DATA NOTE

The genome sequence of the lesser treble-bar moth, *Aplocera efformata* (Guenée, 1857) [version 1; peer review: 3 approved]

Douglas Boyes[1+],
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, Marc Palmada-Flores[2],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]Department of Medicine and Life Sciences (MELIS), Institut de Biologia Evolutiva, Universitat Pompeu Fabra-CSIC, Barcelona, Spain, 08003, Spain

[+] Deceased author

v1 **First published:** 15 Dec 2022, **7**:303
https://doi.org/10.12688/wellcomeopenres.18595.1
Latest published: 15 Dec 2022, **7**:303
https://doi.org/10.12688/wellcomeopenres.18595.1

Abstract

We present a genome assembly from an individual female *Aplocera efformata* (the lesser treble-bar; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 349.5 megabases in span. Most of the assembly (99.97%) is scaffolded into 32 chromosomal pseudomolecules, with W and Z sex chromosomes assembled. The complete mitochondrial genome was also assembled and is 15.4 kilobases in length.

Keywords

Aplocera efformata, lesser treble-bar, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 15 Dec 2022	✓ view	✓ view	✓ view

1. **Jaakko L.O. Pohjoismäki**, University of Eastern Finland, Joensuu, Finland
2. **Niklas Wahlberg**, Lund university, Lund, Sweden
3. **Muzafar Riyaz**, St Xavier's College, Palayamkottai, India

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; **Palmada-Flores M**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194, https://doi.org/10.35802/206194) and the Darwin Tree of Life Discretionary Award (218328, https://doi.org/10.35802/218328). This work received the support of an INPhINIT Retaining Fellowship from "La Caixa" Foundation (ID 100010434) with code LCF/BQ/DR20/11790032.
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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the lesser treble-bar moth, *Aplocera efformata* (Guenée, 1857) [version 1; peer review: 3 approved]** Wellcome Open Research 2022, **7**:303 https://doi.org/10.12688/wellcomeopenres.18595.1

First published: 15 Dec 2022, **7**:303 https://doi.org/10.12688/wellcomeopenres.18595.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Geometroidea; Geometridae; Larentiinae; *Aplocera*; *Aplocera efformata* (Guenée, 1857) (NCBI:txid934917).

Background

The lesser treble-bar, *Aplocera efformata* (Guenée, 1857), is a geometer moth within the subfamily Larentiinae (family Geometridae) composed of carpets, pugs and allies (Waring & Townsend, 2017). It is hard to distinguish from its sister species, the treble-bar (*Aplocera plagiata*), as both are grey with three dark cross-bands in their pointed forewings. However, the lesser treble-bar species is slightly smaller, with a forewing length of 16–19 mm, and displays less intense dark cross-bands and lighter forewings. Its abdomen also has a shorter taper to the apex compared to the very pointed abdomen of the treble-bar (Townsend *et al.*, 2010; Waring & Townsend, 2017).

The lesser treble-bar's range extends from Morocco across southern and central Europe, reaching Anatolia to the east and southern Scandinavia to the north (Bálint *et al.*, 2016).

The preferred habitat of *A. efformata* is hot, dry grasslands, mainly on sandy or calcareous ground, though it is sometimes encountered in regions such as sea-cliffs, woodland rides, abandoned quarries, field margins and gardens. *A. efformata* presents two generations of flight seasons, which are easily disturbed by day, overwinters as larvae and pupates underground (Bálint *et al.*, 2016; Waring & Townsend, 2017).

In Europe, the species has been suffering a decline in population, being threatened by the diminution of their favoured habitat (Bálint *et al.*, 2016). We predict that the Darwin Tree of Life assembly presented here will be an important tool for further examination of its population dynamics.

Genome sequence report

The genome was sequenced from a single female *A. efformata* (Figure 1) collected from Wytham Woods, Berkshire, UK (latitude 51.772, longitude –1.338). A total of 53-fold coverage in Pacific Biosciences single-molecule circular consensus (HiFi) long reads and 128-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected four misjoins which reduced the scaffold number by 7.27%.

