DATA NOTE

The genome sequence of the Common Carpet moth, *Epirrhoe alternata* (Müller, 1764)

[version 1; peer review: 2 approved with reservations]

Douglas Boyes[1+], Liam M. Crowley[2], Finley Hutchinson[3], Denise C. Wawman[2], University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK
[3]University of Exeter, Penryn, England, UK

[+] Deceased author

v1 **First published:** 24 May 2024, **9**:293
https://doi.org/10.12688/wellcomeopenres.22257.1
Latest published: 24 May 2024, **9**:293
https://doi.org/10.12688/wellcomeopenres.22257.1

Abstract

We present a genome assembly from an individual male *Epirrhoe alternata* (the Common Carpet; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 358.5 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 16.99 kilobases in length.

Keywords

Epirrhoe alternata, common carpet, genome sequence, chromosomal, Lepidoptera

Open Peer Review

Approval Status ? ?

	1	2
version 1 24 May 2024	? view	? view

1. **Xiaoyi Dou**, University of Georgia, Athens, USA

2. **Sarah Inwood**, University of Otago, Dunedin, New Zealand

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Tree of Life gateway.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Boyes D: Investigation, Resources; **Crowley LM**: Investigation, Resources; **Hutchinson F**: Investigation, Resources; **Wawman DC**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Boyes D, Crowley LM, Hutchinson F *et al.* **The genome sequence of the Common Carpet moth, *Epirrhoe alternata* (Müller, 1764) [version 1; peer review: 2 approved with reservations]** Wellcome Open Research 2024, **9**:293 https://doi.org/10.12688/wellcomeopenres.22257.1

First published: 24 May 2024, **9**:293 https://doi.org/10.12688/wellcomeopenres.22257.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; Epirrhoe (Müller, 1764) (NCBI:txid190347).

Background

The Common Carpet *Epirrhoe alternata* is a moth in the family Geometridae with a forewing length of 13–14 mm. It is somewhat variable in appearance, with two subspecies found in the United Kingdom: *Epirrhoe alternata alternata* across most of the region and *Epirrhoe alternata obscurata* in the Outer Hebrides (Skinner & Wilson, 2009; Waring *et al.*, 2017). *Epirrhoe alternata alternata* is quite variable in appearance but has a series of brown and greyish-white bands running across the forewing: *E. alternata obscurata* is paler with less distinctive markings (Waring *et al.*, 2017).

Epirrhoe alternata overwinters as a pupa and adults are on the wing from April to early October with a peak in numbers occurring in August, presumably as the result of overlapping generations (Waring *et al.*, 2017). The larvae feed on cleavers *Galium aparine*, lady's bedstraw *Galium verum*, hedge bedstraw *Galium mollugo* and other related plants (Skinner & Wilson, 2009; Waring *et al.*, 2017).

In common with many moth species the numbers of *Epirrhoe alternata* in the British Isles are decreasing (Conrad *et al.*, 2006), but research has shown that restoration of abandoned pastures in Finland did not increase the abundance of this species as it did with some other species of Lepidoptera (Pöyry *et al.*, 2005).

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

Genome sequence report

The genome was sequenced from one male *Epirrhoe alternata* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 52-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data.

The final assembly has a total length of 358.5 Mb in 39 sequence scaffolds with a scaffold N50 of 13.0 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 (99.88%) of the 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). The Z chromosome was identified based on synteny with *Epirrhoe tristata* (GCA_951394285.1). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

Figure 1. Photograph of the *Epirrhoe alternata* (ilEpiAlte1) specimen used for genome sequencing.

The estimated Quality Value (QV) of the final assembly is 66.2 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v completeness of 98.6% (single = 98.1%, duplicated = 0.5%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/190347.

Methods

Sample acquisition and nucleic acid extraction

A male *Epirrhoe alternata* (specimen ID Ox000689, ToLID ilEpiAlte1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.34) on 2020-07-20 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

The specimen used for Hi-C sequencing (specimen ID Ox003078, ToLID ilEpiAlte2) was collected from the same location on 2022-07-22 using a light trap. The specimen was collected by Liam Crowley (University of Oxford) and Finley Hutchinson (University of Exeter), and identified by Finley Hutchinson, and then preserved on dry ice.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation,

Table 1. Genome data for *Epirrhoe alternata*, ilEpiAlte1.1.

