



DATA NOTE

The genome sequence of the Dotted Grey Groundling, *Athrips mouffetella* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes¹⁺, Clare Boyes^{id}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

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

²Independent researcher, Welshpool, Wales, UK

+ Deceased author

V1 First published: 19 Feb 2024, 9:42
<https://doi.org/10.12688/wellcomeopenres.20840.1>
Latest published: 19 Feb 2024, 9:42
<https://doi.org/10.12688/wellcomeopenres.20840.1>

Abstract



We present a genome assembly from an individual female *Athrips mouffetella* (the Dotted Grey Groundling; Arthropoda; Insecta; Lepidoptera; Gelechiidae). The genome sequence is 869.7 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z and W sex chromosomes. The mitochondrial genome has also been assembled and is 15.23 kilobases in length. Gene annotation of this assembly on Ensembl identified 22,889 protein coding genes.

Keywords

Athrips mouffetella, dotted grey groundling, genome sequence, chromosomal, Lepidoptera

Open Peer Review

Approval Status  

	1	2
version 1 19 Feb 2024	 view	 view

1. **Melody Clark** , Natural Environment Research Council, Cambridge, UK
2. **Paul Mireji** , Kenya Agricultural and Livestock Research Organization, Kikuyu, Kenya

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; **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, 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: 2 approved]** Wellcome Open Research 2024, 9:42 <https://doi.org/10.12688/wellcomeopenres.20840.1>

First published: 19 Feb 2024, 9:42 <https://doi.org/10.12688/wellcomeopenres.20840.1>

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; *Athrips*; *Athrips mouffetella* (Linnaeus, 1758) (NCBI:txid1101086).

Background

Athrips mouffetella, the Dotted Grey Groundling, is a micro-moth in the family Gelechiidae. This small moth (forewing length 7–8.5mm) is silvery grey, with a number of black elongated spots running along the forewings. It is single brooded, flying between June and early September, and readily comes to light (Sterling *et al.*, 2012).

The moth is common but local in England and Wales and there is a lone record from Scotland (Palmer & Palmer, 2023). It is widespread throughout Europe, excluding the Iberian Peninsula and there are isolated records to the far east of Asia (GBIF Secretariat, 2023). It has been accidentally introduced to North America (Emmet & Langmaid, 2002).

The larvae feed on the terminal leaves of honeysuckle and snowberry, spinning a silken web. They can be distinguished by their colouration from other web-spinning moth larvae which share these food plants. A *mouffetella* larvae are dark purplish-black or grey with white spots behind the head, and along the side of the body. These spots are brighter on the thorax but more faded on subsequent segments (Emmet & Langmaid, 2002).

The genome of *A. mouffetella* 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 *Athrips mouffetella* based on one female specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one female *Athrips mouffetella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 33-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 51 missing joins or mis-joins and removed 11 haplotypic duplications, reducing the assembly length by 0.51% and the scaffold number by 10.56%, and decreasing the scaffold N50 by 1.32%.

The final assembly has a total length of 869.7 Mb in 126 sequence scaffolds with a scaffold N50 of 29.8 Mb (Table 1). The snailplot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC



Figure 1. Photograph of the *Athrips mouffetella* (ilAthMouf1) specimen used for genome sequencing.

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 (99.2%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the Z and W sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 63.1 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.4% (single = 97.3%, duplicated = 1.1%), 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/1101086>.

Genome annotation report

The *Athrips mouffetella* genome assembly (GCA_947532105.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Athrips_mouffetella_GCA_947532105.1/Info/Index). The resulting annotation includes 23,064 transcribed mRNAs from 22,889 protein-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A female *Athrips mouffetella* (specimen ID Ox000593, ToLID ilAthMouf1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2020-07-05. The specimen used for Hi-C sequencing (specimen ID Ox001679, ToLID ilAthMouf2) was collected from the same location on 2021-07-17. Both

Table 1. Genome data for *Athrips mouffetella*, ilAthMouf1.1.