Figure 1. Image of the female *Aplocera efformata* specimen from which the genome was sequenced. The ilAplEffo1 specimen was used to generate Pacific Biosciences, 10X genomics, Hi-C and RNA-Seq data.

The final assembly has a total length of 349 Mb in 51 sequence scaffolds with a scaffold N50 of 12.5 Mb (Table 1). Most of the assembly sequence (99.97%) was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length) and the W and Z sex chromosomes (Figure 2–Figure 5; Table 2).

Table 1. Genome data for *A. efformata*, ilAplEffo1.1.

Project accession data	
Assembly identifier	ilAplEffo1.1
Species	*Aplocera efformata*
Specimen	ilAplEffo1 (genome assembly, Hi-C, RNA-Seq)
NCBI taxonomy ID	934917
BioProject	PRJEB47323
BioSample ID	SAMEA8603170
Isolate information	Female, thorax tissue (genome assembly), head tissue (Hi-C), abdomen tissue (RNA-Seq)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6939267
10X Genomics Illumina	ERR6688760-ERR6688763
Hi-C Illumina	ERR6688759
PolyA RNA-Seq Illumina	ERR9435023
Genome assembly	
Assembly accession	GCA_921293045.1
Accession of alternate haplotype	GCA_921293035.1
Span (Mb)	350
Number of contigs	55
Contig N50 length (Mb)	12.5
Number of scaffolds	51
Scaffold N50 length (Mb)	12.5
Longest scaffold (Mb)	15.0
BUSCO* genome score	C:98.4%[S:98.1%,D:0.3%],F:0.4%,M:1.2%,n:5,286
Genome annotation	
Number of protein-coding genes	11,393

*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/ilAplEffo1.1/dataset/CAKLCP01.1/busco.

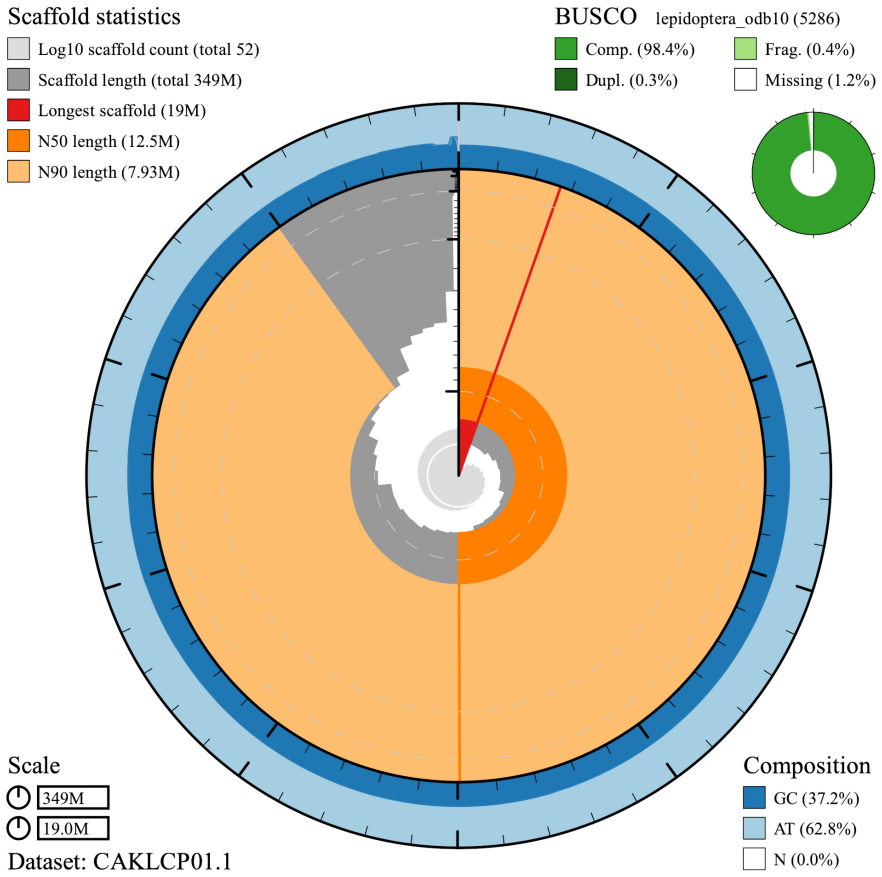

Figure 2. Genome assembly of *Aplocera efformata*, ilAplEffo1.1: metrics. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 349,498,550 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (19,009,616 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (12,527,553 and 7,930,759 bp), respectively. The pale grey spiral shows the cumulative chromosome 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/ilAplEffo1.1/dataset/CAKLCP01.1/snail.

