DATA NOTE

The genome sequence of the Crescent Groundling, *Teleiodes luculella* (Hübner, 1813) [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 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]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

v1 **First published:** 11 Mar 2024, **9**:143
https://doi.org/10.12688/wellcomeopenres.21090.1
Latest published: 11 Mar 2024, **9**:143
https://doi.org/10.12688/wellcomeopenres.21090.1

Abstract

We present a genome assembly from an individual male *Teleiodes luculella* (the Crescent Groundling; Arthropoda; Insecta; Lepidoptera; Gelechiidae). The genome sequence is 454.5 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 15.32 kilobases in length. Gene annotation of this assembly on Ensembl identified 19,943 protein coding genes.

Keywords

Teleiodes luculella, crescent groundling, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 11 Mar 2024	✓ view	✓ view

1. **Wai Lok So**, The Chinese University of Hong Kong, Hong Kong, Hong Kong

2. **Axel Künstner**, University of Lübeck, Lübeck, Germany

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 Crescent Groundling, *Teleiodes luculella* (Hübner, 1813) [version 1; peer review: 2 approved]** Wellcome Open Research 2024, **9**:143 https://doi.org/10.12688/wellcomeopenres.21090.1

First published: 11 Mar 2024, **9**:143 https://doi.org/10.12688/wellcomeopenres.21090.1

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; Gelechioidea; Gelechiidae; Gelechiinae; *Teleiodes*; *Teleiodes luculella* (Hübner, 1813) (NCBI:txid1101163).

Background

Teleiodes luculella (Crescent Groundling) is a micro-moth in the family Gelechiidae, and is found throughout Europe (GBIF Secretariat, 2024). In the UK it is fairly common in oak woods in England and Wales. The moth is small with a forewing length of 5–6 mm. The forewing is dark grey with a semi-circular white costal blotch, often with an orangey-yellow mark, and smaller white costal spots about two-thirds of the way along the wing (Emmet & Langmaid, 2002).

The moth is thought to be single-brooded, flying between May and August Click or tap here to enter text. (Sterling *et al.*, 2012). However, there is some suggestion that the moth may be double-brooded in southern parts of its European range (Palmer & Palmer, 2023). The larvae feed on oak in spun leaves and pupate in leaf litter on the ground (Emmet & Langmaid, 2002).

The genome of *T. luculella* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in Britain and Ireland.

Genome sequence report

The genome was sequenced from one male *Teleiodes luculella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 35-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 324 missing joins or mis-joins and removed 19 haplotypic duplications, reducing the assembly length by 0.59% and the scaffold number by 7.53%, and decreasing the scaffold N50 by 0.78%.

Figure 1. Photograph of the *Teleiodes luculella* (ilTelLucu1) specimen used for genome sequencing.

The final assembly has a total length of 454.5 Mb in 527 sequence scaffolds with a scaffold N50 of 15.3 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 (95.93%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. 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 57.6 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 94.8% (single = 92.1%, duplicated = 2.7%), 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/1101163.

Genome annotation report

The *Teleiodes luculella* genome assembly (GCA_948473455.1) was annotated on Ensembl Rapid Release at the European Bioinformatics Institute (EBI). The resulting annotation includes 20,151 transcribed mRNAs from 19,943 protein-coding genes (Table 1; https://rapid.ensembl.org/Teleiodes_luculella_GCA_948473455.1/Info/Index).

Methods

Sample acquisition and nucleic acid extraction

Teleiodes luculella specimens were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2021-06-16 using a light trap. The specimens were collected and identified by Douglas Boyes (University of Oxford). The specimens were snap-frozen on dry ice. One specimen was used for DNA sequencing (specimen ID Ox001931, ToLID ilTelLucu1) and a second for Hi-C sequencing (specimen ID Ox001932, ToLID ilTelLucu2).

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilTelLucu1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue of the whole organism

Table 1. Genome data for *Teleiodes luculella*, ilTelLucu1.1.

