DATA NOTE

The genome sequence of the Yellow Shell moth, *Camptogramma bilineatum* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Liam M. Crowley[2], Finley Hutchinson[3], 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]Department of Biology, University of Oxford, Oxford, England, UK
[3]University of Exeter, Exeter, England, UK

[+] Deceased author

v1 **First published:** 20 May 2024, **9**:277
https://doi.org/10.12688/wellcomeopenres.21628.1
Latest published: 20 May 2024, **9**:277
https://doi.org/10.12688/wellcomeopenres.21628.1

Abstract

We present a genome assembly from an individual male *Camptogramma bilineatum* (the Yellow Shell; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 442.7 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 17.04 kilobases in length.

Keywords

Camptogramma bilineatum, Yellow Shell 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; **Crowley LM**: Investigation, Resources; **Hutchinson F**: 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, 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, Crowley LM, Hutchinson F *et al.* **The genome sequence of the Yellow Shell moth, *Camptogramma bilineatum* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:277 https://doi.org/10.12688/wellcomeopenres.21628.1

First published: 20 May 2024, **9**:277 https://doi.org/10.12688/wellcomeopenres.21628.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; *Camptogramma*; *Camptogramma bilineatum* (Linnaeus, 1758) (NCBI:txid934908).

Background

The Yellow Shell *Camptogramma bilineata* is a moth in the family Geometridae. It is distinctive, having yellow or yellowish-orange forewings and hindwings with a large number of wavy brown crosslines. It is found in a variety of colour morphotypes. The darker forms, often with brownish bands, are commoner in the British Isles away from the south-east of England (Skinner & Wilson, 2009; Waring *et al.*, 2017). The forms may also vary in size, for example, the one found on the Shetland Isles, *C. bilineata atlantica*, is smaller and darker, than those found further south (Skinner & Wilson, 2009).

Camptogramma bilineata has one brood a year with adults on the wing from June to August in the United Kingdom and Ireland. It can be found in a large range of habitats (Waring *et al.*, 2017), but unlike many of the other moth species in a study in Finland, *C. bilineata* prefers older continuously grazed pastures to those that have been abandoned (Pöyry *et al.*, 2005). This species overwinters as a larva, with larvae found from July until May. They feed on a range of plants including cleavers *Galium aparine* and other bedstraws *Galium* spp., wormwoods *Artemisia* spp., docks and sorrels *Rumex* spp., dandelions *Taraxacum* spp. and common chickweed *Stellaria media* (Skinner & Wilson, 2009; Waring *et al.*, 2017).

We present a chromosomal-level complete genome sequence for a male *Camptogramma bilineata,* based on one specimen collected in Wytham Woods, Oxfordshire, as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from a male *Camptogramma bilineatum* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 33-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 removed one haplotypic duplication.

The final assembly has a total length of 442.7 Mb in 37 sequence scaffolds with a scaffold N50 of 16.8 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.91%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome.

Figure 1. Photograph of the *Camptogramma bilineatum* (ilCamBiln1) specimen used for genome sequencing.

Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). Chromosome Z was assigned by synteny to *Horisme vitalbata* (GCA_951804965.1). 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 66.4 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.4% (single = 98.2%, duplicated = 0.3%), 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/934908.

Methods

Sample acquisition and nucleic acid extraction

A male *Camptogramma bilineatum* (specimen ID Ox000667, ToLID ilCamBiln1) was 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 preserved on dry ice.

The specimen used for Hi-C sequencing (specimen ID Ox003075, ToLID ilCamBiln2) was collected from the same location on 2022-07-22, also using a light trap. The specimen was collected by Liam Crowley (University of Oxford) and Finley Hutchinson (University of Exeter) and identified by Finley Hutchinson and preserved on dry ice.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core

Table 1. Genome data for *Camptogramma bilineatum*, ilCamBiln1.1.