The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.4% (single 98.1%, duplicated 0.3%) using the lepidoptera_odb10 reference set (n = 5,286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Genome annotation report

The GCA_921293045.1 genome was annotated using the Ensembl rapid annotation pipeline (Table 1). The resulting annotation includes 19,297 transcribed mRNAs from 11,393 protein-coding and 1,074 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A single female *A. efformata* specimen (ilAplEffo1) was collected in Wytham Woods, Berkshire, UK (latitude 51.772, longitude –1.338) by Douglas Boyes (University of Oxford), using a light trap. The sample was identified by Douglas Boyes and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The ilAplEffo1 sample was weighed and dissected on dry ice with head tissue set aside for Hi-C sequencing. Thorax tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200 ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the 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

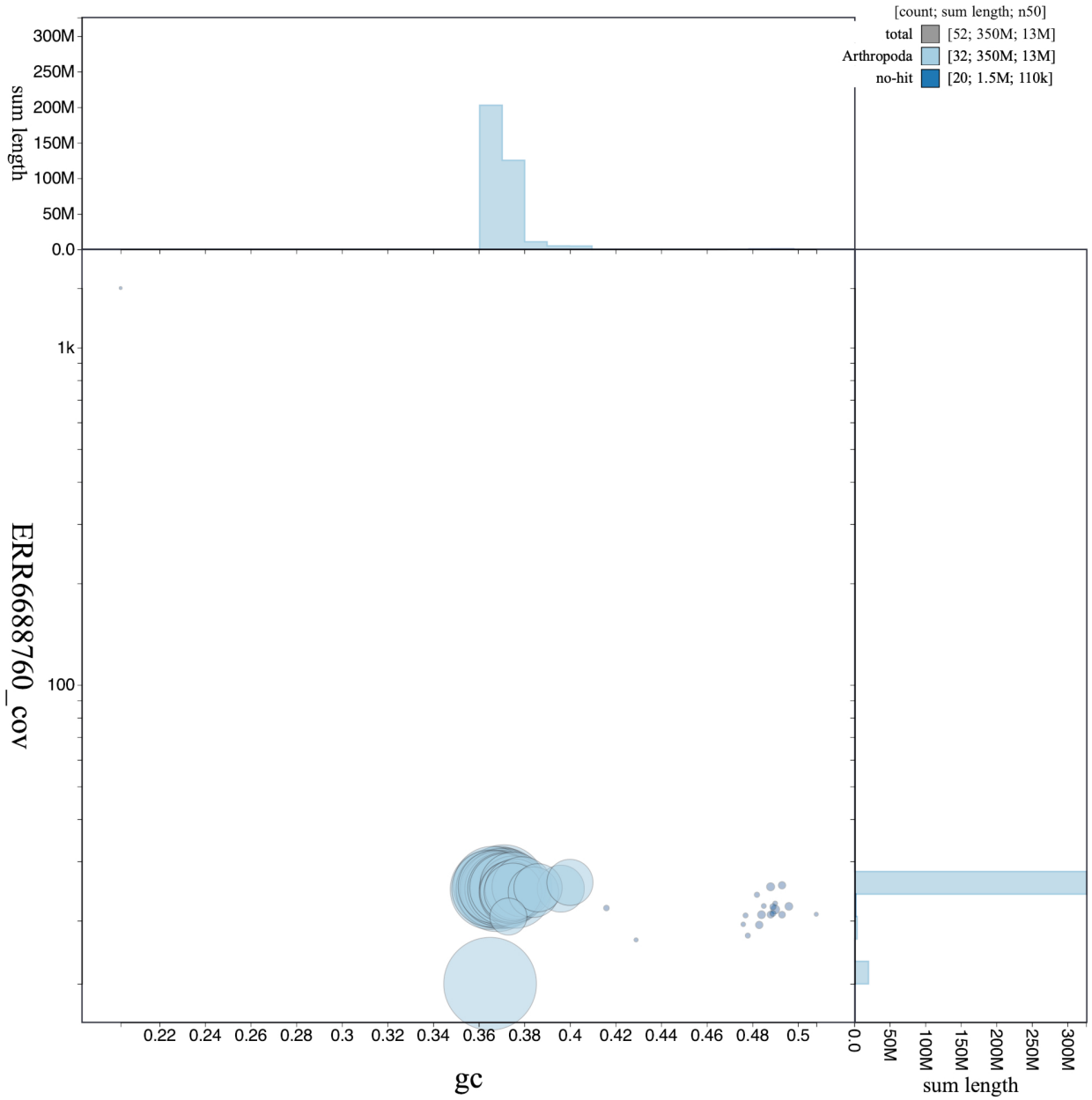

Figure 3. Genome assembly of *Aplocera efformata*, ilAplEffo1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilAplEffo1.1/dataset/CAKLCP01.1/blob.

distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from the abdomen tissue of ilAplEffo1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and the RNA concentration was 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 and 10X Genomics read cloud 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

