



DATA NOTE

# The genome sequence of the Elm Groundling moth, *Carpatolechia fugitivella* (Zeller, 1839) [version 1; peer review: 2 approved]

Douglas Boyes <sup>1+</sup>, Liam M. Crowley <sup>2</sup>, James McCulloch<sup>2</sup>, Clare Boyes <sup>3</sup>,  
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

<sup>1</sup>UK Centre for Ecology & Hydrology, Wallingford, England, UK

<sup>2</sup>University of Oxford, Oxford, England, UK

<sup>3</sup>Independent researcher, Welshpool, Wales, UK

+ Deceased author

**V1** First published: 23 Apr 2024, 9:211  
<https://doi.org/10.12688/wellcomeopenres.21267.1>  
Latest published: 23 Apr 2024, 9:211  
<https://doi.org/10.12688/wellcomeopenres.21267.1>

## Abstract

We present a genome assembly from an individual male *Carpatolechia fugitivella* (the Elm Groundling; Arthropoda; Insecta; Lepidoptera; Gelechiidae). The genome sequence is 493.1 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 15.26 kilobases in length. Gene annotation of this assembly on Ensembl identified 12,721 protein coding genes.

## Keywords

*Carpatolechia fugitivella*, Elm Groundling moth, genome sequence, chromosomal, Lepidoptera



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

## Open Peer Review

Approval Status

	1	2
<b>version 1</b> 23 Apr 2024	 <a href="#">view</a>	 <a href="#">view</a>

1. **Annabel Whibley** , The University of Auckland, Auckland, New Zealand  
Bragato Research Institute, Lincoln, New Zealand

2. **Jerome H L Hui** , The Chinese University of Hong Kong, Hong Kong, Hong Kong

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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **Boyes D:** Investigation, Resources; **Crowley LM:** Investigation, Resources; **McCulloch J:** 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.*

**Copyright:** © 2024 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, Crowley LM, McCulloch J *et al.* **The genome sequence of the Elm Groundling moth, *Carpatolechia fugitivella* (Zeller, 1839) [version 1; peer review: 2 approved]** Wellcome Open Research 2024, 9:211 <https://doi.org/10.12688/wellcomeopenres.21267.1>

**First published:** 23 Apr 2024, 9:211 <https://doi.org/10.12688/wellcomeopenres.21267.1>

## 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; *Carpatolechia*; *Carpatolechia fugitivella* (Zeller, 1839) (NCBI:txid687316).

## Background

*Carpatolechia fugitivella*, the Elm Groundling, is a moth in the family Gelechiidae. It is local in England and Wales, becoming less common in Scotland and Ireland. It is found in central and northern Europe with scattered records as far east as Russia. It is also found in North America (GBIF Secretariat, 2024).

The adult moth (forewing length 6–7.5 mm) is rather non-descript and has greyish-brown forewings with black speckling. The patterning can be variable and there are up to three patches of raised scales on the mid-wing which can be whitish or black (Sterling *et al.*, 2023). The adult moth flies between June and early September in gardens, hedgerows and woodland and also comes to light (Sterling *et al.*, 2023). The eggs, laid on elm and wych elm, are thought to overwinter before hatching the following spring (Emmet & Langmaid, 2002). The caterpillars feed in folded leaves or leaf-spinnings before pupating on the bark of the host tree (Langmaid *et al.*, 2018).

The genome of *Carpatolechia fugitivella* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *Carpatolechia fugitivella* based on one male specimen from Wytham Woods, Oxfordshire, UK.

## Genome sequence report

The genome was sequenced from one male *Carpatolechia fugitivella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 41-fold coverage in



**Figure 1.** Photograph of the *Carpatolechia fugitivella* (ilCarFugi1) specimen used for genome sequencing.

Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 93 missing joins or mis-joins and removed 5 haplotypic duplications, reducing the scaffold number by 13.94%, and increasing the scaffold N50 by 0.46%.