Project accession data		
Assembly identifier	ilEpiAlte1.1	
Species	*Epirrhoe alternata*	
Specimen	ilEpiAlte1	
NCBI taxonomy ID	190347	
BioProject	PRJEB55954	
BioSample ID	SAMEA7701550	
Isolate information	ilEpiAlte1: whole organism (PacBio sequencing) ilEpiAlte2: whole organism (Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	66.2	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:98.6%[S:98.1%,D:0.5%],F:0.4%,M:1.0%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.88%	*≥ 95%*
Sex chromosomes	ZZ	*localised homologous pairs*
Organelles	Mitochondrial genome: 16.99 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences Sequel IIe	ERR10224902	
Hi-C Illumina	ERR12102371	
Genome assembly		
Assembly accession	GCA_963565295.1	
Accession of alternate haplotype	GCA_963565885.1	
Span (Mb)	358.5	
Number of contigs	40	
Contig N50 length (Mb)	13.0	
Number of scaffolds	39	
Scaffold N50 length (Mb)	13.0	
Longest scaffold (Mb)	16.78	

* 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 v5.4.3. 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/Epirrhoe_alternata/dataset/GCA_963565295.1/busco.

and clean-up. In sample preparation, the ilEpiAlte1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023).

Tissue of 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

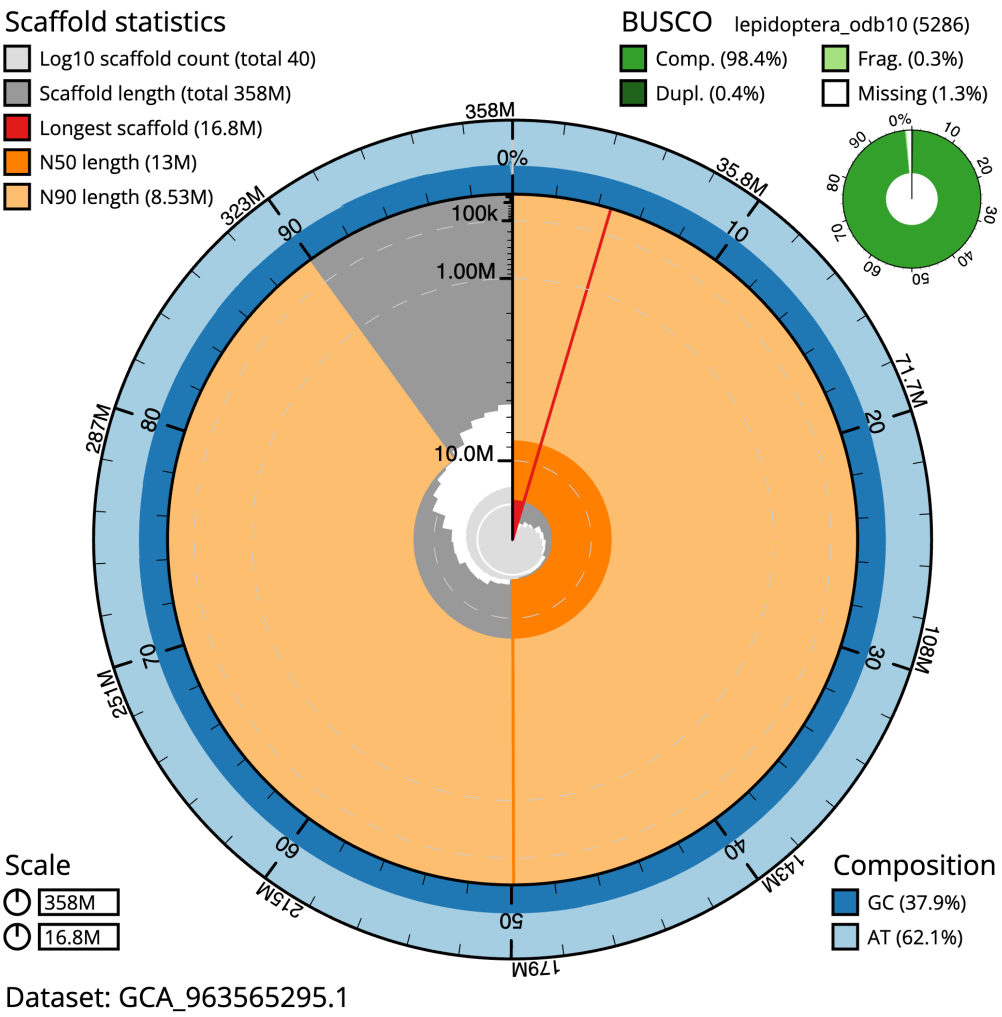

Figure 2. Genome assembly of *Epirrhoe alternata*, ilEpiAlte1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 358,495,576 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (16,783,756 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (13,027,933 and 8,529,277 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/Epirrhoe_alternata/dataset/GCA_963565295.1/snail.