Project accession data		
Assembly identifier	ilCamBiln1.1	
Species	*Camptogramma bilineatum*	
Specimen	ilCamBiln1	
NCBI taxonomy ID	934908	
BioProject	PRJEB55948	
BioSample ID	SAMEA7701528	
Isolate information	ilCamBiln1, male: whole organism (DNA sequencing) ilCamBiln2: abdomen (Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	66.4	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:98.4%[S:98.2%,D:0.3%], F:0.3%,M:1.3%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.91%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 17.04 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences Sequel IIe	ERR10224899	
Hi-C Illumina	ERR11679368	
Genome assembly		
Assembly accession	GCA_958496255.1	
Accession of alternate haplotype	GCA_958496295.1	
Span (Mb)	442.7	
Number of contigs	47	
Contig N50 length (Mb)	16.3	
Number of scaffolds	37	
Scaffold N50 length (Mb)	16.8	
Longest scaffold (Mb)	22.43	

* 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/ilCamBiln1_1/dataset/ilCamBiln1_1/busco.

Laboratory includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilCamBiln1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). 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

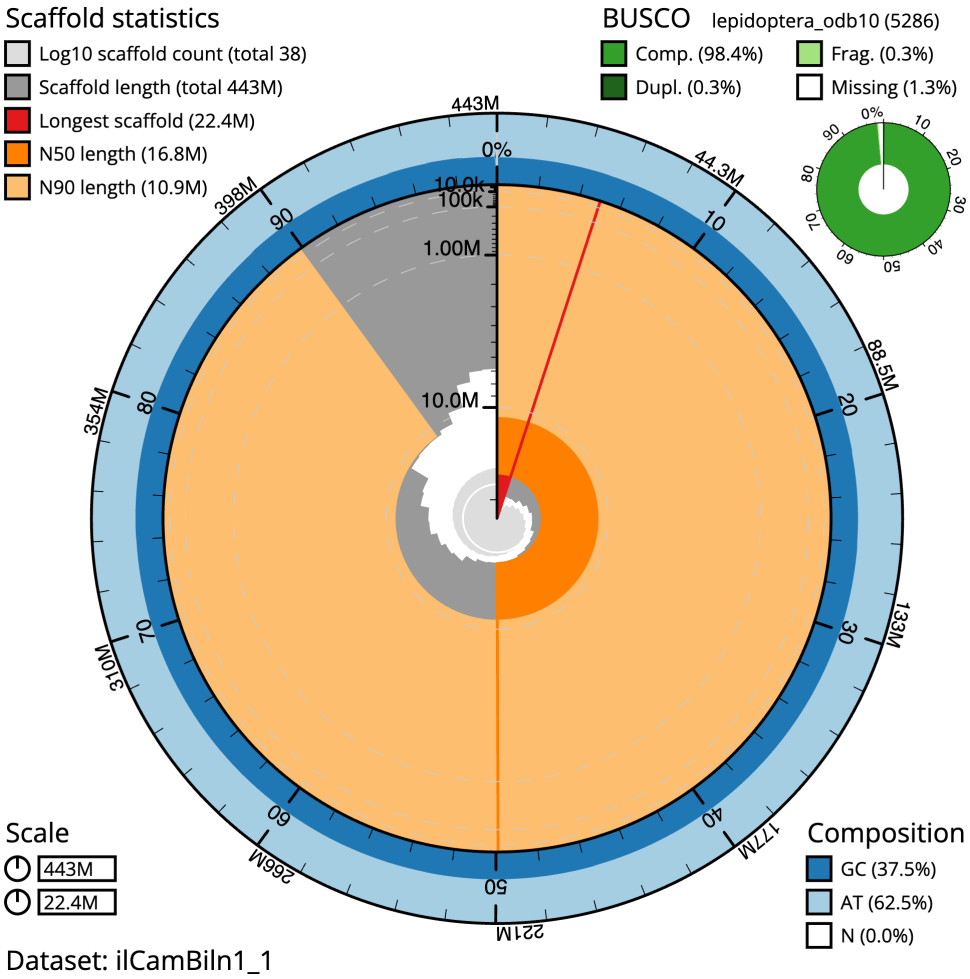

Figure 2. Genome assembly of *Camptogramma bilineatum*, ilCamBiln1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 442,695,304 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (22,426,387 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (16,753,518 and 10,868,945 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/ilCamBiln1_1/dataset/ilCamBiln1_1/snail.

ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

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

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 a Pacific Biosciences Sequel IIe instrument. Hi-C data were also generated from abdomen tissue of ilCamBiln2 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly and curation

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

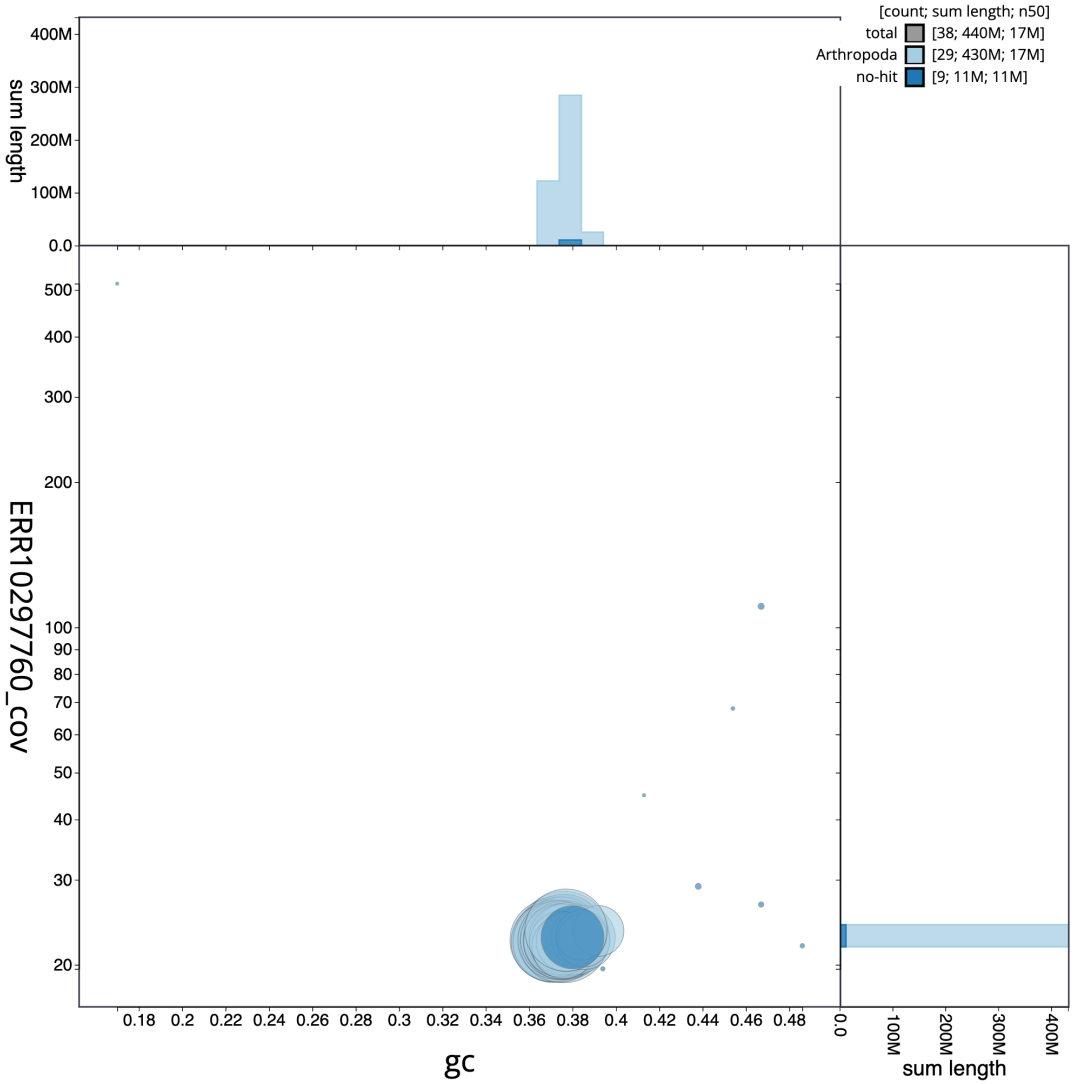

Figure 3. Genome assembly of *Camptogramma bilineatum*, ilCamBiln1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilCamBiln1_1/dataset/ilCamBiln1_1/blob.

(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 (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.

