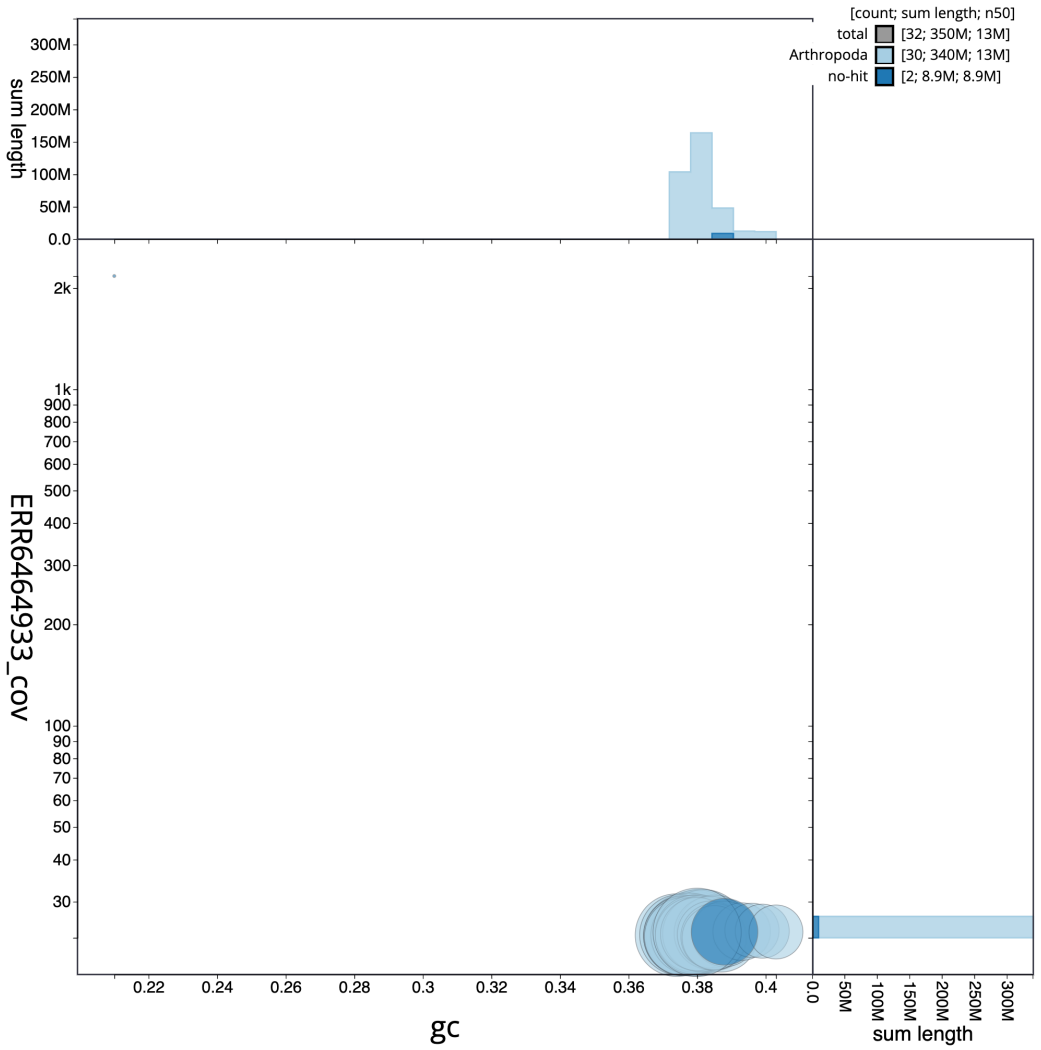

Figure 3. Genome assembly of *Ptilodon capucinus*, ilPtiCapc1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilPtiCapc1.1/dataset/ilPtiCapc1_1.1/blob.

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 tissue of ilPtiCapc1 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 (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.*, 2022), 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.