and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

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

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences Sequel IIe instrument. Hi-C data were also generated from tissue of ilEpiAlte2 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly and curation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was performed using JBrowse2 (Diesh *et al.*, 2023), HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using

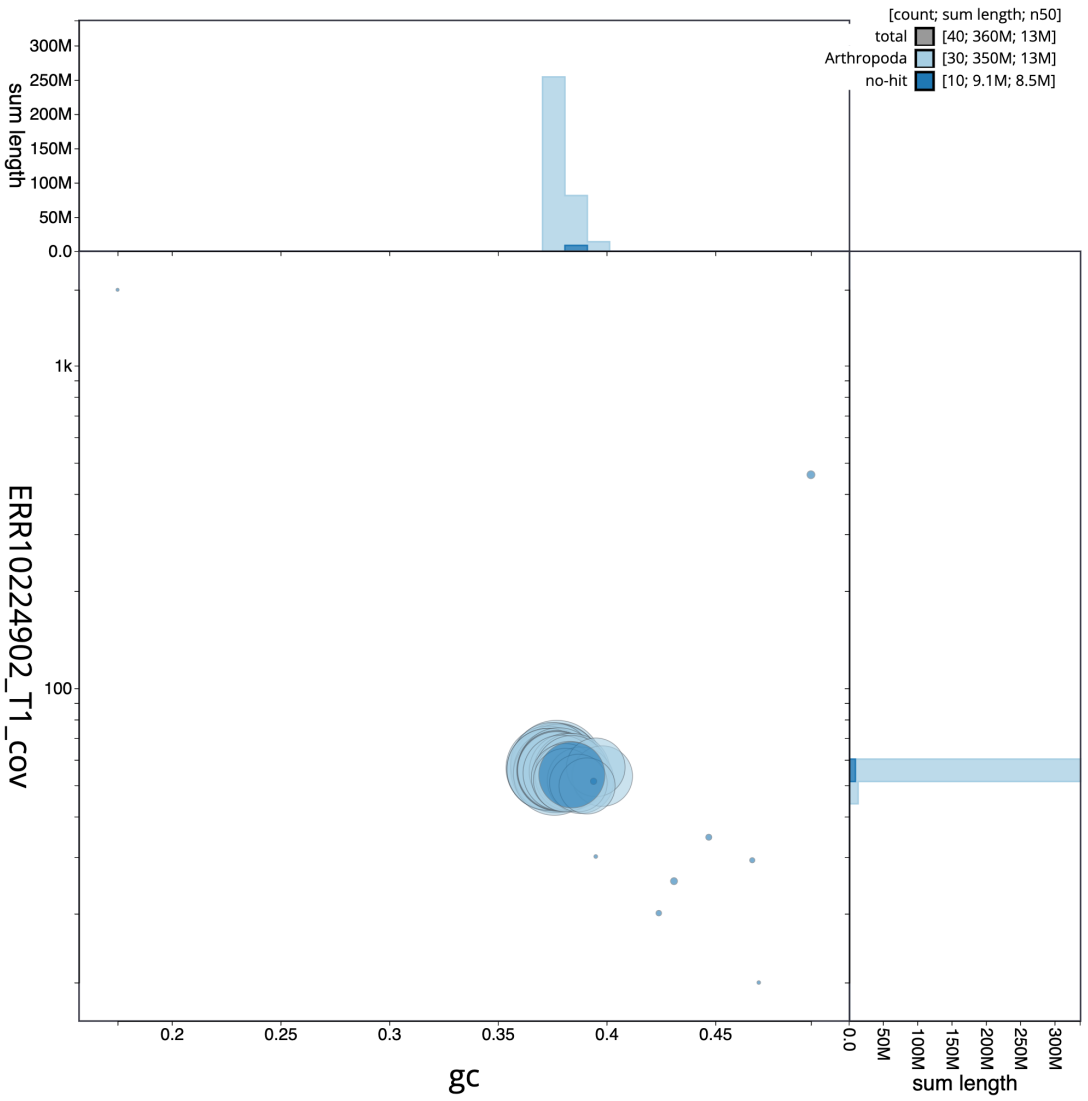

Figure 3. Genome assembly of *Epirrhoe alternata*, ilEpiAlte1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Epirrhoe_alternata/dataset/GCA_963565295.1/blob.