The final assembly has a total length of 493.1 Mb in 141 sequence scaffolds with a scaffold N50 of 17.2 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 (99.15%) 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). The Z chromosome was assigned based on synteny with *Athrips mouffetella* (GCA\_947532105.1) (Boyes *et al.*, 2024). 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 58.2 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 97.9% (single = 97.1%, duplicated = 0.8%), 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/687316>.

## Genome annotation report

The *Carpatolechia fugitivella* genome assembly (GCA\_951230895.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 23,512 transcribed mRNAs from 12,721 protein-coding and 1,593 non-coding genes (Table 1; [https://rapid.ensembl.org/Carpatolechia\\_fugitivella\\_GCA\\_951230895.1/Info/Index](https://rapid.ensembl.org/Carpatolechia_fugitivella_GCA_951230895.1/Info/Index)).

## Methods

### Sample acquisition and nucleic acid extraction

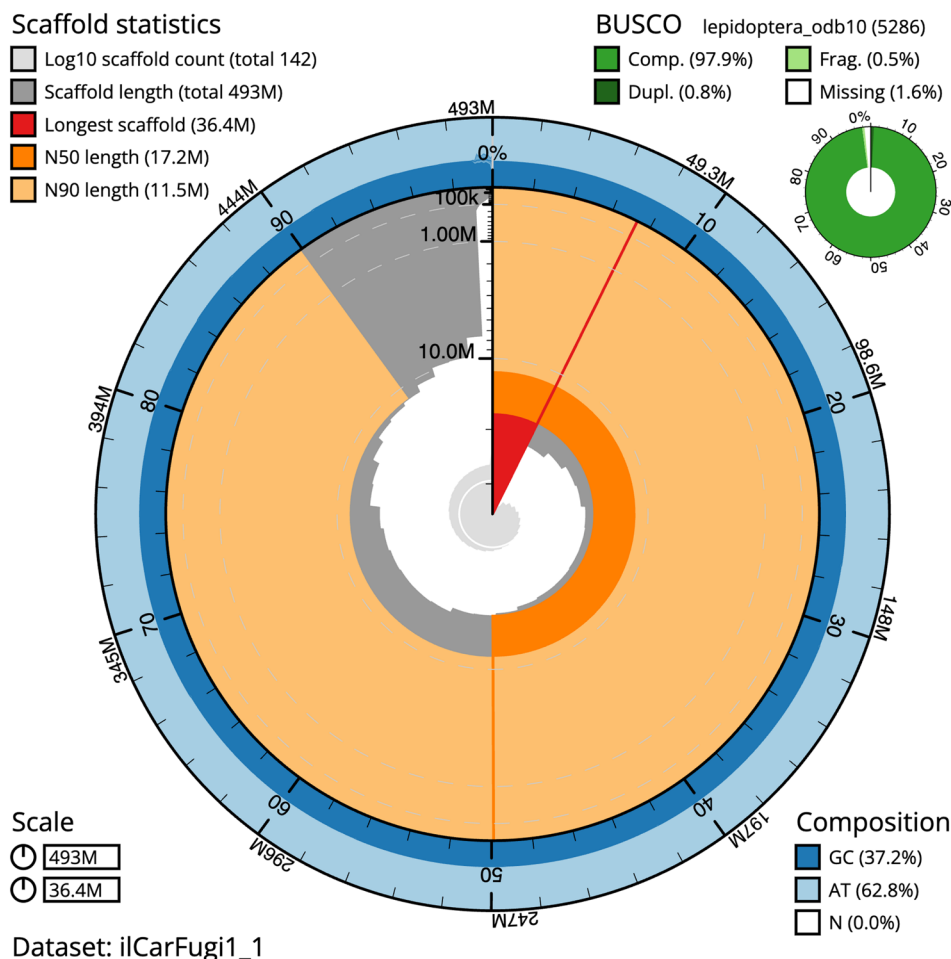
The specimen used for DNA sequencing, a male *Carpatolechia fugitivella* (specimen ID Ox000831, ToLID ilCarFugi1), was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2020-08-01 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved by snap-freezing on dry ice. The specimen used for Hi-C and RNA sequencing (specimen ID Ox002539, ToLID ilCarFugi2) was collected from the same location on 2022-07-2, also using a light trap. The specimen was collected by Liam Crowley and James McCulloch (University of Oxford), identified by James McCulloch, and then preserved on dry ice.