Project accession data		
Assembly identifier	ilAthMouf1.1	
Species	<i>Athrips mouffetella</i>	
Specimen	ilAthMouf1	
NCBI taxonomy ID	1101086	
BioProject	PRJEB55448	
BioSample ID	SAMEA7701457	
Isolate information	ilAthMouf1 ilAthMouf2	
Assembly metrics*		Benchmark
Consensus quality (QV)	63.1	≥ 50
<i>k</i> -mer completeness	100.0%	≥ 95%
BUSCO**	C:98.4%[S:97.3%,D:1.1%],F:0.4%,M:1.3%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.2%	≥ 95%
Sex chromosomes	ZW	localised homologous pairs
Organelles	Mitochondrial genome: 15.23 kb	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10100490, ERR10100491	
Hi-C Illumina	ERR10113284	
Genome assembly		
Assembly accession	GCA_947532105.1	
Accession of alternate haplotype	GCA_947532115.1	
Span (Mb)	869.7	
Number of contigs	273	
Contig N50 length (Mb)	9.6	
Number of scaffolds	126	
Scaffold N50 length (Mb)	29.8	
Longest scaffold (Mb)	57.33	
Genome annotation		
Number of protein-coding genes	22,889	
Number of gene transcripts	23,064	

* 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/CANNWQ01/dataset/CANNWQ01/busco>.

specimens were collected using a light trap and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

Protocols developed by the Wellcome Sanger Institute (WSI) Tree of Life core laboratory have been deposited on protocols.io (Denton *et al.*, 2023b). The workflow for high