Project accession data		
Assembly identifier	ilTelLucu1.1	
Species	*Teleiodes luculella*	
Specimen	ilTelLucu1	
NCBI taxonomy ID	1101163	
BioProject	PRJEB58075	
BioSample ID	SAMEA10979195	
Isolate information	ilTelLucu1 (DNA sequencing) ilTelLucu2 (Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.6	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:94.8%[S:92.1%,D:2.7%], F:1.0%,M:4.2%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	95.93%	*≥ 95%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.32 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10662023	
Hi-C Illumina	ERR10659247	
Genome assembly		
Assembly accession	GCA_948473455.1	
Accession of alternate haplotype	GCA_948473415.1	
Span (Mb)	454.5	
Number of contigs	1,914	
Contig N50 length (Mb)	0.5	
Number of scaffolds	527	
Scaffold N50 length (Mb)	15.3	
Longest scaffold (Mb)	30.47	
Genome annotation		
Number of protein-coding genes	19,943	
Number of gene transcripts	20,151	

* 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/CAOKYR01/dataset/CAOKYR01/busco.

was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol

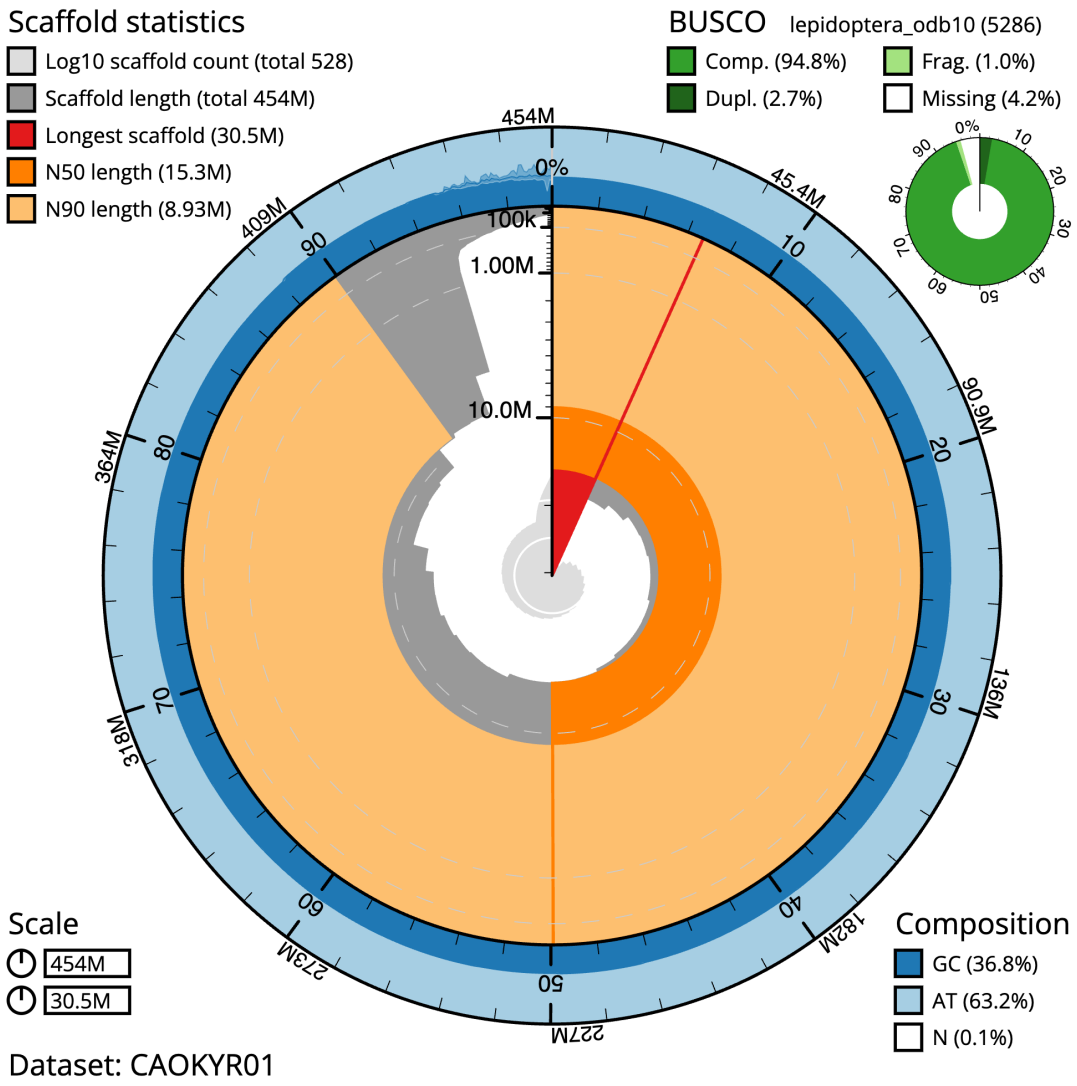

Figure 2. Genome assembly of *Teleiodes luculella*, ilTelLucu1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 454,474,787 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (30,471,558 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (15,281,747 and 8,929,433 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/CAOKYR01/dataset/CAOKYR01/snail.