**Table 1. Genome data for *Carpatolechchia fugitivella*, ilCarFugi1.1.**

Project accession data		
Assembly identifier	ilCarFugi1.1	
Species	<i>Carpatolechchia fugitivella</i>	
Specimen	ilCarFugi1	
NCBI taxonomy ID	687316	
BioProject	PRJEB61343	
BioSample ID	SAMEA7746638	
Isolate information	ilCarFugi1, whole organism (DNA sequencing) ilCarFugi2, whole organism (Hi-C and RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	58.2	≥ 50
k-mer completeness	99.99%	≥ 95%
BUSCO**	C:97.9%[S:97.1%,D:0.8%], F:0.5%,M:1.6%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.15%	≥ 95%
Sex chromosomes	Z	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome: 15.26 kb	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11242130	
Hi-C Illumina	ERR11242553	
PolyA RNA-Seq Illumina	ERR11242554	
Genome assembly		
Assembly accession	GCA_951230895.1	
Accession of alternate haplotype	GCA_951230885.1	
Span (Mb)	493.1	
Number of contigs	941	
Contig N50 length (Mb)	1.0	
Number of scaffolds	141	
Scaffold N50 length (Mb)	17.2	
Longest scaffold (Mb)	36.6	
Genome annotation		
Number of protein-coding genes	12,721	
Number of non-coding genes	1,593	
Number of gene transcripts	23,512	

\* 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/ilCarFugi1\\_1/dataset/ilCarFugi1\\_1/busco](https://blobtoolkit.genomehubs.org/view/ilCarFugi1_1/dataset/ilCarFugi1_1/busco).



