



DATA NOTE

The genome sequence of the Light Emerald, *Campaea margaritaria* (Linnaeus, 1761) [version 1; peer review: 2 approved]

Douglas Boyes¹⁺, Marianne Eagles²,
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

¹UK Centre for Ecology & Hydrology, Wallingford, England, UK

²Independent researcher, Crawley Down, England, UK

+ Deceased author

V1 First published: 18 Oct 2023, **8**:476
<https://doi.org/10.12688/wellcomeopenres.20136.1>
Latest published: 18 Oct 2023, **8**:476
<https://doi.org/10.12688/wellcomeopenres.20136.1>

Abstract

We present a genome assembly from an individual male *Campaea margaritaria* (the Light Emerald; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 335.2 megabases in span. The whole assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.9 kilobases in length. Gene annotation of this assembly on Ensembl identified 16,403 protein coding genes.

Keywords

Campaea margaritaria, Light Emerald, genome sequence, chromosomal, Lepidoptera



This article is included in the [Tree of Life](#) gateway.

Open Peer Review

Approval Status

	1	2
version 1		
18 Oct 2023	view	view
<hr/>		
1. Saskia Wutke , University of Eastern Finland, Joensuu, Finland		
2. Jaakko L.O. Pohjoismäki , University of Eastern Finland, Joensuu, Finland		

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; **Eagles M:** 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.

Copyright: © 2023 Boyes D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Boyes D, Eagles M, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Light Emerald, *Campaea margaritaria* (Linnaeus, 1761) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:476 <https://doi.org/10.12688/wellcomeopenres.20136.1>

First published: 18 Oct 2023, 8:476 <https://doi.org/10.12688/wellcomeopenres.20136.1>

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; Geometroidae; Geometridae; Ennominae; *Campaea*; *Campaea margaritaria* (Linnaeus, 1761) (NCBI:txid934813).

Background

The Emerald moths are a large group of moths which belong to two distinct sub-families of the Geometridae family: most belong to Geometrinae, while the Light Emerald, *Campaea margaritaria*, belongs to Ennominae (Murillo-Ramos *et al.*, 2019). A newly emerged *C. margaritaria* has a pale green colour that fades to white within days, as the green wing pigment geoverdin is unstable (Cook *et al.*, 1994). This Geometrid moth also has a distinctive hooked red tip to the forewing (Waring *et al.*, 2017).

A common resident that is well distributed and frequently found throughout the Atlantic Archipelago of Britain and Ireland (GBIF Secretariat, 2023), *C. margaritaria* is not currently a species under threat (Butterfly Conservation, 2023). It is most abundant in broadleaved woodland but also found in scrub, hedgerows, parks, gardens and in urban areas, often disturbed while resting under leaves. There is one generation in northern England, Scotland and Ireland and two generations in the southern half of Britain, with the moth being seen from late May to early October (Waring *et al.*, 2017). The *C. margaritaria* larva, which has a distinctive fringe on the underside of its body, feeds on a variety of deciduous trees, including apple, beech, birch, elm, hawthorn, hazel and oak as well as several species of *Prunus* (Waring *et al.*, 2017). The larva lives on the foliage of deciduous trees, overwintering along the stems of the foodplant (Kimber, 2023).

A pilot study analysing the diet of bats in Belgium orchards applied DNA metabarcoding to find evidence of *C. margaritaria* in *Plecotus auritus* faecal samples (Dekeukeleire *et al.*, 2020). The completed genome sequence might be used in areas of research such as this, continuing work to identify biological pest suppressors such as bats on invasive agricultural pests. We present a chromosomally complete genome sequence for *Campaea margaritaria* based on one male specimen from Wytham Woods, Oxfordshire as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from one male *Campaea margaritaria* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 48-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 115-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly



Figure 1. Photograph of the *Campaea margaritaria* (ilCamMarg1) specimen used for genome sequencing.

curation corrected 6 missing joins or mis-joins, reducing the scaffold number by 16.22%.

The final assembly has a total length of 335.2 Mb in 31 sequence scaffolds with a scaffold N50 of 12.1 Mb (Table 1). A summary of the assembly statistics is shown in Figure 2, 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. The whole 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 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 60.3 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.3% (single = 97.9%, duplicated = 0.4%), using the lepidoptera_odb10 reference set (*n* = 5,286).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at <https://links.tol.sanger.ac.uk/species/934813>.