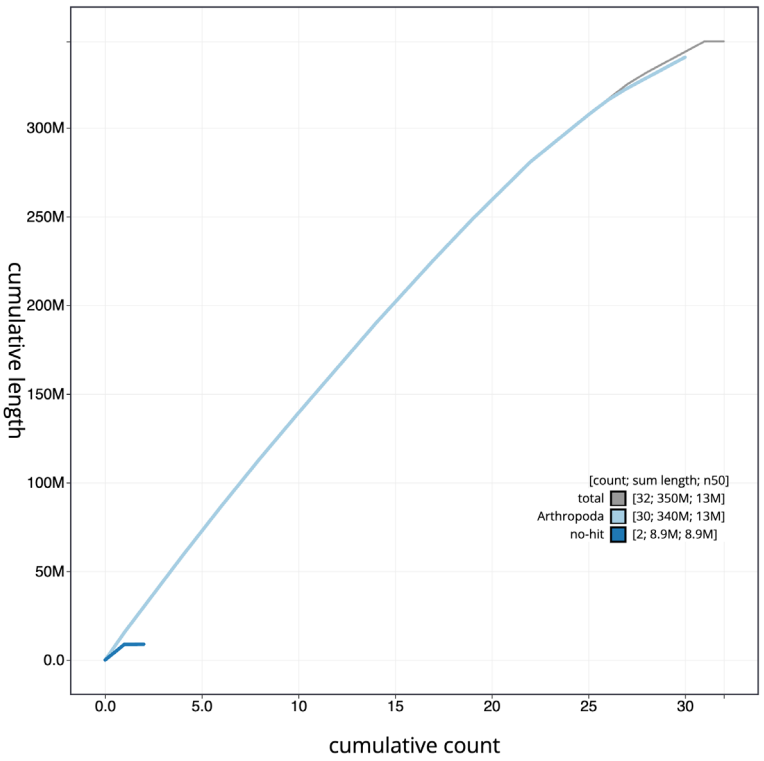

Figure 4. Genome assembly of *Ptilodon capucinus*, ilPtiCapc1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilPtiCapc1.1/dataset/ilPtiCapc1_1.1/cumulative.

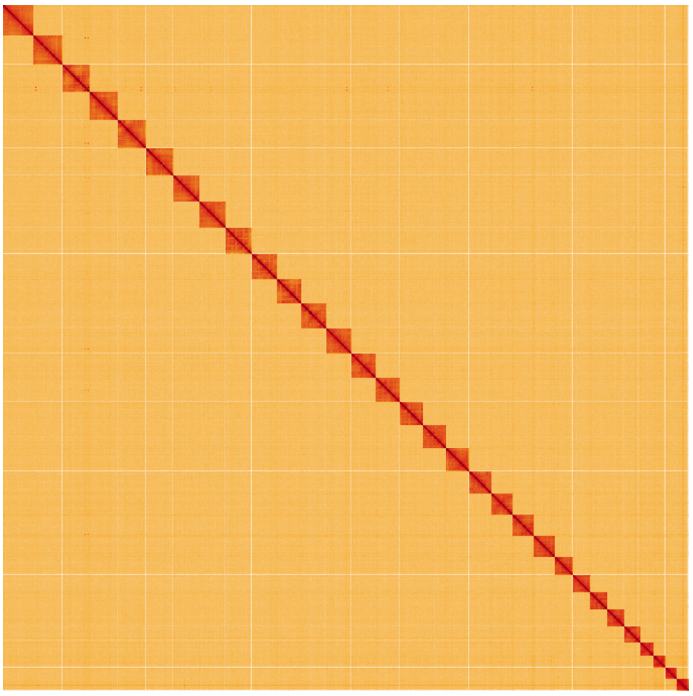

Figure 5. Genome assembly of *Ptilodon capucinus*, ilPtiCapc1.1 alternate haplotype: Hi-C contact map of the ilPtiCapc1.1 alternate haplotype assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=BqET2dZSTnSs7zhjHqvgWg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Ptilodon capucinus*, ilPtiCapc1.

INSDC	Chromosome	Size (Mb)	GC%
OU611808.1	1	14.62	37.9
OU611809.1	2	14.27	38.1
OU611810.1	3	14.25	38
OU611811.1	4	14.13	38.2
OU611812.1	5	13.69	37.4
OU611813.1	6	13.67	37.7
OU611814.1	7	13.22	37.8
OU611815.1	8	13.05	37.5
OU611816.1	9	12.78	37.6
OU611817.1	10	12.72	38
OU611818.1	11	12.7	37.6
OU611819.1	12	12.64	37.9
OU611820.1	13	12.45	37.8
OU611821.1	14	12.15	37.6
OU611822.1	15	11.83	37.9
OU611823.1	16	11.8	37.9
OU611824.1	17	11.54	38.1
OU611825.1	18	11.4	38.5
OU611826.1	19	10.78	38.7
OU611827.1	20	10.77	38.4
OU611828.1	21	10.74	38
OU611829.1	22	8.99	38.4
OU611830.1	23	8.92	38.5
OU611831.1	24	8.88	38.8
OU611832.1	25	8.63	38.8
OU611833.1	26	8.26	38.5
OU611834.1	27	6.64	39.3
OU611835.1	28	5.88	39.6
OU611836.1	29	5.84	39.9
OU611837.1	30	5.82	40.3
OU611807.1	Z	15.61	38
OU611838.1	MT	0.02	21.2

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 *Ptilodon capucinus* assembly (GCA_914767695.1). 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