Evaluation of the final assembly

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

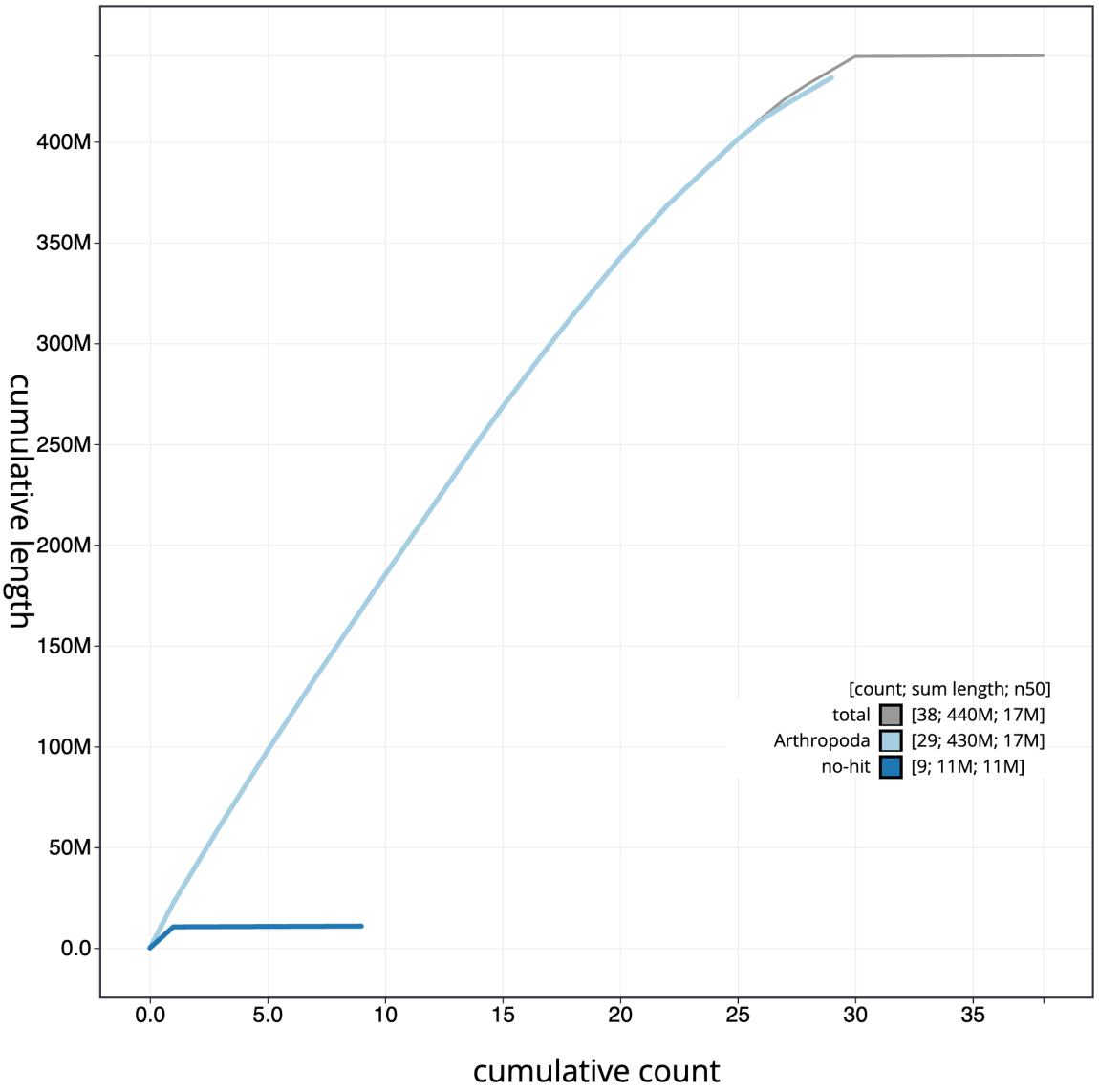

Figure 4. Genome assembly of *Camptogramma bilineatum*, ilCamBiln1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilCamBiln1_1/dataset/ilCamBiln1_1/cumulative.