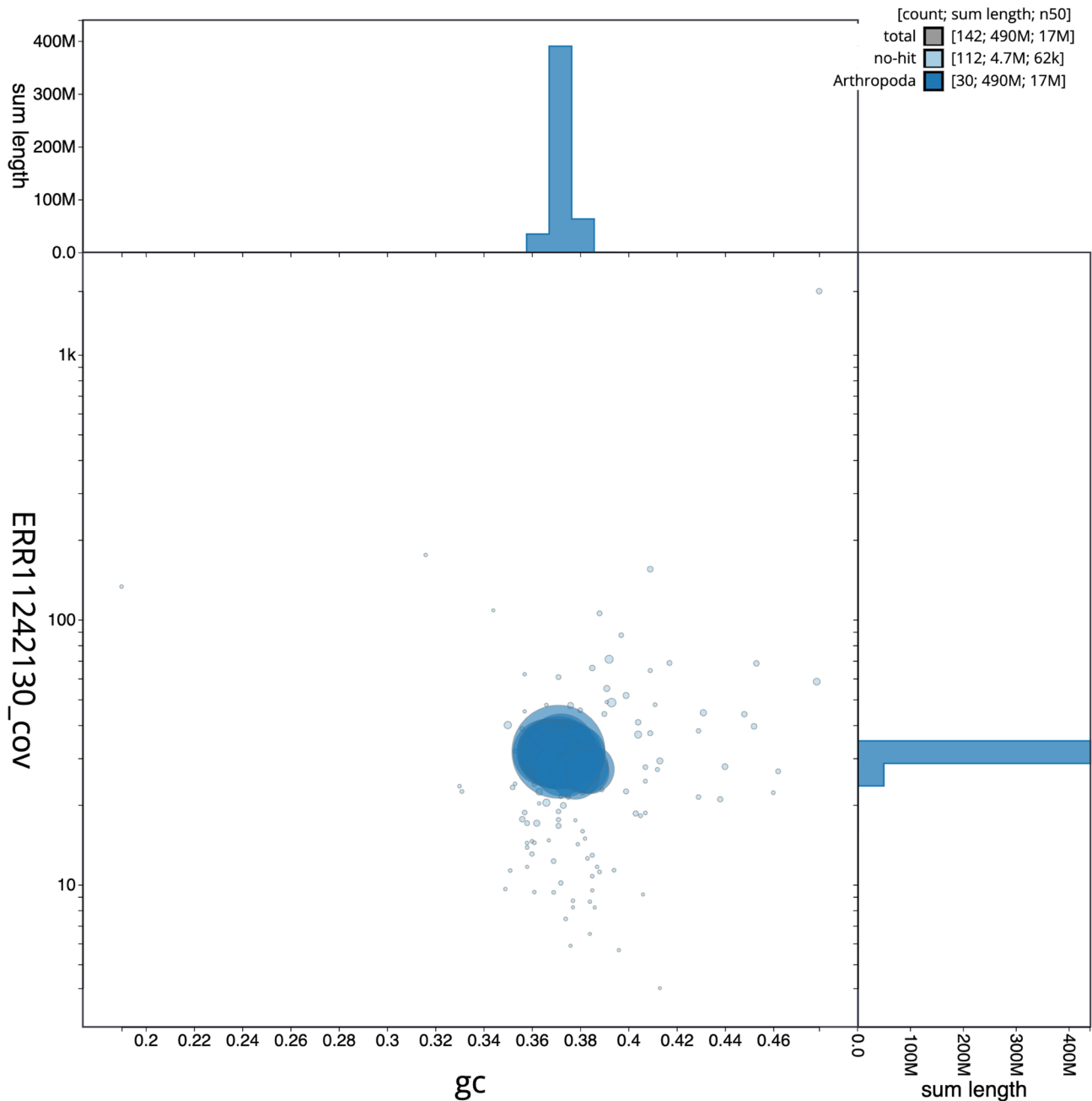
**Figure 2. Genome assembly of *Carpatolechia fugitivella*, ilCarFugi1.1: metrics.** The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 493,085,460 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (36,393,606 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (17,187,121 and 11,524,727 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/ilCarFugi1\\_1/dataset/ilCarFugi1\\_1/snail](https://blobtoolkit.genomehubs.org/view/ilCarFugi1_1/dataset/ilCarFugi1_1/snail).

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 ilCarFugi1 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 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.

RNA was extracted from tissue of ilCarFugi2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mir-Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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



**Figure 3. Genome assembly of *Carpatolechia fugitivella*, ilCarFugi1.1: 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/ilCarFugi1\\_1/dataset/ilCarFugi1\\_1/blob](https://blobtoolkit.genomehubs.org/view/ilCarFugi1_1/dataset/ilCarFugi1_1/blob).