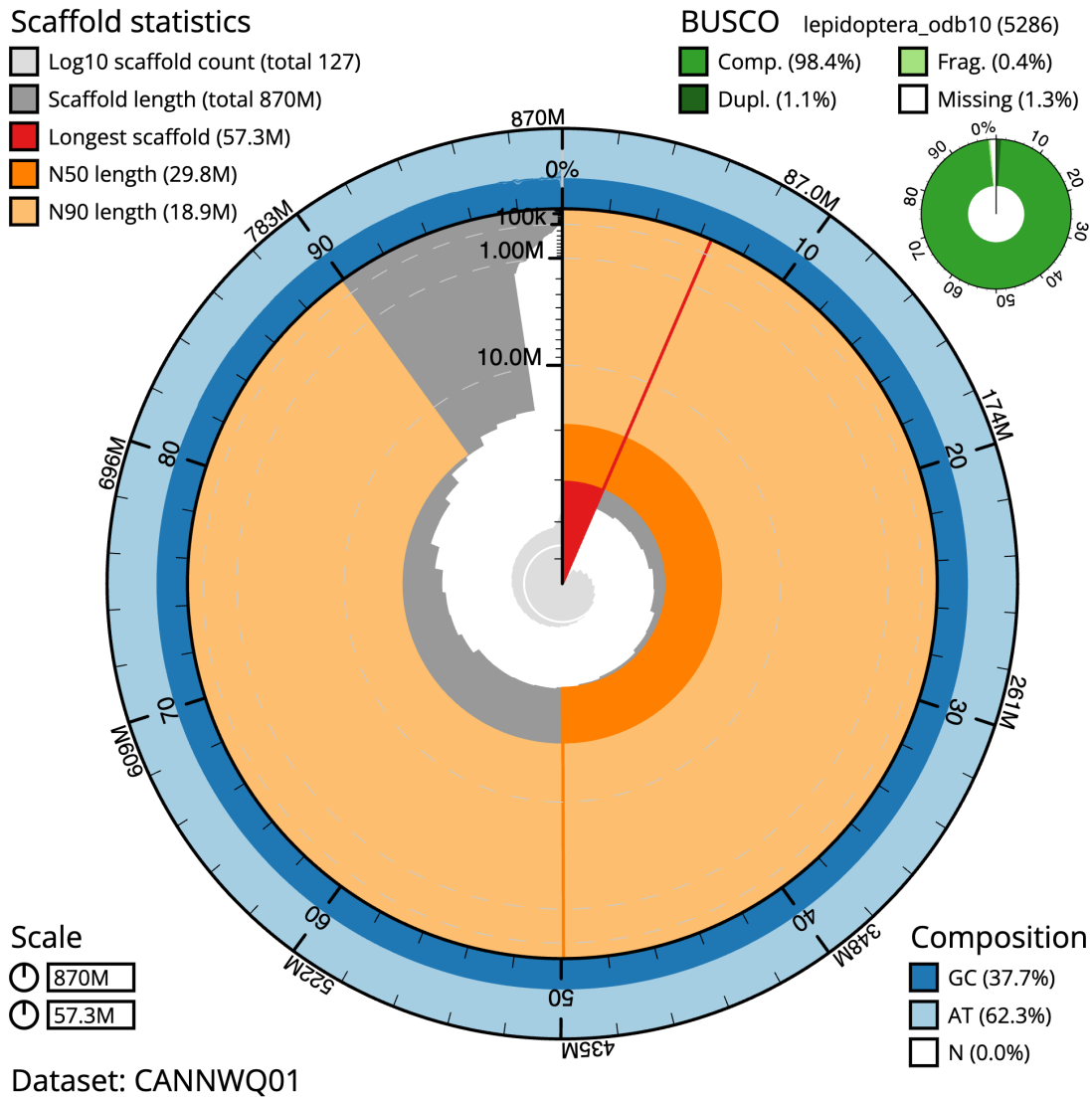


Figure 2. Genome assembly of *Athrips mouffetella*, ilAthMouf1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 869,731,197 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (57,334,238 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (29,779,897 and 18,863,855 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/CANNWQ01/dataset/CANNWQ01/snail>.

molecular weight (HMW) DNA extraction at the WSI includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilAthMouf1 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). HMW 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