Genome annotation report

The *Campaea margaritaria* genome assembly (GCA_912999815.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Campaea_margaritaria_GCA_912999815.1/Info/Index). The resulting annotation includes 16,555 transcribed mRNAs from 16,403 protein-coding genes.

Table 1. Genome data for *Campaea margaritaria*, ilCamMarg1.1.

Project accession data		
Assembly identifier	ilCamMarg1.1	
Assembly release date	2021-08-18	
Species	<i>Campaea margaritaria</i>	
Specimen	ilCamMarg1	
NCBI taxonomy ID	934813	
BioProject	PRJEB45206	
BioSample ID	SAMEA7701535	
Isolate information	ilCamMarg1, male: abdomen (DNA sequencing); head and thorax (Hi-C scaffolding)	
Assembly metrics*		Benchmark
Consensus quality (QV)	60.3	≥ 50
k-mer completeness	100%	≥ 95%
BUSCO**	C:98.3%,S:97.9%,D:0.4%, F:0.4%,M:1.3%,n:5,286	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	ERR6454734, ERR6939214	
10X Genomics Illumina	ERR6054360, ERR6054361, ERR6054359, ERR6054362	
Hi-C Illumina	ERR6054358	
Genome assembly		
Assembly accession	GCA_912999815.1	
Accession of alternate haplotype	GCA_912999775.1	
Span (Mb)	335.2	
Number of contigs	38	
Contig N50 length (Mb)	11.9	
Number of scaffolds	31	
Scaffold N50 length (Mb)	12.1	
Longest scaffold (Mb)	18.7	
Genome annotation		
Number of protein-coding genes	16,403	
Number of gene transcripts	16,555	

* 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/ilCamMarg1.1/dataset/ilCamMarg1_1.1/busco.