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

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: *Ptilodon capucinus* (coxcomb prominent). Accession number PRJEB46308; https://identifiers.org/ena.embl/PRJEB46308. (Wellcome Sanger Institute, 2021)

The genome sequence is released openly for reuse. The *Ptilodon capucinus* 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

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

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

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

Marren P: **Emperors, Admirals, and Chimney Sweepers**. Toller Fratrum,

Dorset: Little Toller, 2019.
Reference Source

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths**. Newbury: NatureBureau, 2019.
Reference Source

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: 17 April 2023).
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev)**. *Zenodo*. 2023b; (Accessed: 17 April 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.** *bioRxiv.* [Preprint], 2022.
Publisher Full Text

Vasimuddin Md, 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 Coxcomb Prominent, *Ptilodon capucinus* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB46308, 2022.

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 21 May 2024

https://doi.org/10.21956/wellcomeopenres.21643.r74963

✔ **James Mallet**
Harvard University, Cambridge, Massachusetts, USA

There's nothing much to say. The paper reports a genome assembly of the relevant species. Most of the reads mapped to "31 chromosomal pseudomolecules," as they should, given that the mode of chromosome numbers in the Lepidoptera is 31. The mitochondrial genome was also assembled successfully. As far as I can see the identity of the chromosomes was not verified against those of other species. The Z chromosome was reportedly identified as OU611807.1, but it is not explained how, since the sequence data was from a male specimen, and the Z would therefore normally be at the same coverage as autosomes. However, this seems typical for the work in this series of papers, and the chromosomes would certainly be identifiable via comparative study in the future.

Overall, an impressive and successful assembly which will be useful for further and ongoing work on the comparative genomics of Lepidoptera.

Is the rationale for creating the dataset(s) clearly described?
Partly

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative genomics, evolutionary biology of the Lepidoptera

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 14 March 2024

https://doi.org/10.21956/wellcomeopenres.21643.r74958

Vlad Dincă
University of Oulu, Oulu, Finland

In this study, the authors have successfully generated a chromosome-level genome assembly of a male *Ptilodon capucinus* (Lepidoptera, Notodontidae). The mitochondrial genome has been assembled as well.

Although I am not an expert in several of the methodological approaches used, as far as I can tell, the manuscript is technically sound and uses methodologies and pipelines that are fairly well-established through use by the Sanger Institute.

The genome appears to be of high quality and has a BUSCO v5.3.2 completeness of 98.7%.
It is a pity that the W sex chromosome is lacking (a male was sequenced).

Reference genomes such as this one represent a valuable resource for the scientific community and each new addition provides new opportunities for research.

A few additional comments are included below.

Numerous sources list the species as "*Ptilodon capucina*". Although this study is obviously not focused on nomenclatural aspects, I wonder whether a short explanation (in the Background section) of this name discrepancy is necessary (perhaps it is a case of gender agreement issue, which is the subject of notable debate in some groups – e.g. European butterflies). Briefly addressing this may be useful given that we are dealing with the first reference genome for this species and it is good to avoid any potential ambiguities, even if slight.

I see Ptilodontinae is listed as subfamily. However, various sources seem to list this species under Notodontinae. I tried to find more information about this, but I was not able to find much, at least not based on DNA data. What I could find does not seem to provide strong support for Ptilodontinae, although these studies used either limited DNA data (Kobayashi & Nonaka 2016), or were not focused on this issue (St Laurent R, et al. 2023 [Ref 1]).
It is possible that I may have missed some key study (Notodontidae are not my main area of taxonomic expertise). Nevertheless, I wanted to mention this subfamily aspect that may also need a short clarification/mention in the Background section.

I wonder how much is known in terms of genetic structure for *P. capucinus*. Even if limited data is

available (e.g. mtDNA, a few nDNA markers etc), if any notable aspect is known (e.g. diverged lineages), it could be worth briefly mentioning in the Background.

Finally, it may also be good to mention whether *Wolbachia* has been detected in the sample analysed (I assume not, but it may be useful to mention that).
https://www.academia.edu/32479616/Molecular_phylogeny_of_the_Notodontidae_Subfamilies_inferred_from_2

References

Kobayashi, H., Nonaka, M. (2016) Molecular phylogeny of the Notodontidae: subfamilies inferred from 28S rRNA sequences (Lepidoptera, Noctuoidea, Notodontidae). *Tinea*, 23, 1–83.

References

1. St Laurent R, Goldstein P, Miller J, Markee A, et al.: Phylogenetic systematics, diversification, and biogeography of Cerurinae (Lepidoptera: Notodontidae) and a description of a new genus. *Insect Systematics and Diversity*. 2023; **7** (2). Publisher Full Text

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Lepidoptera speciation, taxonomy and phylogeography, with a focus on Western Palearctic butterflies.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.