(Oatley *et al.*, 2023). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 31 (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

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'

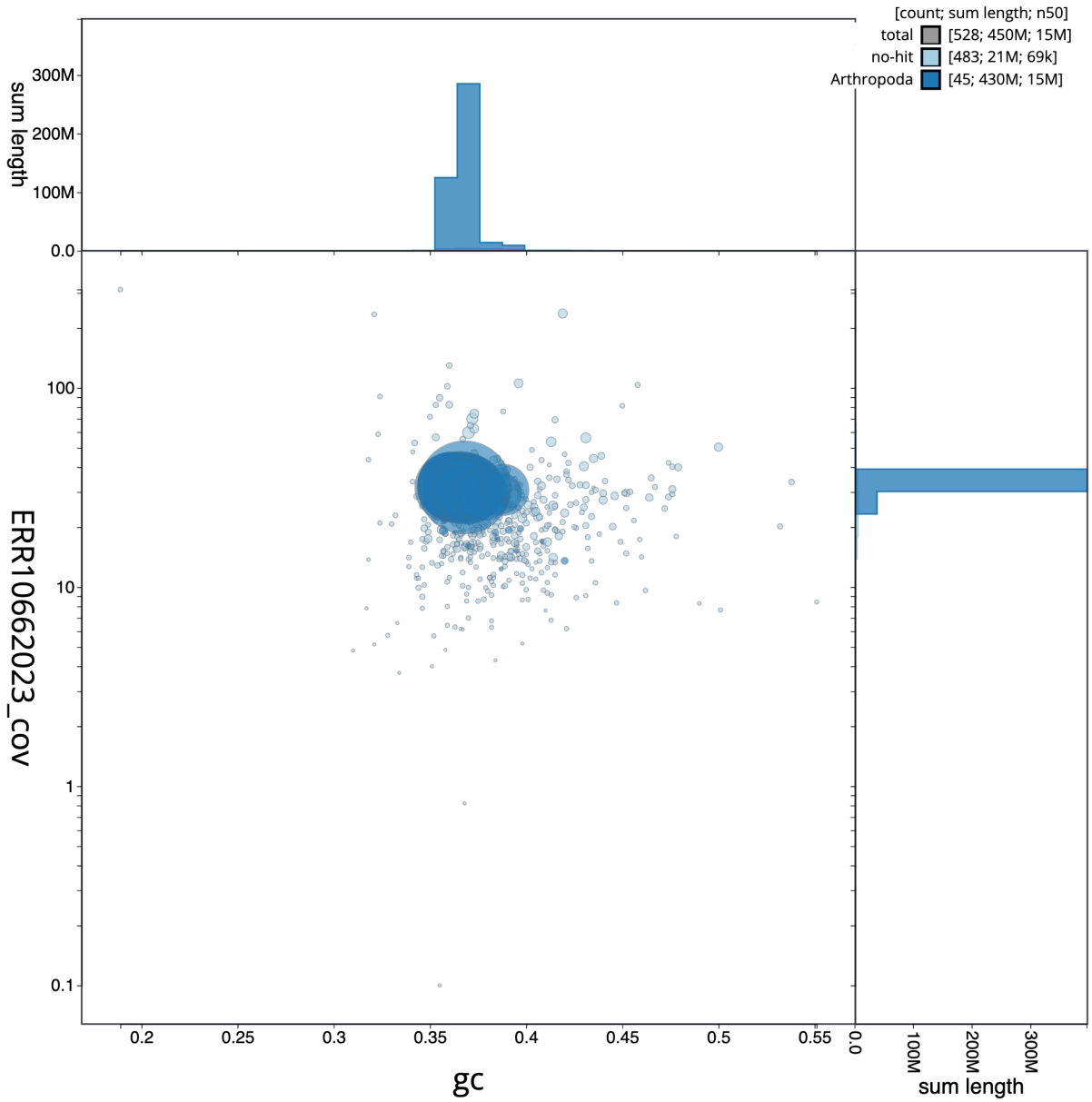

Figure 3. Genome assembly of *Teleiodes luculella*, ilTelLucu1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAOKYR01/dataset/CAOKYR01/blob.

instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL II instrumens. Hi-C data were also generated from the whole organism of ilTelLucu2 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (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

