



DATA NOTE

The genome sequence of the Red-green Carpet, *Chloroclysta siterata* (Hufnagel, 1767) [version 1; peer review: 2 approved]

Douglas Boyes¹⁺, Peter W.H. Holland²,
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

²University of Oxford, Oxford, England, UK

+ Deceased author

V1 First published: 10 May 2023, 8:206
<https://doi.org/10.12688/wellcomeopenres.19301.1>
Latest published: 10 May 2023, 8:206
<https://doi.org/10.12688/wellcomeopenres.19301.1>

Abstract

We present a genome assembly from an individual male *Chloroclysta siterata* (the Red-green Carpet; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 437.9 megabases in span. Most of the assembly is scaffolded into 21 chromosomal pseudomolecules including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.7 kilobases in length. Gene annotation of this assembly on Ensembl identified 11,814 protein coding genes.

Keywords

Chloroclysta siterata, Red-green Carpet, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1 10 May 2023	 view	 view
1. XiaoLong Liu , Hubei University, Wuhan, China		
2. Hana Konvičková , Biology Centre CAS, Ceske Budejovice, Czech Republic		

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; **Holland PWH:** 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.*

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, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Red-green Carpet, *Chloroclysta siterata* (Hufnagel, 1767) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:206 <https://doi.org/10.12688/wellcomeopenres.19301.1>

First published: 10 May 2023, 8:206 <https://doi.org/10.12688/wellcomeopenres.19301.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Geometroidea; Geometridae; Larentiinae; *Chloroclysta*; *Chloroclysta siterata* (Hufnagel, 1767) (NCBI:txid934828).

Background

The Red-green Carpet, *Chloroclysta siterata*, is a delicately patterned moth in the family Geometridae. The name ‘carpet moth’ for this and closely related species has no relation to their diet; instead, it derives from the forewing markings which comprise a series of irregular bands, considered to resemble the ornate patterns on woven rugs and carpets. In *C. siterata*, these bands are alternating shades of dark and light green, suffused with ruby red streaks. The wings are held flat against the surface at rest and may provide cryptic camouflage on bark and lichen. The green colouration is vivid in freshly emerged specimens but fades with age; this may be the basis for the specific name *siterata*, meaning ‘pertaining to corn’, a plant which also changes colour (Emmet, 1991)

The species is widespread across northern and eastern Europe, and in the UK is found predominantly in southern England and Wales (NBN Atlas Partnership, 2021; GBIF Secretariat, 2022). Both distribution and abundance have shown large long-term increases – over six-fold since 1970 (Randle *et al.*, 2019). *C. siterata* has an unusual life-cycle, being univoltine but with two flight periods. Most records of the adult moth in Britain and Ireland are from September to November, when they will visit ivy blooms or are attracted to light. These adults mate, and the mated females overwinter as adults, emerging to give a second flight period from April to June (Newman, 1869). In Italy, Austria and Finland, overwintering in caves has been recorded (Moog *et al.*, 2021; Soderholm, 2022; Teobladelli, 2008). After egg-laying, the larvae develop through July and August, feeding on foliage of deciduous trees - primarily oak but also apple, cherry, rose, rowan, blackthorn and birch (South, 1961; Waring *et al.*, 2017).

A genome sequence of *C. siterata* will be useful in analyses of molecular adaptations to polyphagy and as part of wider comparative studies into genome evolution in the Lepidoptera.

Genome sequence report

The genome was sequenced from one male *Chloroclysta siterata* (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude -1.34). A total of 39-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 two missing joins or mis-joins.

The final assembly has a total length of 437.9 Mb in 25 sequence scaffolds with a scaffold N50 of 24.9 Mb



Figure 1. Photographs of the *Chloroclysta siterata* (ilChlSite2) specimen used for genome sequencing.

(Table 1). Most (99.97%) of the assembly sequence was assigned to 20 chromosomal-level scaffolds, representing 19 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 70.1 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.2% (single = 97.8%, 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/934828>.

Genome annotation report

The *Chloroclysta siterata* genome assembly GCA_932294275.1 (ilChlSite2.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Chloroclysta_siterata_GCA_932294275.1/Info/Index). The resulting annotation includes 19,880 transcribed mRNAs from 11,814 protein-coding and 1,416 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Chloroclysta siterata* (ilChlSite2) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude -1.34) on 8 October 2020. The specimen was taken from woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The specimen was identified by the collector and snap-frozen on dry ice.