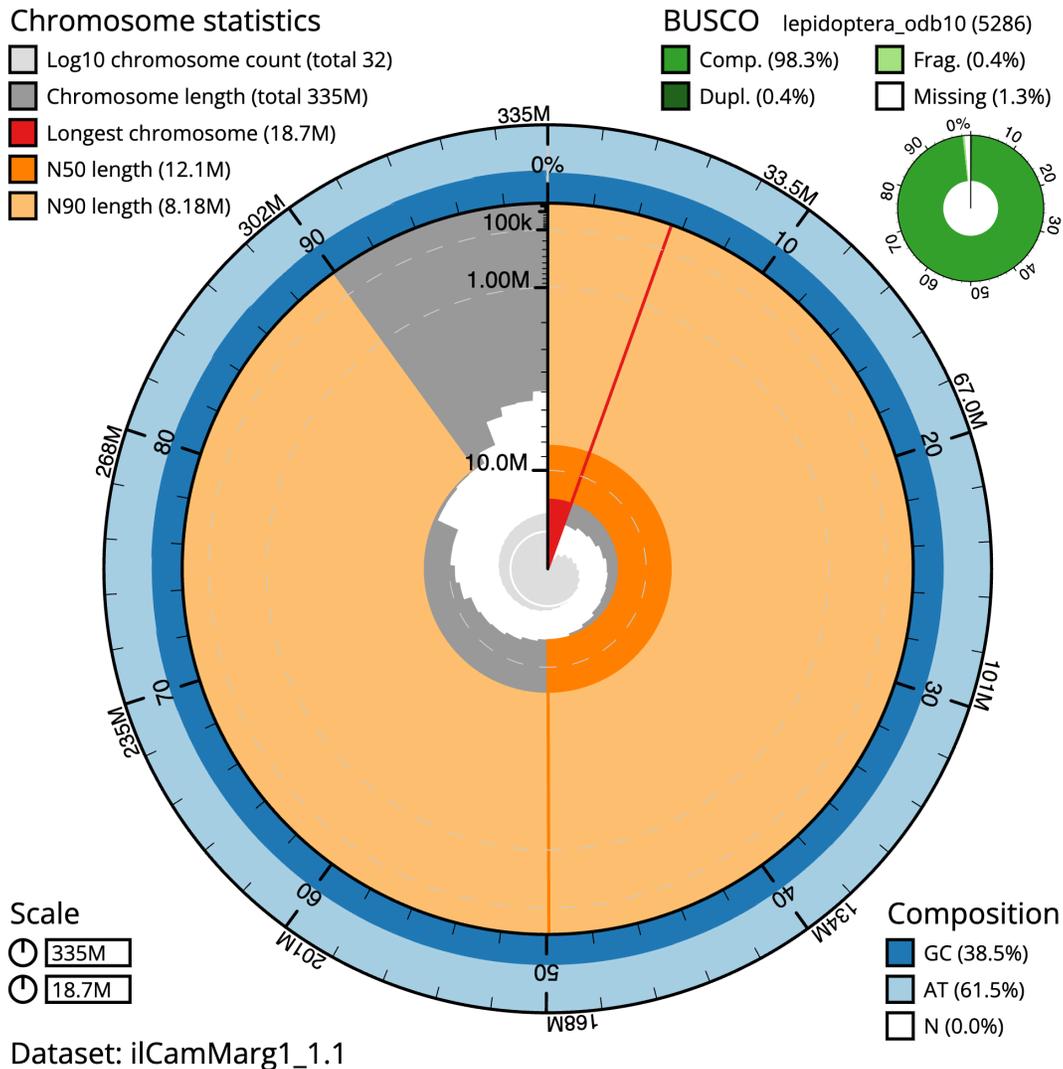


Figure 2. Genome assembly of *Campaea margaritaria*, ilCamMarg1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 335,226,495 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (18,728,427 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,093,622 and 8,177,772 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/ilCamMarg1.1/dataset/ilCamMarg1_1.1/snail.

Methods

Sample acquisition and nucleic acid extraction

A male *Campaea margaritaria* (specimen ID Ox000674, ToLID ilCamMarg1) 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.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilCamMarg1 sample was

weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Abdomen tissue 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