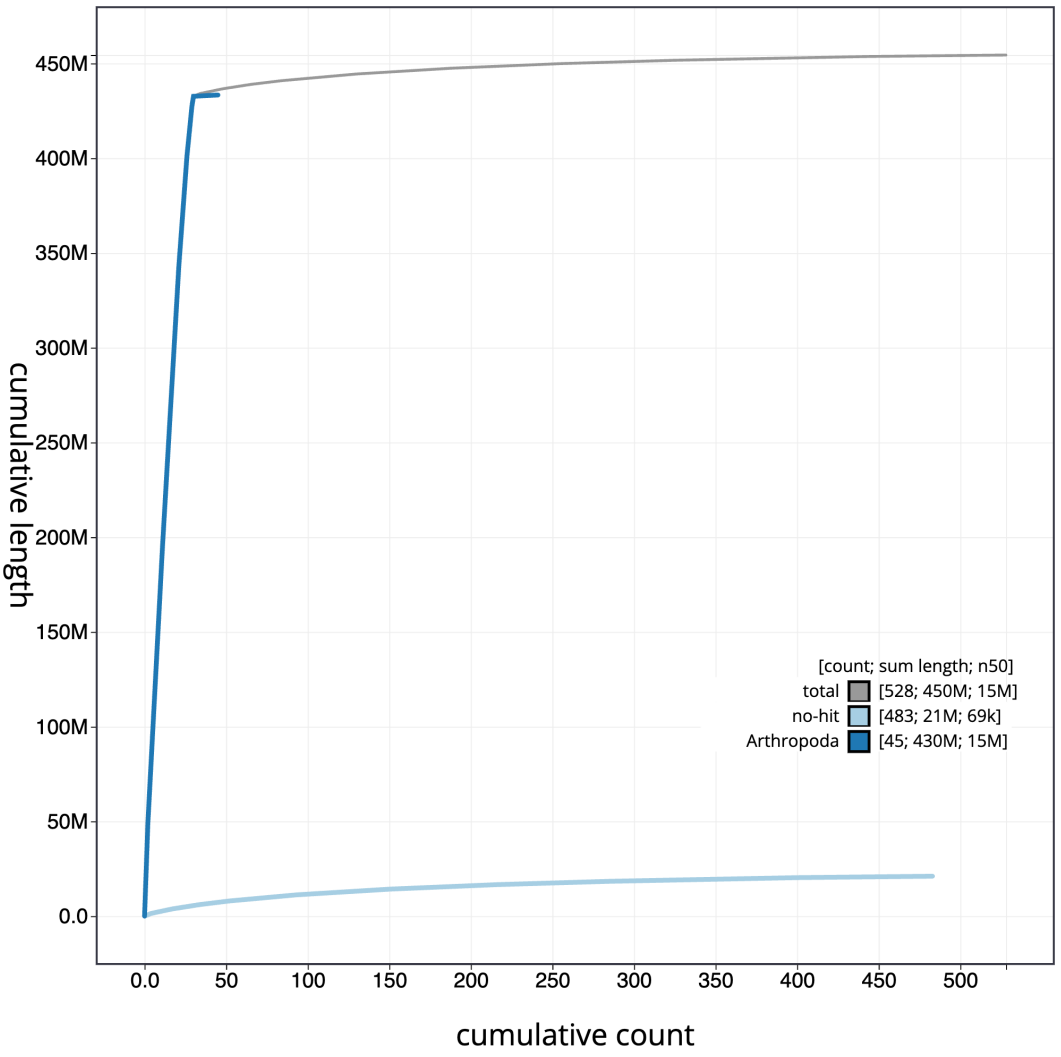

Figure 4. Genome assembly of *Teleiodes luculella*, ilTelLucu1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAOKYR01/dataset/CAOKYR01/cumulative.

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 *Teleiodes luculella* assembly (GCA_948473455.1) in Ensembl Rapid Release at the EBI.