Table 1. Genome data for *Chloroclysta siterata*, ilChlSite2.1.

Project accession data		
Assembly identifier	ilChlSite2.1	
Species	<i>Chloroclysta siterata</i>	
Specimen	ilChlSite2	
NCBI taxonomy ID	934828	
BioProject	PRJEB50736	
BioSample ID	SAMEA8603199	
Isolate information	ilChlSite2, thorax (genome sequencing), head (Hi-C scaffolding), abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	70.1	≥ 50
<i>k</i> -mer completeness	100%	≥ 95%
BUSCO**	C:98.2%[S:97.8%,D:0.4%], F:0.4%,M:1.4%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.97%	≥ 95%
Sex chromosomes	Z chromosome	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR8575370, ERR8575371	
Hi-C Illumina	ERR8571652	
PolyA RNA-Seq Illumina	ERR8571653	
Genome assembly		
Assembly accession	GCA_932294275.1	
<i>Accession of alternate haplotype</i>	GCA_932294285.1	
Span (Mb)	437.9	
Number of contigs	29	
Contig N50 length (Mb)	20.6	
Number of scaffolds	25	
Scaffold N50 length (Mb)	24.9	
Longest scaffold (Mb)	31.6	
Genome annotation		
Number of protein-coding genes	11,814	
Number of non-coding genes	1,416	
Number of transcripts	19,880	

*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/ilChlSite2.1/dataset/CAK0AF01/busco>.