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.

Final assembly evaluation

The final assembly was post-processed and evaluated with the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a), "sanger-tol/genomenote" (Surana *et al.*, 2023b), and "sanger-tol/blobtoolkit" (Muffato *et al.*, 2024). The pipeline sanger-tol/readmapping aligns the Hi-C reads with bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The sanger-tol/genomenote pipeline transforms the Hi-C alignments into a contact map with BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020), which is then visualised with HiGlass (Kerpedjiev *et al.*, 2018). It also provides statistics about the assembly with the NCBI datasets (Sayers *et al.*, 2024) report, computes *k*-mer completeness and QV consensus quality values with FastK and MerquryFK, and a completeness assessment with BUSCO (Manni *et al.*, 2021).

The sanger-tol/blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads with SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO

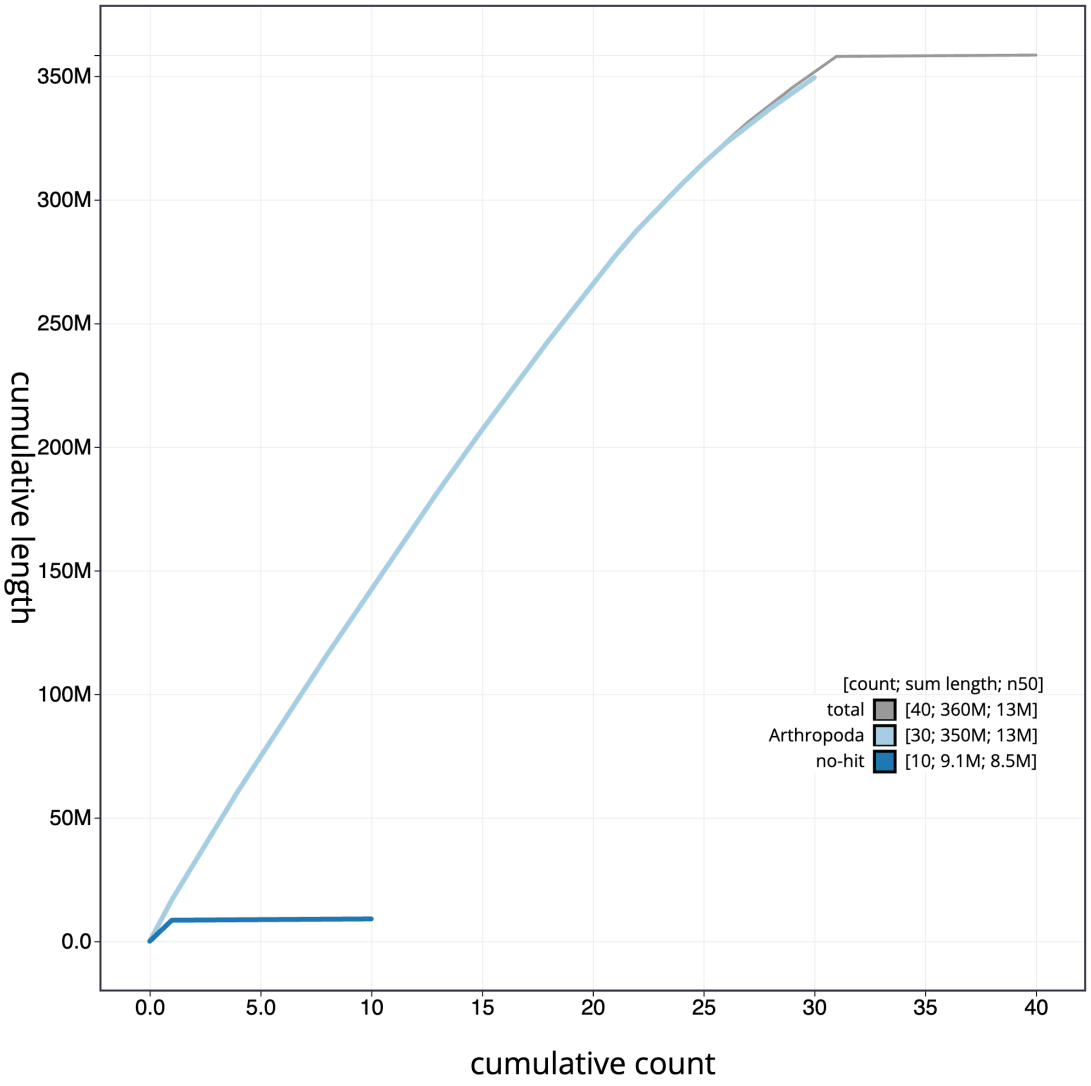

Figure 4. Genome assembly of *Epirrhoe alternata*, ilEpiAlte1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Epirrhoe_alternata/dataset/GCA_963565295.1/cumulative.

lineage, the pipeline aligns the BUSCO genes to the Uniprot Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND (Buchfink *et al.*, 2021) blastp. The genome is also split into chunks according to the density of the BUSCO genes from the closest taxonomically lineage, and each chunk is aligned to the Uniprot Reference Proteomes database with DIAMOND blastx. Genome sequences that have no hit are then chunked with seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). All those outputs are combined with the blobtools suite into a blobdir for visualisation.