'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),

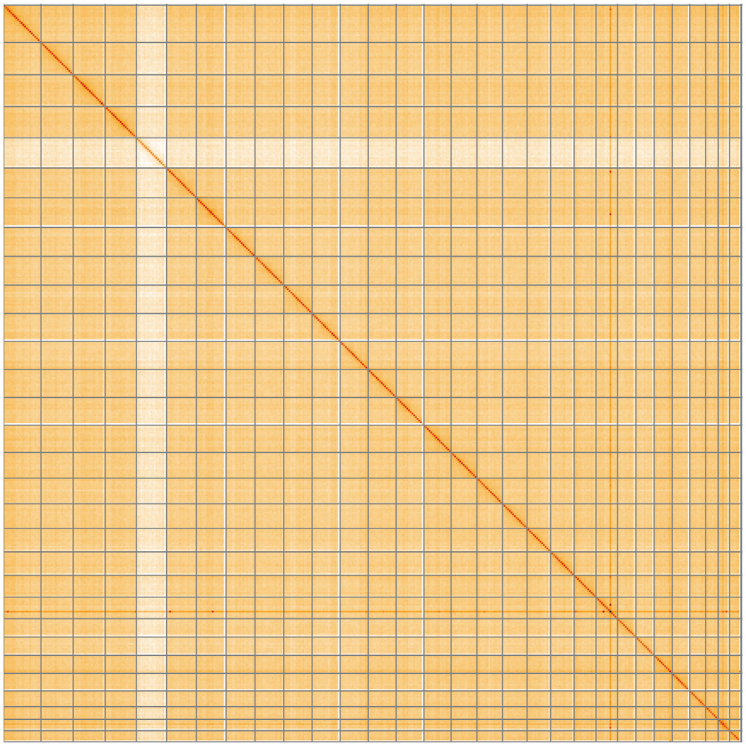

Figure 5. Genome assembly of *Camptogramma bilineatum*, ilCamBiln1.1: Hi-C contact map of the ilCamBiln1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Sulp9mTWS2SK0paaX5ODgw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Camptogramma bilineatum*, ilCamBiln1.

INSDC accession	Chromosome	Length (Mb)	GC%
OY292450.1	1	22.43	37.5
OY292451.1	2	19.3	37.5
OY292452.1	3	19.2	37.5
OY292453.1	4	18.66	37.5
OY292455.1	5	17.86	37.0
OY292456.1	6	17.82	37.5
OY292457.1	7	17.37	37.0
OY292458.1	8	17.22	37.0
OY292459.1	9	17.03	37.0
OY292460.1	10	16.85	37.0
OY292461.1	11	16.84	37.5
OY292462.1	12	16.75	37.5
OY292463.1	13	16.66	37.5
OY292464.1	14	16.3	37.5
OY292465.1	15	15.52	37.5
OY292466.1	16	15.28	37.5
OY292467.1	17	14.8	37.5
OY292468.1	18	14.14	37.5
OY292469.1	19	14.06	38.0
OY292470.1	20	13.14	38.0
OY292471.1	21	12.96	37.5
OY292472.1	22	11.01	37.5
OY292473.1	23	10.91	37.5
OY292474.1	24	10.87	38.5
OY292475.1	25	10.52	38.0
OY292476.1	26	9.48	37.5
OY292477.1	27	7.58	38.5
OY292478.1	28	6.84	39.0
OY292479.1	29	6.78	38.5
OY292454.1	Z	18.17	37.5
OY292480.1	MT	0.02	17.5

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	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	1.2a.2	https://github.com/c-zhou/yahs

and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Camptogramma bilineatum* (yellow shell). Accession number PRJEB55948; https://identifiers.org/ena.embl/PRJEB55948 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Camptogramma bilineatum* 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

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**

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
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**

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**

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.
Publisher Full Text

Pöyry J, Lindgren S, Salminen J, *et al.*: **Responses of butterfly and moth species to restored cattle grazing in semi-natural grasslands.** *Biol Conserv.* 2005; **122**(3): 465–478.
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**

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**

Skinner B, Wilson D: **Colour identification guide to moths of the British Isles.** First edit. Stenstrup, Denmark: Apollo Books, 2009.
Reference Source

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 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 Yellow Shell moth, *Camptogramma bilineatum* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB55948, 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**