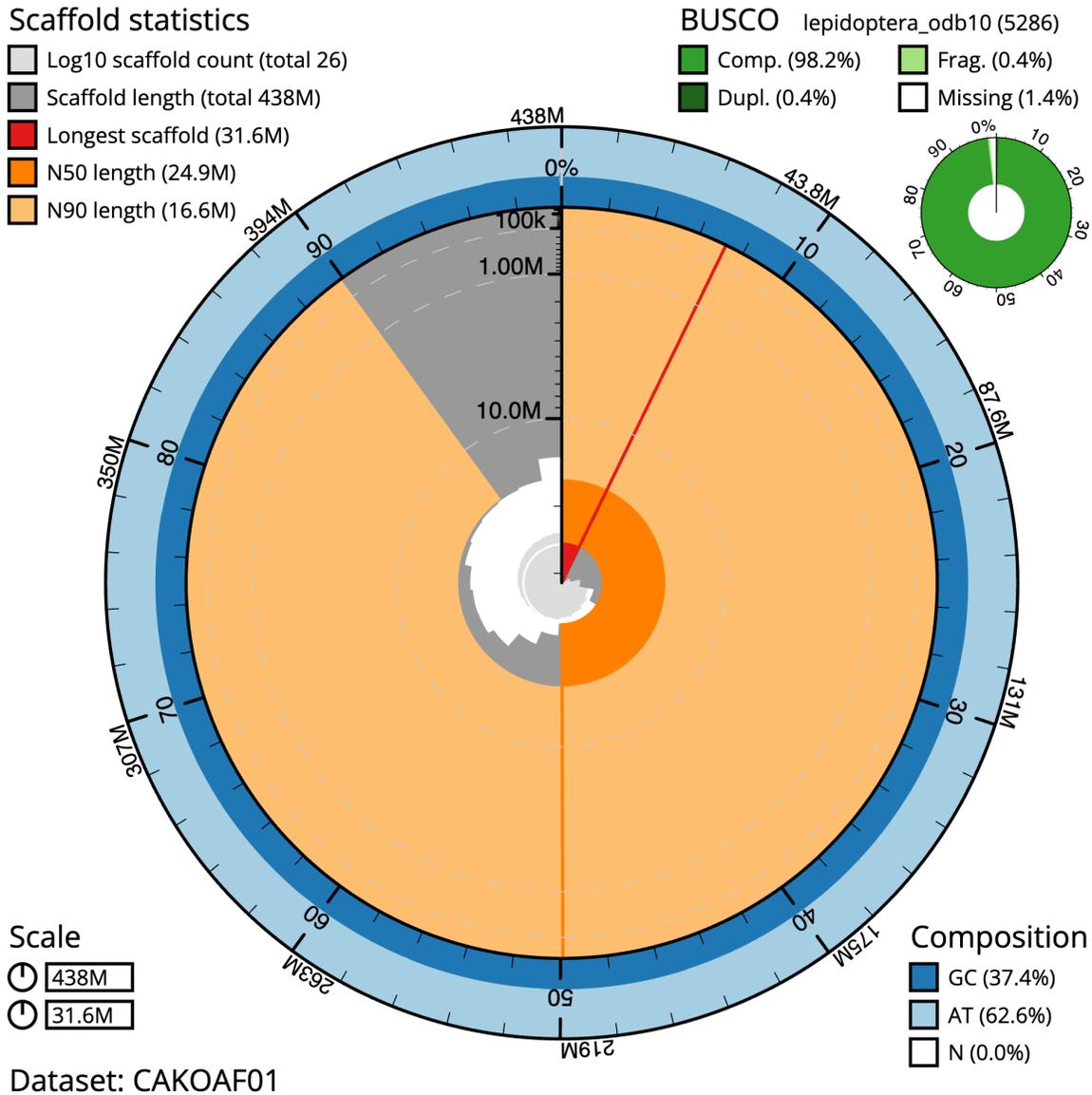


Figure 2. Genome assembly of *Chloroclysta siterata*, iChlSite2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 437,887,824 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (31,593,026 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (24,933,739 and 16,571,927 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/iChlSite2.1/dataset/CAKOAF01/snail>.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The iChlSite2 sample was weighed and dissected on dry ice with head 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. 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