All three pipelines were developed using the nf-core tooling (Ewels *et al.*, 2020), use MultiQC (Ewels *et al.*, 2016), and make extensive use of the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), and the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 3 contains a list of relevant software tool versions and sources.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'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

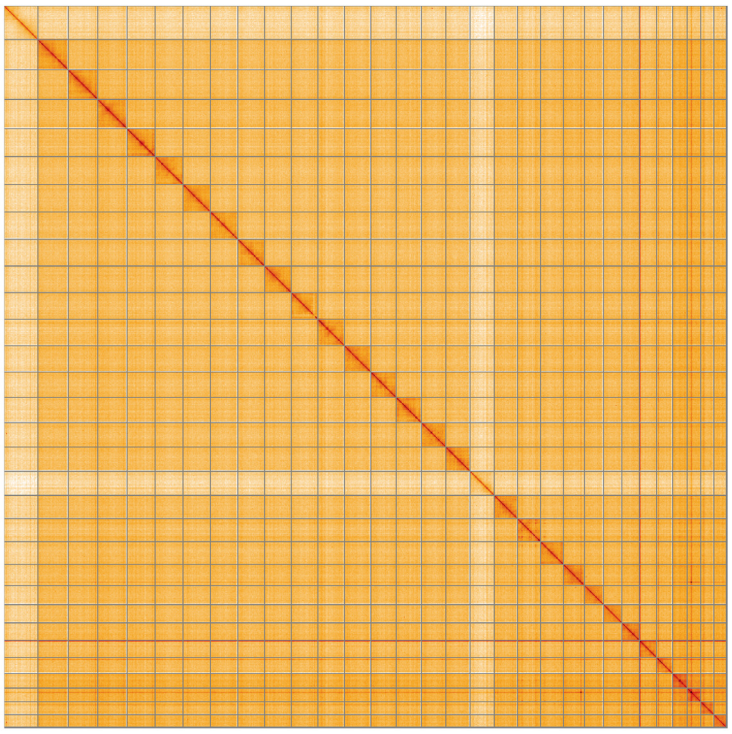

Figure 5. Genome assembly of *Epirrhoe alternata*, ilEpiAlte1.1: Hi-C contact map of the ilEpiAlte1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=aQNjFNvqRQ2dZ5-nWk-tyQ.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Epirrhoe alternata*, ilEpiAlte1.

INSDC accession	Chromosome	Length (Mb)	GC%
OY751376.1	1	15.02	37.5
OY751377.1	2	14.66	38.0
OY751378.1	3	14.5	38.0
OY751379.1	4	13.92	38.0
OY751380.1	5	13.8	37.5
OY751381.1	6	13.62	37.5
OY751382.1	7	13.56	37.5
OY751383.1	8	13.26	37.5
OY751384.1	9	13.23	37.5
OY751385.1	10	13.21	37.5
OY751386.1	11	13.19	37.5
OY751387.1	12	13.03	38.0
OY751388.1	13	12.59	38.0
OY751389.1	14	12.49	37.5
OY751390.1	15	12.09	37.5
OY751391.1	16	12.0	37.5
OY751392.1	17	11.98	37.5
OY751393.1	18	11.52	38.0
OY751394.1	19	11.43	38.5
OY751395.1	20	11.3	38.0
OY751396.1	21	10.5	38.5
OY751397.1	22	9.46	38.5
OY751398.1	23	9.04	38.0
OY751399.1	24	8.68	39.0
OY751400.1	25	8.53	38.5
OY751401.1	26	7.96	38.0
OY751402.1	27	7.19	40.0
OY751403.1	28	6.75	39.5
OY751404.1	29	6.5	38.5
OY751405.1	30	6.17	39.0
OY751375.1	Z	16.78	37.5
OY751406.1	MT	0.02	17.5