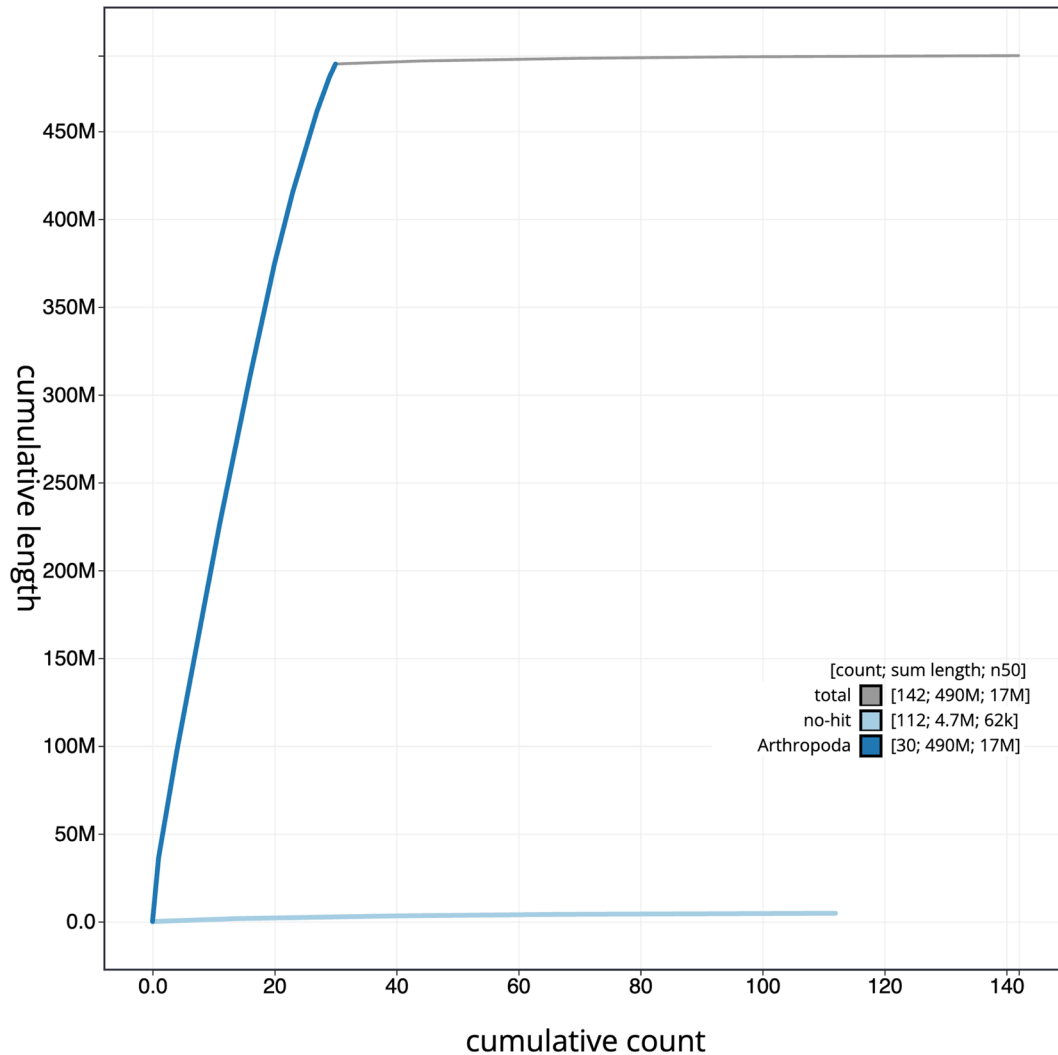
### Sequencing

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 Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data

were also generated from whole organism tissue of ilCarFugi2 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



**Figure 4. Genome assembly of *Carpatolechia fugitivella*, ilCarFugi1.1: 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/ilCarFugi1\\_1/dataset/ilCarFugi1\\_1/cumulative](https://blobtoolkit.genomehubs.org/view/ilCarFugi1_1/dataset/ilCarFugi1_1/cumulative).

scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, 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.