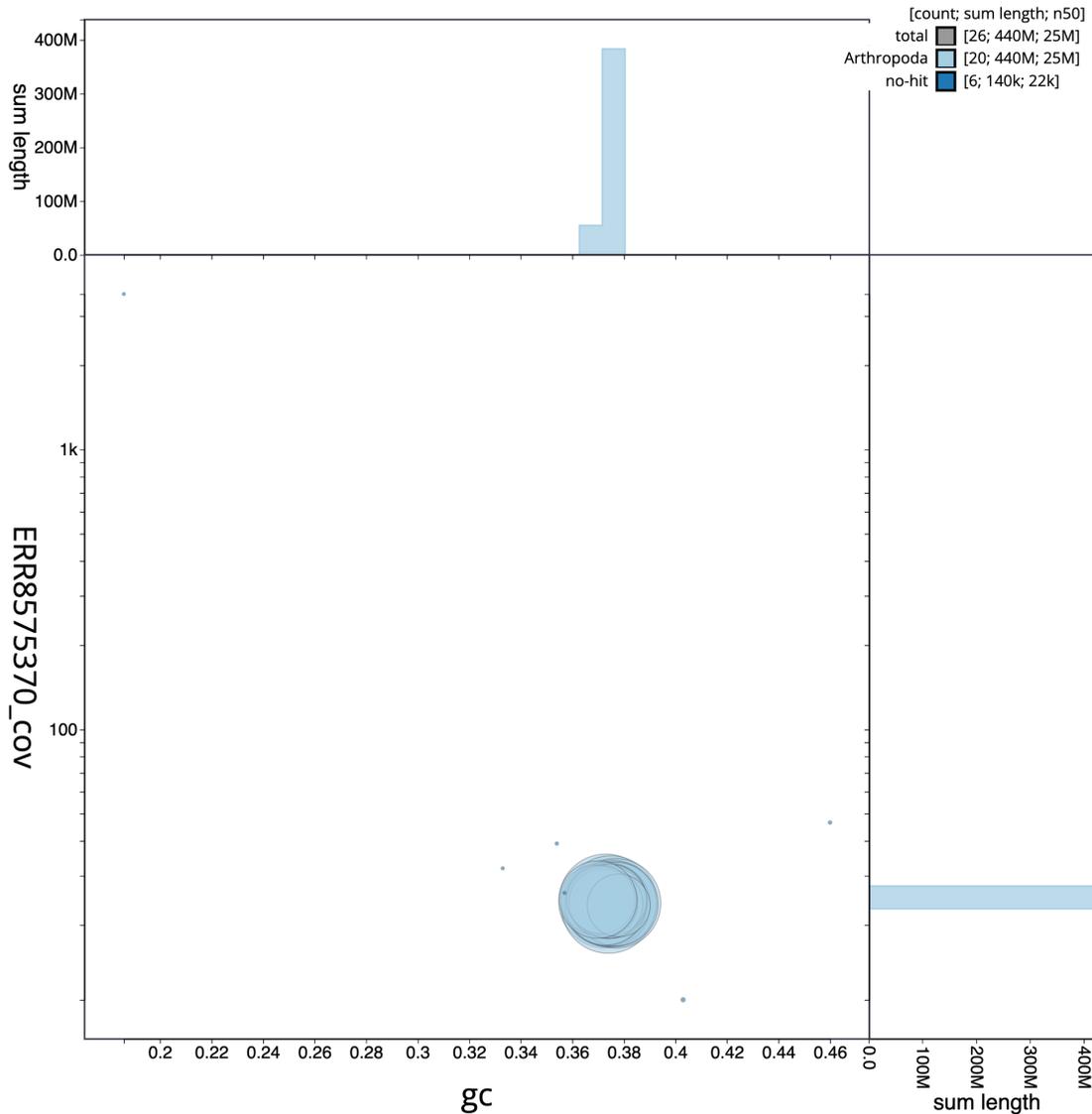


Figure 3. Genome assembly of *Chloroclysta siterata*, ilChlSite2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilChlSite2.1/dataset/CAKOAF01/blob>.

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 ilChlSite2 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μ l RNase-free water and its concentration assessed using a 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 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 were performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina HiSeq 4000 (RNA-Seq) instruments.