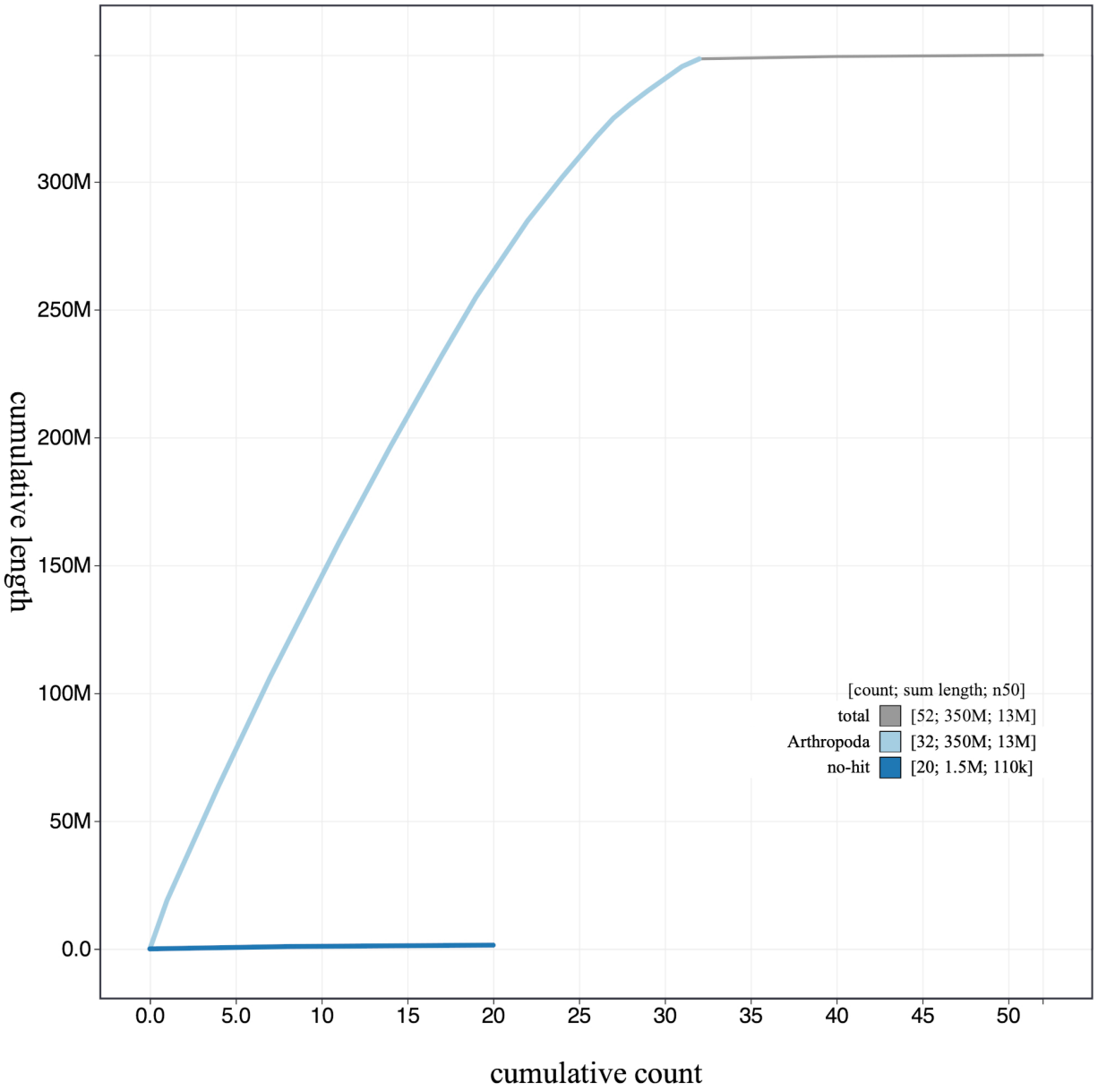

Figure 4. Genome assembly of *Aplocera efformata*, ilAplEffo1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilAplEffo1.1/dataset/CAKLCP01.1/cumulative.

the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina NovaSeq 6000 and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were also generated from the remaining head tissue of ilAplEffo1 using the Arima v2 Hi-C kit and sequenced on an Illumina NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performs annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020), generating BUSCO scores. Table 3 contains a list of all software tool versions used, where appropriate.

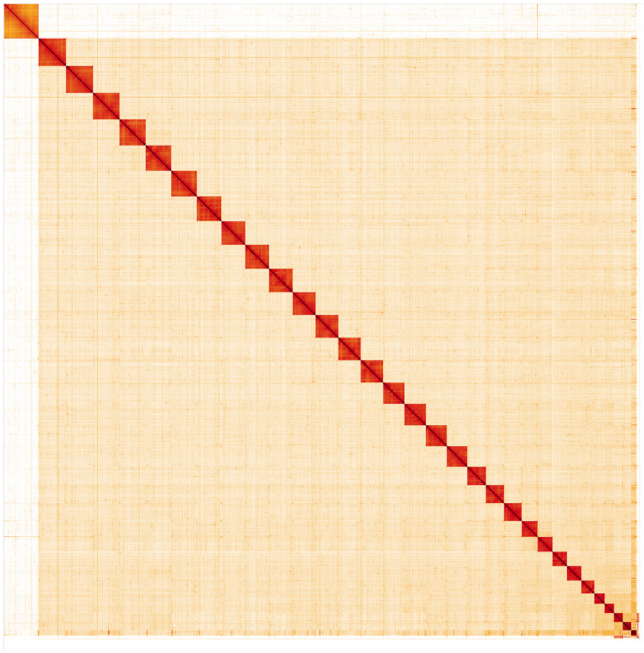

Figure 5. Genome assembly of *Aplocera efformata*, ilAplEffo1.1: Hi-C contact map. Hi-C contact map of the ilAplEffo1.1 