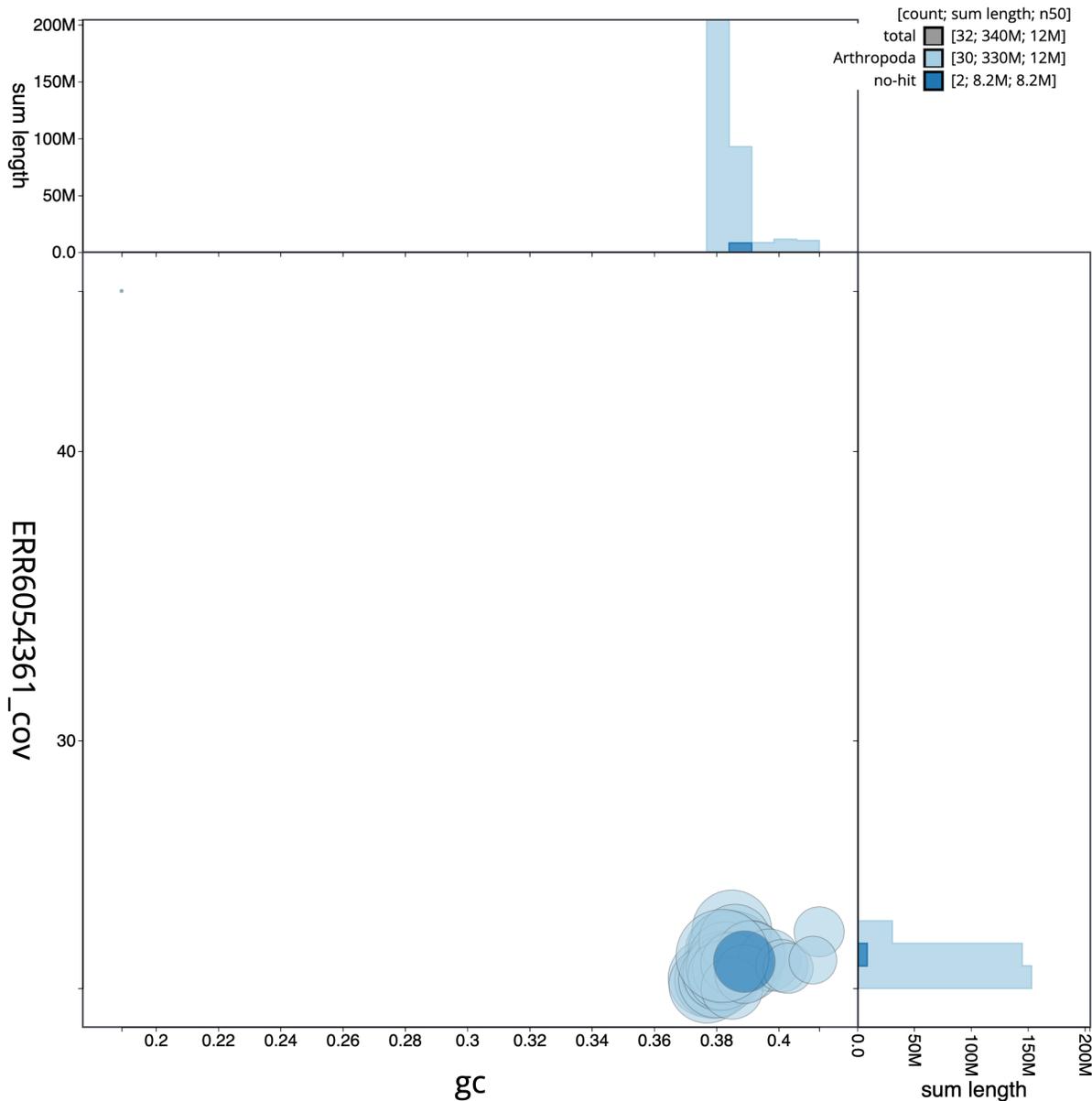


Figure 3. Genome assembly of *Campaea margaritaria*, ilCamMarg1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilCamMarg1.1/dataset/ilCamMarg1_1.1/blob.

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.

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 (10X) instruments. Hi-C data were also generated from head and thorax tissue of ilCamMarg1 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