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

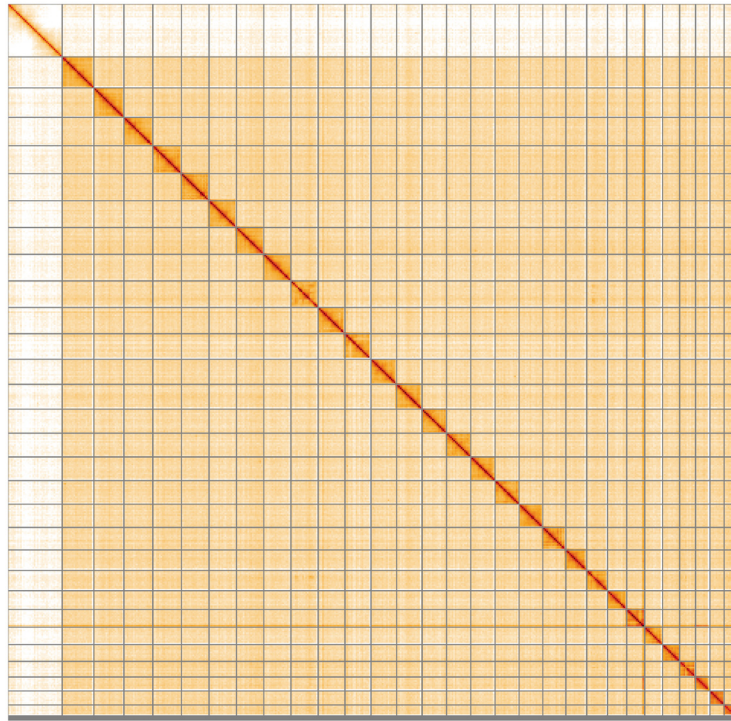
were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a) and “sanger-tol/genomenote” (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et 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 Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Carpatolechia fugitivella* assembly (GCA\_951230895.1) in Ensembl Rapid Release at the EBI. Annotation was created primarily through





**Figure 5. Genome assembly of *Carpatolechia fugitivella*, iCarFugi1.1: Hi-C contact map of the iCarFugi1.1 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=JhviKSQ4Q4iS0Tc8jZbMpg>.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Carpatolechia fugitivella*, iCarFugi1.**

INSDC accession	Chromosome	Length (Mb)	GC%
OX579646.1	1	21.25	37.0
OX579647.1	2	20.21	37.5
OX579648.1	3	19.44	37.0
OX579649.1	4	19.19	37.0
OX579650.1	5	18.51	37.5
OX579651.1	6	18.4	36.5
OX579652.1	7	18.38	37.0
OX579653.1	8	18.26	37.0
OX579654.1	9	18.22	37.0
OX579655.1	10	18.01	37.0
OX579656.1	11	17.5	37.0
OX579657.1	12	17.19	37.0
OX579658.1	13	17.11	37.0
OX579659.1	14	16.41	36.5

INSDC accession	Chromosome	Length (Mb)	GC%
OX579660.1	15	16.3	37.0
OX579661.1	16	16.27	37.0
OX579662.1	17	16.15	37.5
OX579663.1	18	15.91	37.0
OX579664.1	19	15.49	37.5
OX579665.1	20	13.88	37.5
OX579666.1	21	14.07	37.5
OX579667.1	22	12.93	37.0
OX579668.1	23	12.16	38.0
OX579669.1	24	11.9	37.0
OX579670.1	25	11.52	37.0
OX579671.1	26	10.68	38.0
OX579672.1	27	9.62	38.5
OX579673.1	28	9.55	38.0
OX579674.1	29	7.45	38.5
OX579645.1	Z	36.39	37.0
OX579675.1	MT	0.02	19.5



**Table 3. Software tools: versions and sources.**

Software tool	Version	Source
BlobToolKit	4.2.1	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.3.2	<a href="https://gitlab.com/ezlab/busco">https://gitlab.com/ezlab/busco</a>
Hifiasm	0.16.1-r375	<a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a>
HiGlass	1.11.6	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Mercury	MercuryFK	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>
MitoHiFi	3	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
PretextView	0.2	<a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a>
purge_dups	1.2.5	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
sanger-tol/genomenote	v1.0	<a href="https://github.com/sanger-tol/genomenote">https://github.com/sanger-tol/genomenote</a>
sanger-tol/readmapping	1.1.0	<a href="https://github.com/sanger-tol/readmapping/tree/1.1.0">https://github.com/sanger-tol/readmapping/tree/1.1.0</a>
YaHS	1.2a.2	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

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](#)).