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=DFH-u6PYSz6oMane3MisUg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *A. efformata*, ilAplEffo1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OV121039.1	1	15.04	37.1
OV121040.1	2	15.03	36.9
OV121041.1	3	14.63	36.8
OV121042.1	4	14.27	37.1
OV121043.1	5	14.13	37
OV121044.1	6	14.06	36.5
OV121045.1	7	13.39	36.6
OV121046.1	8	13.18	36.7
OV121047.1	9	13.08	36.7
OV121048.1	10	12.96	36.5
OV121049.1	11	12.55	37
OV121050.1	12	12.53	37
OV121051.1	13	12.41	36.9
OV121052.1	14	12.18	36.8
OV121053.1	15	11.81	37
OV121054.1	16	11.73	36.7
OV121055.1	17	11.5	37.1
OV121056.1	18	11.44	37.6
OV121057.1	19	10.22	37.1
OV121058.1	20	9.95	37.5
OV121059.1	21	9.61	37.5
OV121060.1	22	8.7	37.4
OV121061.1	23	8.21	37.4
OV121062.1	24	8	37.6
OV121063.1	25	7.93	37.9
OV121064.1	26	7.3	37.5
OV121065.1	27	4.84	39.6
OV121066.1	28	5.59	38.4
OV121067.1	29	5.09	38.6
OV121068.1	30	4.67	40
OV121069.1	W	2.98	37.3
OV121038.1	Z	19.01	36.5
OV121070.1	MT	0.02	20.4
-	Unplaced	1.45	48.4