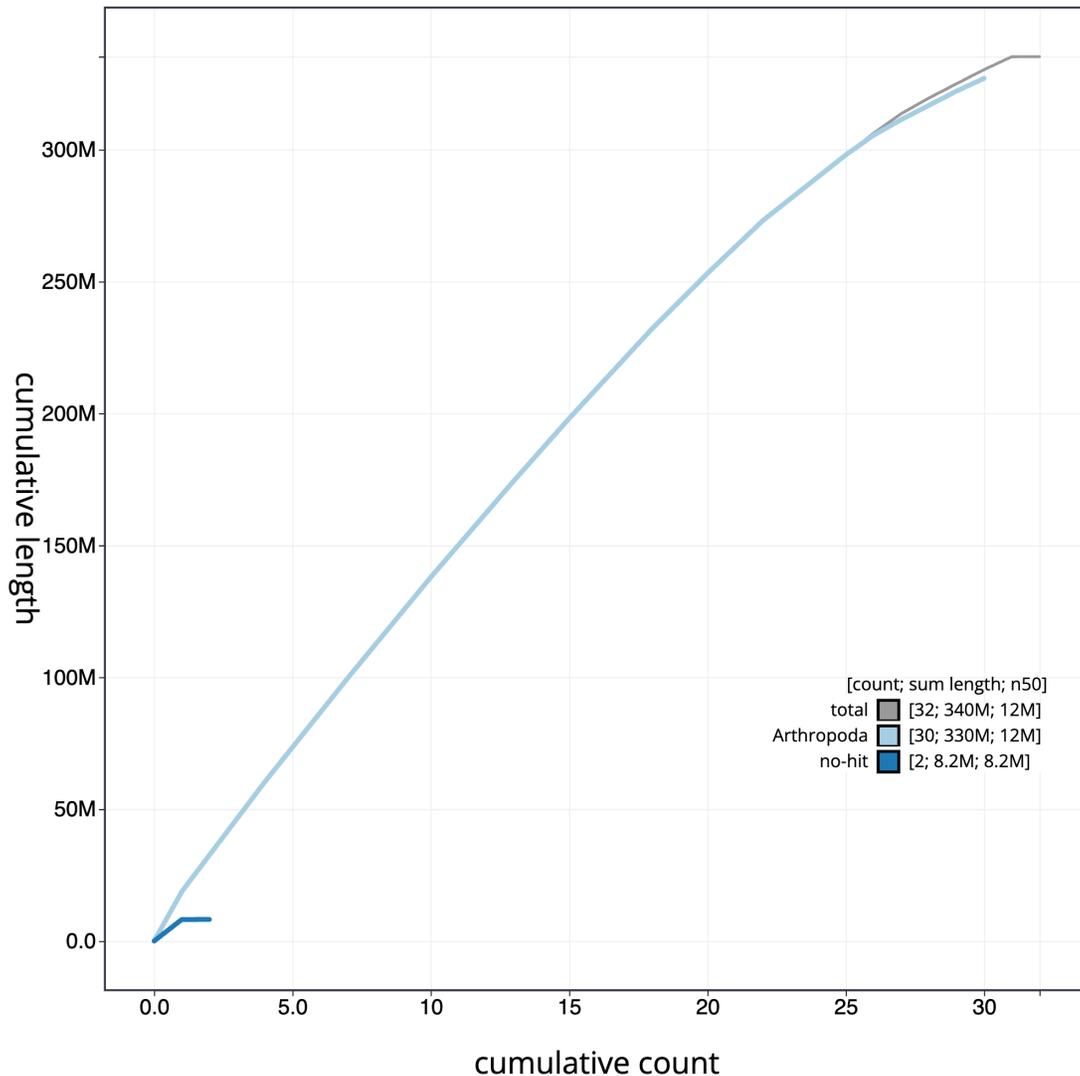


Figure 4. Genome assembly of *Campaea margaritaria*, ilCamMarg1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscodegenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilCamMarg1.1/dataset/ilCamMarg1_1.1/cumulative.

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.*, 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.

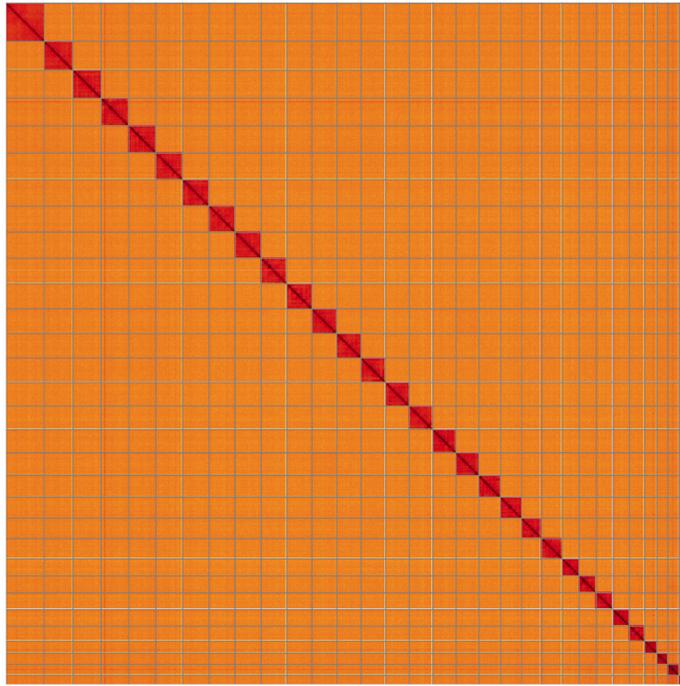


Figure 5. Genome assembly of *Campaea margaritaria*, ilCamMarg1.1: Hi-C contact map of the ilCamMarg1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=fCFg_877SRicnS1Ytr6jWQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Campaea margaritaria*, ilCamMarg1.

INSDC accession	Chromosome	Length (Mb)	GC%
OU538789.1	1	14.13	38.3
OU538790.1	2	13.82	38.4
OU538791.1	3	13.63	38.5
OU538792.1	4	13.33	38.6
OU538793.1	5	13.08	38.2
OU538794.1	6	12.99	37.7
OU538795.1	7	12.88	37.9
OU538796.1	8	12.79	38.1
OU538797.1	9	12.54	37.7
OU538798.1	10	12.33	38
OU538799.1	11	12.12	38.3
OU538800.1	12	12.09	38.3
OU538801.1	13	11.94	38.3
OU538802.1	14	11.69	38.1
OU538803.1	15	11.47	38.2

INSDC accession	Chromosome	Length (Mb)	GC%
OU538804.1	16	11.36	38.4
OU538805.1	17	11.21	38.6
OU538806.1	18	10.63	39.1
OU538807.1	19	10.34	38.2
OU538808.1	20	10	39
OU538809.1	21	9.98	39.1
OU538810.1	22	8.37	38.5
OU538811.1	23	8.34	39.7
OU538812.1	24	8.22	38.5
OU538813.1	25	8.18	38.9
OU538814.1	26	7.42	38.9
OU538815.1	27	5.93	40.1
OU538816.1	28	5.42	40.3
OU538817.1	29	5.38	41.3
OU538818.1	30	4.88	41.1
OU538788.1	Z	18.73	38.2
OU538819.1	MT	0.02	18.9