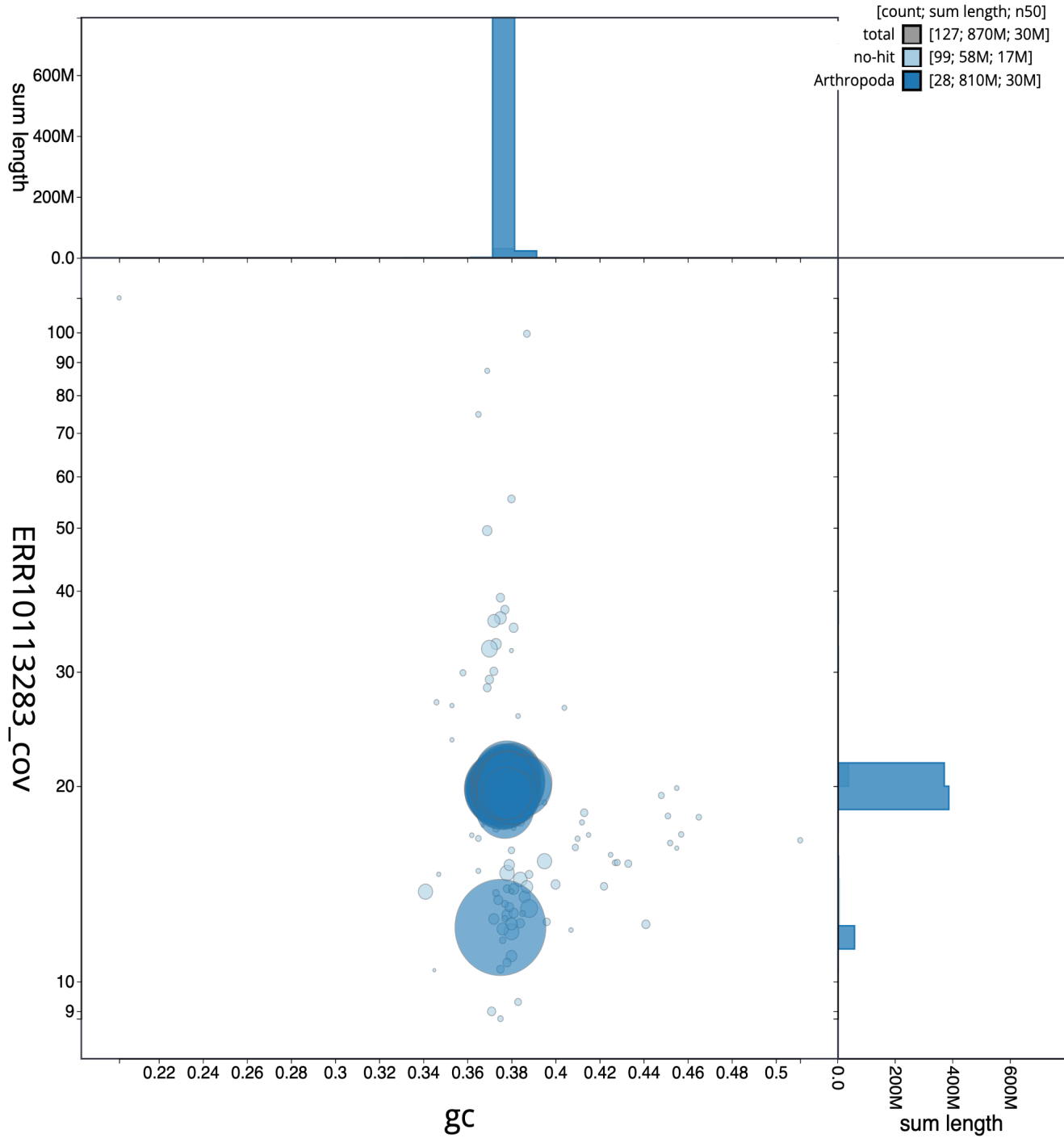


Figure 3. Genome assembly of *Athrips mouffetella*, ilAthMouf1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CANNWQ01/dataset/CANNWQ01/blob>.

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.

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

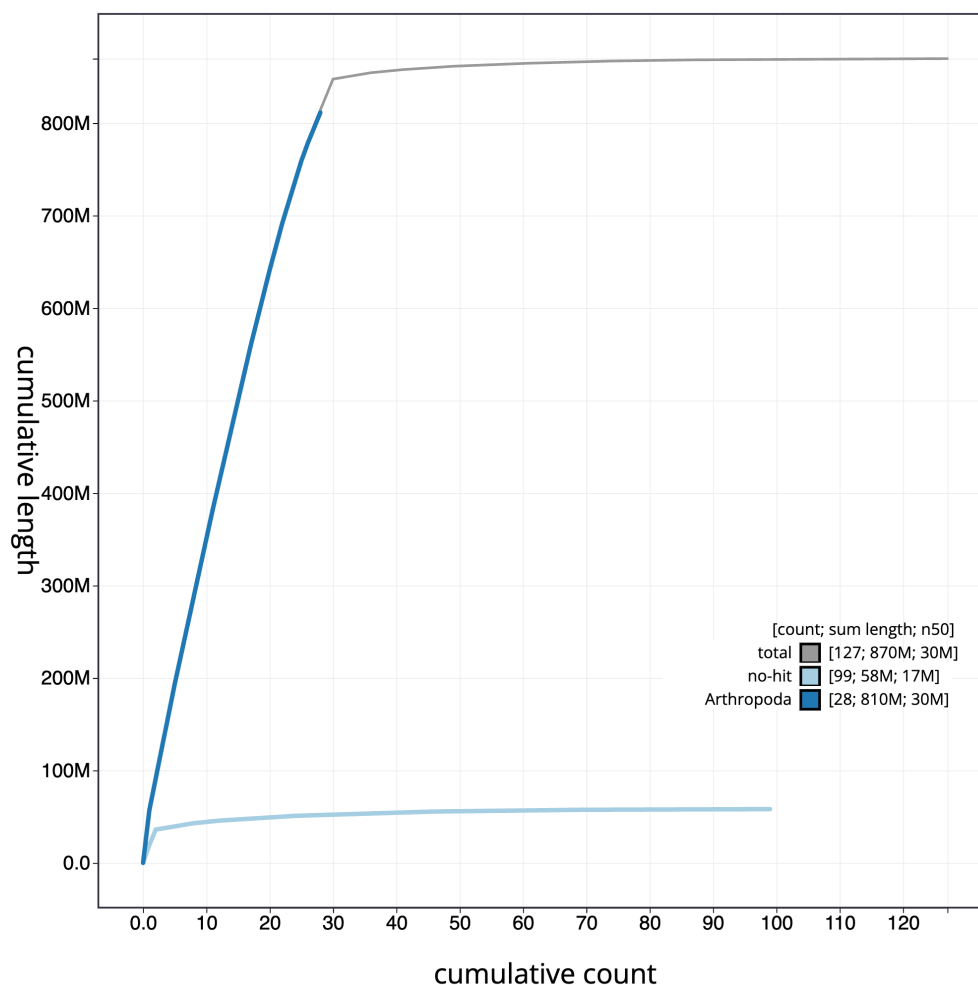


Figure 4. Genome assembly of *Athrips mouffetella*, ilAthMouf1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CANNWQ01/dataset/CANNWQ01/cumulative>.