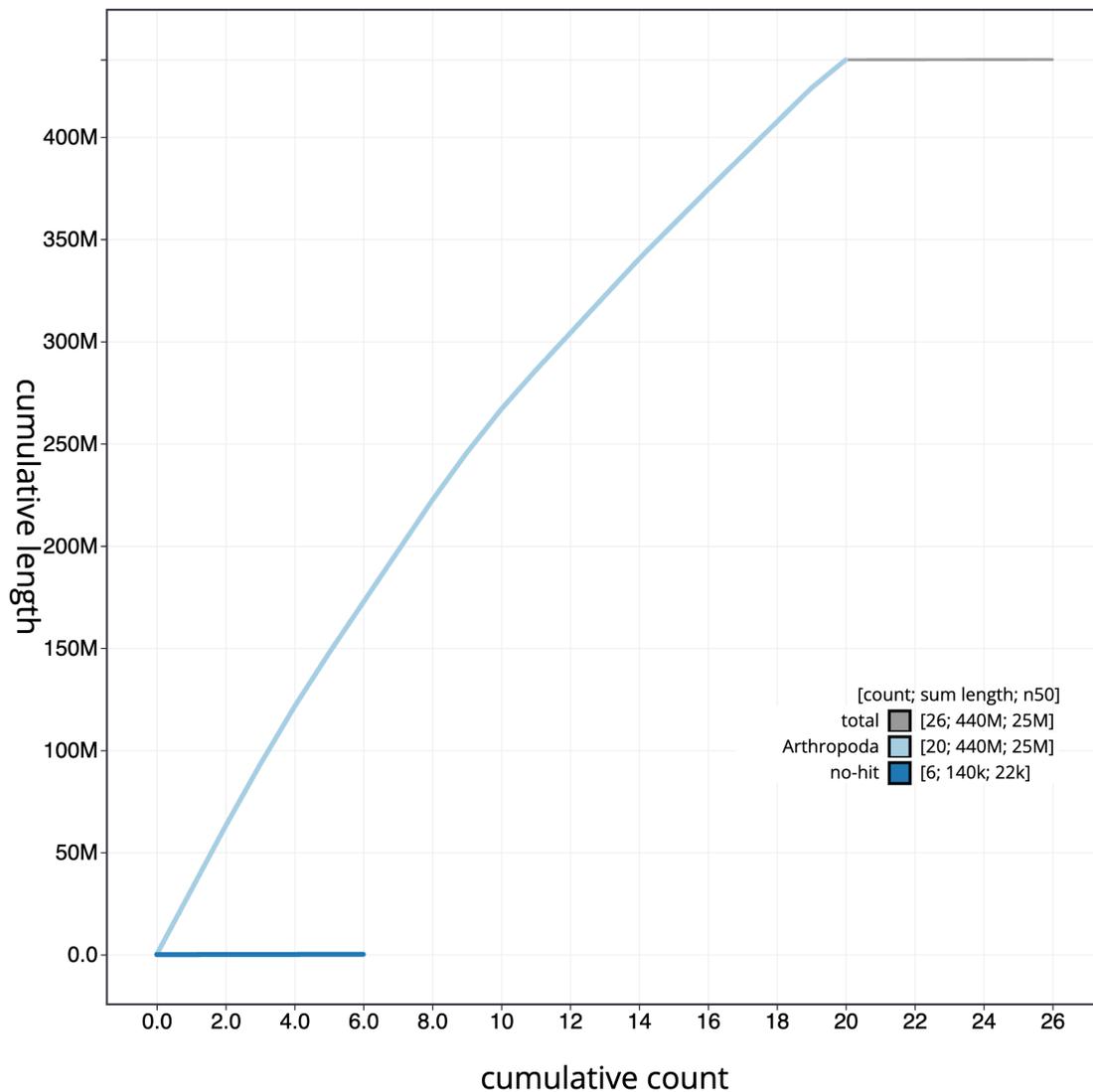


Figure 4. Genome assembly of *Chloroclysta siterata*, iChlSite2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iChlSite2.1/dataset/CAKOAF01/cumulative>.

Hi-C data were also generated from head tissue of iChl-Site2 using the Arima v2 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 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. To evaluate the assembly, MerquryFK was used to estimate consensus quality (QV) scores and *k*-mer completeness (Rhie *et al.*, 2020). 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 software tool versions and sources.

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Chloroclysta*

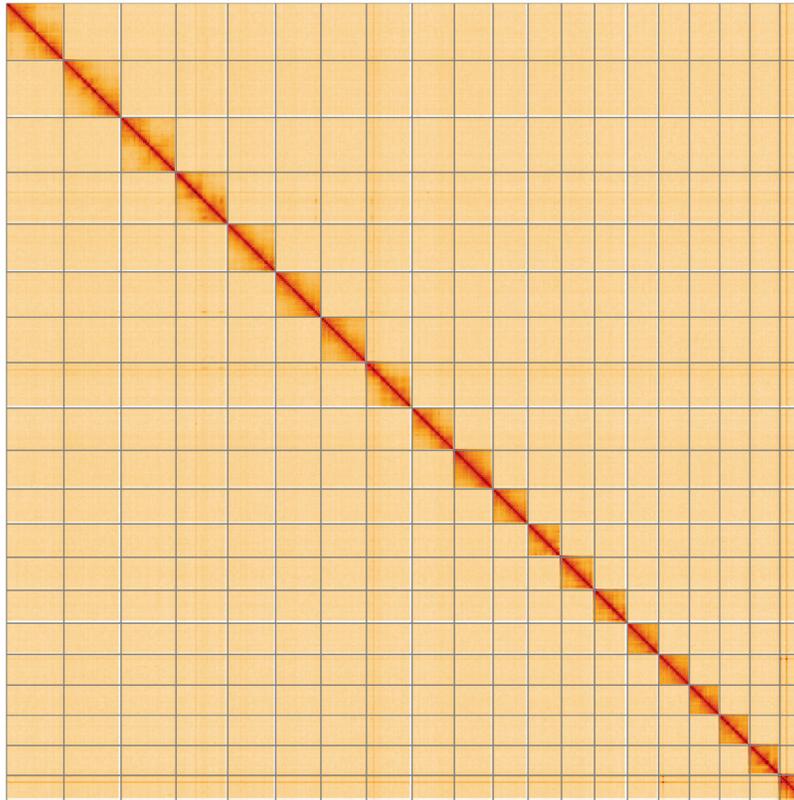


Figure 5. Genome assembly of *Chloroclysta siterata*, ilChlSite2.1: Hi-C contact map of the ilChlSite2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/?d=V7ANhS-_Rd6CIbouTb-IJA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Chloroclysta siterata*, ilChlSite2.

INSDC accession	Chromosome	Size (Mb)	GC%
OW028588.1	1	31.59	37.3
OW028589.1	2	31.29	37.4
OW028590.1	3	29.98	37.5
OW028591.1	4	28.42	37.7
OW028592.1	5	26.17	37.6
OW028593.1	6	24.96	37.4
OW028594.1	7	24.94	37.3
OW028595.1	8	24.93	37.7
OW028596.1	9	23.21	37.3

INSDC accession	Chromosome	Size (Mb)	GC%
OW028598.1	10	19.06	37.4
OW028599.1	11	18.29	37.5
OW028600.1	12	18.24	37.5
OW028601.1	13	17.95	37.5
OW028602.1	14	17.07	36.9
OW028603.1	15	16.84	37.4
OW028604.1	16	16.7	37.2
OW028605.1	17	16.57	37.2
OW028606.1	18	16.39	37.1
OW028607.1	19	13.87	37.8
OW028597.1	Z	21.26	37
OW028608.1	MT	0.02	18.7

Table 3. Software tools: sources and versions.