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**

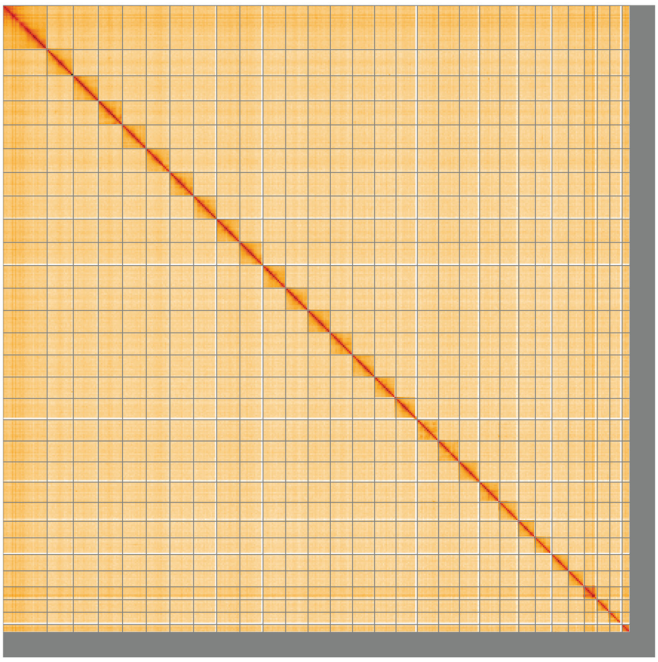

Figure 5. Genome assembly of *Teleiodes luculella*, ilTelLucu1.1: Hi-C contact map of the ilTelLucu1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=ZLnB8ozxSOygY9xC6AzPNg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Teleiodes luculella*, ilTelLucu1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX419583.1	1	18.05	36.5
OX419584.1	2	17.32	37.0
OX419585.1	3	16.3	36.5
OX419586.1	4	16.59	36.5
OX419587.1	5	16.48	36.5
OX419588.1	6	16.29	36.0
OX419589.1	7	16.0	36.0
OX419590.1	8	15.98	37.0
OX419591.1	9	15.98	36.5
OX419592.1	10	15.71	36.0
OX419593.1	11	15.28	36.5
OX419594.1	12	15.4	36.0
OX419595.1	13	15.37	36.5
OX419596.1	14	15.28	36.0
OX419597.1	15	14.74	36.5
OX419598.1	16	14.8	36.5
OX419599.1	17	14.62	37.0
OX419600.1	18	14.35	37.0
OX419601.1	19	13.99	37.0
OX419602.1	20	13.96	37.0
OX419603.1	21	13.08	37.5
OX419604.1	22	11.32	37.0
OX419605.1	23	11.46	37.0
OX419606.1	24	11.42	37.0
OX419607.1	25	11.06	36.5
OX419608.1	26	8.58	37.5
OX419609.1	27	8.93	39.0
OX419610.1	28	8.51	38.0
OX419611.1	29	5.38	38.5
OX419582.1	Z	30.47	37.0
OX419612.1	MT	0.02	19.0

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	1.2a	https://github.com/c-zhou/yahs

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: *Teleiodes luculella* (crescent groundling). Accession number PRJEB58075; https://identifiers.org/ena.embl/PRJEB58075 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Teleiodes luculella* 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.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

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor® 3 for LI PacBio.** *Protocols.Io.* 2023.
Publisher 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

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

Emmet AM, Langmaid JR: **The Moths and Butterflies of Great Britain and Ireland.** Gelechiidae, Colchester: Harley Books, 2002.
Reference Source

GBIF Secretariat: ***Teleiodes luculella* (Hübner, 1813).** GBIF Backbone Taxonomy, 2024; [Accessed 4 February 2024].
Reference Source

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

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2.** *Protocols.Io.* 2023.
Publisher Full Text

Palmer C, Palmer S: **Teleiodes luculella, Gelechiid Recording Scheme.** 2023. [Accessed 4 February 2024].
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

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

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

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 Grounding, *Teleiodes luculella* (Hübner, 1813).** European Nucleotide Archive. [dataset], accession number PRJEB58075, 2023.

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* Edited by C. Alkan, 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 12 August 2024

https://doi.org/10.21956/wellcomeopenres.23330.r90454

✔ **Axel Künstner**
University of Lübeck, Lübeck, Germany

The manuscript presents the genome assembly of the Crescent Groundling, *Teleiodes luculella*. The manuscript exhibits clear writing and sound methodologies. The raw data for this study was submitted to ENA.

Besides one minor comment (see below), I have no further complaints.

Minor comment:

In the Background section, there is the following unassociated sentence: 'Click or tap here to enter 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: Bioinformatics, Cancer research, Microbiome

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 07 August 2024

https://doi.org/10.21956/wellcomeopenres.23330.r90450

Wai Lok So
School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong

The authors sequenced the Crescent Groundling, *Teleiodes luculella,* yielding a high-quality chromosomal-level genome, together with the sex-chromosome and the mitochondrial genome. The methods are standard, sound, and logical. Sufficient details have been given regarding sample preparation, genome sequencing, and assembly. The genome and other associated materials have been properly deposited, which is accessible and reusable for future users.

Though the genome sequencing part is well done, I would like to see a bit more background on this moth species, especially its special biology or ecological significance. The inclusion of additional biological details of this species in the "Background" would be greatly appreciated.

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: myriapod biology, invertebrate biology, evolution, genomics, molecular biology, soil biodiversity, invertebrate endocrinology

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.