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	3.3.2	(Challis *et al.*, 2020)
freebayes	1.3.1-17-gaa2ace8	(Garrison & Marth, 2012)
Hifiasm	0.15.3	(Cheng *et al.*, 2021)
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
MitoHiFi	2.0	(Uliano-Silva *et al.*, 2021)
PretextView	0.1.x	(Harry, 2022)
purge_dups	1.2.3	(Guan *et al.*, 2020)
SALSA2	2.2	(Ghurye *et al.*, 2019)

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *A. efformata* assembly (GCA_921293045.1). 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)).

Data availability

European Nucleotide Archive: *Aplocera efformata* (lesser treble-bar). Accession number PRJEB47323; https://identifiers.org/ena.embl/PRJEB47323 (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *A. efformata* 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 the 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.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

Bálint Z, Katona G, Ronkay L: **Data to the Knowledge of the Macrolepidoptera Fauna of the Sălaj-region, Transylvania, Romania (Arthropoda: Insecta).** *Studia Univ Vasile Goldis.* 2016; **26**(1): 59–74.
Reference Source

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

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

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing**. 2012.
Publisher Full Text

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps**. 2022; (Accessed: 19 October 2022).
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153.
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, 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**

Rao SS, 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**

Townsend MC, Clifton J, Goodey B: **British and Irish Moths: An Illustrated Guide to Selected Difficult Species**. Butterfly Conservation, 2010.
Reference Source

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: **MitoHiFi**. 2021; (Accessed: 19 October 2022).
Reference Source

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: **Field Guide to the Moths of Great Britain and Ireland: Third Edition**. Bloomsbury, 2017.
Reference Source

Wellcome Sanger Institute: **The genome sequence of the lesser treble-bar moth, Aplocera efformata (Guenée, 1857), European Nucleotide Archive [dataset]**. 2022; accession number PRJEB47323.

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 1

Reviewer Report 08 November 2023

https://doi.org/10.21956/wellcomeopenres.20620.r69217

✔ **Muzafar Riyaz**
Xavier Research Foundation, St Xavier's College, Palayamkottai, Tamil Nadu, India

The article titled "The genome sequence of the lesser treble-bar moth, Aplocera efformata (Guenée, 1857)" presents a comprehensive analysis of the genome of Aplocera efformata, shedding light on the genetic makeup of this lesser-known species within the Lepidoptera order. The study provides valuable insights into the evolutionary and ecological aspects of this geometer moth, making it a significant contribution to the field of genomics and entomology.

The authors meticulously detail the methodology employed in genome sequencing, which involved a combination of advanced technologies such as Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries. The assembly process, outlined in the article, is notable for its depth and precision, resulting in a high-quality genome assembly of A. efformata. The inclusion of Hi-C data in scaffolding the assembly ensures the chromosomal-level accuracy, which is crucial for understanding the genetic architecture and chromosomal organization of the species.

The article provides a thorough overview of the lesser treble-bar moth, its taxonomy, habitat, and distinguishing features, setting the stage for the importance of genomic analysis in understanding its population dynamics and evolutionary history. The detailed genome annotation, involving alignment of transcriptomic data and protein-to-genome alignments, enhances our understanding of the genetic elements within this species, including protein-coding and non-coding genes.