Software tool	Version	Source
BlobToolKit	4.0.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhy123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Mercury	MercuryFK	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
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

siterata assembly (GCA_932294275.1). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

Ethics and compliance issues

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. 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. All efforts are undertaken to minimise the suffering of animals used for sequencing. 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: *Chloroclysta siterata*. Accession number PRJEB50736; <https://identifiers.org/ena.embl/PRJEB50736>. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Chloroclysta siterata* 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

- Aken BL, Ayling S, Barrell D, et al.: **The Ensembl gene annotation system. Database (Oxford).** 2016; **2016**: baw093.
[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](#)

Emmet AM: **The Scientific Names of the British Lepidoptera - their History and Meaning**. Colchester: Harley Books, 1991.

[Reference Source](#)

GBIF Secretariat: ***Chloroclysta siterata* (Hufnagel, 1767)**. Red-green Carpet. *GBIF Backbone Taxonomy*. Checklist dataset, 2022; (Accessed: 24 December 2022).

[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): g1aa153.

[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, 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](#)

Moog O, Christian E, Eis R: **Increased cave use by butterflies and moths: a response to climate warming?** *International Journal of Speleology*. 2021; **50**(1): 15–24.

[Publisher Full Text](#)

NBN Atlas Partnership: ***Chloroclysta siterata* (Hufnagel, 1767)**. NBN Atlas, 2021; (Accessed: 24 December 2022).

[Reference Source](#)

Newman E: **An Illustrated Natural History of British Moths**. London: Hardwicke and Bogue, 1869.

[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.*: **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](#)

Soderholm H: **Red-green carpet moth (*Chloroclysta siterata*) hibernating in a natural cave, wild Finland**. alamy, 2022; (Accessed: 17 March 2023).

[Reference Source](#)

South R: **Moths of the British Isles. New edition**. London: Frederick Warne and Co, 1961.

Teobladelli A: **Elenco aggiornato dei Lepidotteri rinvenuti all'interno di alcune caverne nelle Marche**. In: D. Fiacchini, G. Carotti, and G. Fusco (eds) *Convegno Biospeleologia dell'Appennino: studi e ricerche su Anfibi e Invertebrati, con particolare riferimento all'Appennino UmbroMarchigiano*. 2008; 66–70.

[Reference Source](#)

Uliano-Silva M, Ferreira JGRN, Krashennikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads**. *bioRxiv*. [Preprint], 2022.

[Publisher Full Text](#)

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge**. *Nucleic Acids Res*. 2019; **47**(D1): D506–D515.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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 Red-green Carpet, *Chloroclysta siterata* (Hufnagel, 1767)**. European Nucleotide Archive. [dataset], accession number PRJEB50736, 2022.

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 09 May 2024

<https://doi.org/10.21956/wellcomeopenres.21384.r81602>

© 2024 Konvičková H. 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.



Hana Konvičková 

Institute of Entomology, Biology Centre CAS, Ceske Budejovice, Czech Republic

The paper is clearly written and I appreciate its briefness with no excess information. The authors aimed to sequence a genome of widespread moth *Chloroclysta siterata*. This species is polyphagous and the authors hope to reveal the molecular mechanisms enabling it. They used modern method of HiFi sequencing. The method and processing the data are accurate.

I have only two minor comments:

- 1) Figure 1 - there is only one photograph, although the title says "photographs"
- 2) The origin of the sample is mentioned twice (Genome sequence report and Methods), which I believe is not necessary.

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: population genetics of animals, with focus on 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 20 February 2024

<https://doi.org/10.21956/wellcomeopenres.21384.r72557>

© 2024 Liu X. 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.



XiaoLong Liu

Hubei University, Wuhan, China

In the Background part, detailed introduction to the morphological characteristics, lifestyle, and distribution range of *Chloroclysta siterata*. Detailed description of the Methods part and the protocols sound appropriate. Clearly described the basic principles of creating datasets. Genome data for *C. siterata* were showed in the table and raw sequence data and the assembly have been deposited in INSDC databases.

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: Insect Physiology Biochemistry and Molecular Biology, Transcriptomics

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.