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.0.7	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	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

Genome annotation

The BRAKER2 pipeline (Brûna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Campaea margaritaria* assembly (GCA_912999815.1) 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)

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: *Campaea margaritaria* (light emerald). Accession number PRJEB45206; <https://identifiers.org/ena.embl/PRJEB45206>. (Wellcome Sanger Institute, 2021) The genome sequence is released openly for reuse. The *Campaea margaritaria* 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](#)
- Butterfly Conservation: **Light Emerald *Campaea margaritata*.** *Butterfly-conservations.org.* 2023; [Accessed 13 September 2023]. [Reference Source](#)
- 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](#)
- Cook MA, Harwood LM, Scoble MJ, et al.: **The chemistry and systematic importance of the green wing pigment in emerald moths.** *Entomol Exp Appl.* 1994; **22**(1): 43–51. [Publisher Full Text](#)
- Dekeukeleire D, Janssen R, Delbroek R, et al.: **First molecular evidence of an invasive agricultural pest, *Drasophila suzukii*, in the diet of a common bat *Pipistrellus pipistrellus*, in Belgian orchards.** *Barbastella.* 2020; **13**(1): 109–115. [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](#)
- Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012; [Accessed 26 July 2023]. [Publisher Full Text](#)
- GBIF Secretariat: ***Campaea margaritaria* (Linnaeus, 1761).** *GBIF Backbone Taxonomy.* 2023; [Accessed 13 September 2023]. [Reference Source](#)
- Ghurye J, Rhie A, Walenz BP, et al.: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guan D, McCarthy SA, Wood J, et al.: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022]. [Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): gjaa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, et al.: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kimber I: **Light Emerald *Campaea margaritaria* (Linnaeus, 1761).** UKMoths. 2023; [Accessed 13 September 2023]. [Reference Source](#)
- Manni M, Berkeley MR, Seppely 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](#)
- Murillo-Ramos L, Brehm G, Sihvonen P, et al.: **A comprehensive molecular phylogeny of Geometridae (Lepidoptera) with a focus on enigmatic small subfamilies.** *PeerJ.* 2019; **7**: e7386. [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.: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simão FA, Waterhouse RM, Ioannidis P, et al.: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212. [PubMed Abstract](#) | [Publisher Full Text](#)
- Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a; [Accessed 21 July 2023]. [Publisher Full Text](#)
- Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023]. [Publisher Full Text](#)
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova 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](#)
- 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 Light Emerald, *Campaea margaritaria* (Linnaeus, 1761).** European Nucleotide Archive. [dataset], accession number PRJEB45206, 2021.

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 16 February 2024

<https://doi.org/10.21956/wellcomeopenres.22298.r72644>

© 2024 Pohjoismäki J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jaakko L.O. Pohjoismäki 

Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland

I have no reservations regarding the presented work. The genome assembly and annotation follows the DToL pipeline. The species is a great addition to the growing list of moth genomes. However, it would have been great to have the heterogametic female sex sampled.

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: molecular biology, biodiversity genetics, taxonomy, entomology

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 02 February 2024

<https://doi.org/10.21956/wellcomeopenres.22298.r72643>

© 2024 Wutke S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Saskia Wutke 

Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland

This data note reports the genome of the Light Emerald (*Campaea margaritaria*), a geometrid moth. Genome sequencing, assembly and annotation is based PacBio HiFi, 10X Genomics and Hi-C data. Overall, the dataset is presented clearly and concisely. The methods are state-of-the-art. Yet, reproducibility would benefit from further details concerning the settings of the employed software tools. Moreover, the relevance of this genome report could be explained more in the background section. For example, it is mentioned briefly that the genome can be used for work on invasive agricultural pests but it is not mentioned anywhere if *Campaea margaritaria* falls into this category. Nevertheless, this genome provides a valuable data resource for future research on insect comparative genomics, biodiversity and evolution.

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?

Partly

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: evolutionary genetics, insect genomics, phylogenomics

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.