One of the key strengths of the article lies in its commitment to transparency and accessibility. The authors have made the genome sequence openly available, promoting further research and ensuring the data's usability by the scientific community. The comprehensive data availability section, including accession numbers and relevant links, enhances the reproducibility and credibility of the study. Additionally, the article effectively integrates visualizations, such as Hi-C contact maps and genome assembly metrics, enhancing the reader's understanding of the methodology and results. These visual aids contribute to the clarity of the presented information.

The genome sequence of the lesser treble-bar moth, Aplocera efformata (Guenée, 1857)" stands out as a well-executed genomic study, providing valuable insights into the lesser treble-bar moth's genetic makeup. The meticulous methodology, detailed annotations, and transparent data sharing make this article a significant resource for researchers in genomics, evolutionary biology, and entomology.

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, Phylogenomics, Mitogenomics, and Evolutionary Biology

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 November 2023

https://doi.org/10.21956/wellcomeopenres.20620.r69223

Niklas Wahlberg
Lund university, Lund, Sweden

The report on the genome of *Aplocera efformata* (Lepidoptera: Geometridae) follows the standard format for genome notes, of which more than 200 have been published already for butterflies and moths. This is an amazing resource for the scientific community! In this particular case, the genome of *A. efformata* will be interesting to compare to the closely related *A. plagiata*, from which it is difficult to distinguish (as mentioned in the article).

The methods section has been written clearly, and it is good that the format has developed, with all the programs being listed along with the version that was used for assembling the genome. I have nothing more to add, it seems that Lepidoptera genomes are easy to sequence and easy to assemble.

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 am a founding member of the Psyche project, which aims to sequence the genomes of all Lepidoptera in Europe.

Reviewer Expertise: Phylogenomics, systematics, Lepidoptera, Geometridae

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 28 March 2023

https://doi.org/10.21956/wellcomeopenres.20620.r55797

Jaakko L.O. Pohjoismäki
Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland

The genome release of the lesser treble-bar moth by Douglas Boyes and the DToL crew is another nice addition to the growing number of Lepidopteran reference genomes. Also, the presented species is not that trivial and could offer some interesting insights into taxonomy and biology of these moths.

The provided metrics seem in agreement with what is required from reference genomes, but my understanding of the bioinformatic work is not sufficient to evaluate the used approach further. Instead, I try to contribute some thoughts for the taxonomic and faunistic parts of the report.

Background
You could point out that *A. efformata* is a monophagous species on *Hypericum*. This might be relevant, as monophagy vs polyphagy could be somehow reflected by the insect genomes (adaptations to plant secondary compounds etc). Perhaps something to address in the future? If you have some speculation to offer about this (adaptive signatures, specialized detoxifying

enzymes etc), please feel free to add a couple of sentences at the end of the section.

[...] geometer moth [...] -> geometrid moth

[...] decline in population [...] The sentence reads odd. Please reformulate.

Methods:
The trapping date is not provided.

A representative of the heterogametic sex was analyzed, which is great as it is not trivial with night active moths (males more often caught on light). However, please assign a voucher specimen for the genome and provide details of its location (name of the collection, storing institute storing) for future reference.

I would have chosen thorax for RNA as there are probably more tissue types present. Moreover, the female abdomen is often full of eggs (note that the specimen looks relatively young, wings and scales intact, so probably has not laid its eggs yet) that mainly have a limited set of maternal RNAs for the early embryonic development. Was this reflected by the RNA-seq data? If so, comment this in the genome annotation report.

As pointed out in the background chapter, *A. efformata* is very similar to *plagiata*. While I trust the skill of your taxonomic expert, I nevertheless extracted the *Co1* sequence from the mitochondrial genome you provided and it seems to agree with *efformata*. The authors could point out that the species identity can be confirmed in this way using DNA-barcoding (95.5 % pairwise identity between the *Co1* sequence from the two species, so a nice separation). It would also not harm to mention here the BINs for the different *Aplocera* species in BOLD for comparison.

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: Molecular biology, molecular ecology and genetics, taxonomy, DNA barcoding

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.