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
Blast	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/genomenote	1.1.1	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.2.1	https://github.com/sanger-tol/readmapping
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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: *Epirrhoe alternata* (common carpet). Accession number PRJEB55954; https://identifiers.org/ena.embl/PRJEB55954 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Epirrhoe alternata* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.7125292.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the universal protein knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
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

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Conrad KF, Warren MS, Fox R, *et al.*: **Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis.** *Biol Conserv.* 2006; **132**(3): 279–291.
Publisher Full Text

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

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

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
PubMed Abstract | Publisher 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

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

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
Reference Source

Muffato M, Butt Z, Challis R, *et al.*: **sanger-tol/blobtoolkit: v0.3.0 – Poliwag.** 2024.
Publisher Full Text

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

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

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
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

Sayers EW, Cavanaugh M, Clark K, *et al.*: **GenBank 2024 update.** *Nucleic Acids Res.* 2024; **52**(D1): D134–D137.
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

Skinner B, Wilson D: **Colour identification guide to the moths of the British isles.** Leiden, Netherlands: Brill, 2009.
Reference Source

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

Vasimuddin Md, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems**. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Waring P, Townsend M, Lewington R: **Field guide to the moths of Great Britain and Ireland: Third edition.** Bloomsbury Wildlife Guides, 2017.
Reference Source

Wellcome Sanger Institute: **The genome sequence of the Common Carpet moth, *Epirrhoe alternata* (Müller, 1764).** European Nucleotide Archive, [dataset], accession number PRJEB55954, 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 04 September 2024

https://doi.org/10.21956/wellcomeopenres.24534.r94487

?

Sarah Inwood
University of Otago, Dunedin, Otago, New Zealand

The authors sequenced and assembled a chromosomal scale genome assembly for the Common Carpet moth for *Epirrhoe alternata*. This assembly used HiFi long reads from a male individual, and was scaffolded using Hi-C data from a second sample of unspecified sex, resulting in a high-quality assembly, alongside a mitochondrial genome assembly, using technically sound methodology.

Minor comments:
The significance of having a genome available for *E. alternata* is not made immediately clear to readers.

Despite the mention of two subspecies of *E. alternata* in the United Kingdom, it is not specified which subspecies was sequenced, which is instead left to the reader to infer. Is much known about potential genetic differences between the two subspecies in the United Kingdom?

The sex of the second sample used for Hi-C sequencing is not reported.

For multiple software tools in Table 3, e.g. BUSCO, multiple version numbers are reported. It is unclear to the reader the reason for multiple versions of the same tool to have been used, and at which stages of the analyses different versions were used. This reduces the replicability of this workflow. Alongside this, the authors state the use of MitoHifi, which can run MitoFinder OR MITOS, but do not state which tool resulted in their deposited mitochondrial genome assembly.

The significance of this assembly, the research gap this fills, and the importance of this data are not clearly stated.

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: Genomics, bioinformatics, host-parasite interactions, parasitoid wasps, insects, viruses, biocontrol

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, however I have significant reservations, as outlined above.

Reviewer Report 06 August 2024

https://doi.org/10.21956/wellcomeopenres.24534.r89297

?

Xiaoyi Dou
Department of Entomology, University of Georgia, Athens, USA

The paper presents a chromosomal-level genome assembly for Epirrhoe alternata. The assembly, based on a single male specimen, provides a comprehensive genetic map that includes both nuclear and mitochondrial DNA sequences. The work is significant for understanding the genetic makeup of this species, contributing to the broader goals of the insect biology. However, some essential parts need to be added

- In abstract, it is better to have a brief description of the methods and an overall conclusion
- In introduction, it is important to claim the research gap and significant of this study.
- A flowchart to summarize the workflow would be better
- A conclusion or discussion showing the advances or importance of the workflow or data would be btter.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, endocrinology, infochemicals

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, however I have significant reservations, as outlined above.