#### 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: *Carpatolechia fugitivella* (elm groundling). Accession number PRJEB61343; <https://identifiers.org/ena.embl/PRJEB61343> (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Carpatolechia fugitivella* 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](#)
- 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](#)
- Boyes D, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab, et al.: **The genome sequence of the Dotted Grey Groundling, *Athrips mouffetella* (Linnaeus, 1758) [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2024; **9**: 42.  
[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](#)
- Chow W, Brugger K, Caccamo M, et al.: **gEVAL—a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.  
[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](#)
- do Amaral RJV, Bates A, Denton A, et al.: **Sanger Tree of Life RNA Extraction: Automated MagMax™ mirVana.** *protocols.io.* 2023.  
[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: ***Carpatolechia fugitivella* (Zeller, 1839).** *GBIF Backbone Taxonomy.* 2024; [Accessed 28 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.* 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](#)
- Langmaid JR, Palmer S, Young MR: **A Field Guide to the Smaller Moths of Great Britain and Ireland.** 3rd ed. British Entomological and Natural History Society, 2018.  
[Reference Source](#)
- Manni M, Berkeley MR, Seppy 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](#)
- 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](#)
- 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](#)
- Sterling P, Parsons M, Lewington R: **Field Guide to the Micro-moths of Great Britain and Ireland.** London: Bloomsbury Publishing, 2023.  
[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](#)
- Ulliano-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](#)
- 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](#)
- 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 Elm Groundling moth, *Carpatolechia fugitivella* (Zeller, 1839).** European Nucleotide Archive, [dataset], accession number PRJEB61343, 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](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 12 August 2024

<https://doi.org/10.21956/wellcomeopenres.23522.r92046>

© 2024 Hui 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.



**Jerome H L Hui** 

The Chinese University of Hong Kong, Hong Kong, Hong Kong

In this data note, Boyes and colleagues reported the genomic resource of a male moth *Carpatolechia fugitivella* (Zeller, 1839). According to the NatureSpot, UKMoths, and NBN, this species can be found in many places in Britain. Prior to this report, molecular data available for this species is limited to mitochondrial cytochrome oxidase subunit 1 (COI) gene sequences deposited to the NCBI database. Therefore, this new high-quality genome resource is valuable and will be useful for further studies, ranging from revealing their population structures, studying cross-kingdom interaction (e.g. insects/yeasts/plants), to understanding their evolution with other insects.

This genome resource is excellent from the summary statistics, with high BUSCO number scores, high sequence continuity (scaffold N50), and majority of sequences contained on the 29 pseudochromosomes (plus sex chromosome and mitochondrion). All in all, this is another Darwin Tree of Life's valuable contribution.

**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:** I have published with Peter Holland more than three years ago, and confirm that this potential conflict of interest did not affect my ability to write an objective and unbiased

review of the article.

**Reviewer Expertise:** Genomics, evolution, invertebrates

**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 06 August 2024

<https://doi.org/10.21956/wellcomeopenres.23522.r90424>

© 2024 Whibley A. 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.



**Annabel Whibley** 

<sup>1</sup> The University of Auckland, Auckland, Auckland, New Zealand

<sup>2</sup> Bragato Research Institute, Lincoln, New Zealand

In this Data Note, Boyes, Crowley, McCulloch, Boyes and colleagues present a reference genome assembly and gene annotation of the Elm Groundling Moth (*Carpatolechia fugitivella*). The reference is high-quality, was constructed using appropriate tools and with comprehensive reporting of the sample collection, data generation and analysis and all associated metadata. Links to data accessions have been tested and are working.

**Minor comments:**

Background: "It is local in England and Wales" reads oddly to me. Perhaps "It is locally common/abundant"?

Please specify longitude and latitude when giving co-ordinates.

Please use consistent formatting for dates (Sample acquisition for one sample is "2020-08-01" and for second is "2022-07-2").

I will continue to query whether this templated detail is correct: "in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA." This ratio of beads is not size-selective, I believe this should be 0.6x, as in the protocols.io guidelines

**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:** Genomics, Bioinformatics, Evolution

**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.**

---