SEQUEL II (HiFi) instrument. Hi-C data were also generated from whole organism tissue of ilAthMouf2 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 scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the

final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a) and “sanger-tol/genomenote” (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

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

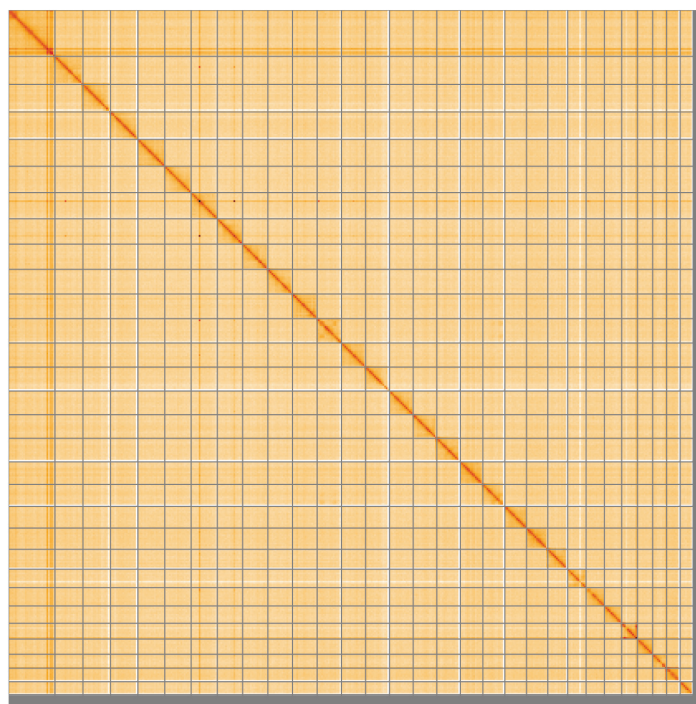


Figure 5. Genome assembly of *Athrips mouffetella*, iAthMouf1.1: Hi-C contact map of the iAthMouf1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/!/?d=Pk3PR6WiSgCVbfOA0levGg>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Athrips mouffetella*, iAthMouf1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX383957.1	1	34.7	38.0
OX383958.1	2	34.12	37.5
OX383959.1	3	33.88	38.0
OX383960.1	4	33.42	37.5
OX383961.1	5	32.59	37.5
OX383962.1	6	32.39	37.5
OX383963.1	7	31.38	37.5
OX383964.1	8	31.16	37.5
OX383965.1	9	29.53	37.5
OX383966.1	10	30.56	37.5
OX383967.1	11	30.56	38.0
OX383968.1	12	30.18	37.5
OX383969.1	13	29.78	38.0
OX383970.1	14	29.37	37.5
OX383971.1	15	29.26	37.5

INSDC accession	Chromosome	Length (Mb)	GC%
OX383972.1	16	29.23	37.5
OX383973.1	17	27.85	38.0
OX383974.1	18	27.12	38.0
OX383975.1	19	26.93	37.5
OX383976.1	20	26.26	38.0
OX383977.1	21	24.61	38.0
OX383978.1	22	22.81	37.5
OX383979.1	23	22.75	38.5
OX383980.1	24	21.46	37.5
OX383981.1	25	19.47	38.0
OX383982.1	26	18.86	38.0
OX383983.1	27	17.31	38.0
OX383984.1	28	16.62	38.5
OX383985.1	29	16.1	38.0
OX383986.1	W	1.48	39.0
OX383956.1	Z	57.33	37.5
OX383987.1	MT	0.02	20.5

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Mercury	MercuryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Athrips mouffetella* assembly (GCA_947532105.1) in Ensembl Rapid Release.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the ‘**Darwin Tree of Life Project Sampling Code of Practice**’, which can be found in full on the Darwin Tree of Life website [here](https://www.darwintreeoflife.org/). 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: *Athrips mouffetella*. Accession number PRJEB55448; <https://identifiers.org/ena.embl/PRJEB55448> (Wellcome Sanger Institute, 2022). The genome sequence is released openly for reuse. The *Athrips mouffetella* 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](#)
- Allio R, Schomaker-Bastos A, Romiguier J, et al.: **MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bernt M, Donath A, Jühling F, et al.: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Brůna T, Hoff KJ, Lomsadze A, et al.: **BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom Bioinform.* 2021; **3**(1): lqaa108.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- 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](#)
- Emmet AM, Langmaid JR: **The Moths and Butterflies of Great Britain and Ireland.** Gelechiidae. Colchester: Harley Books, 2002.
[Reference Source](#)
- GBIF Secretariat: ***Athrips mouffetella* (Linnaeus, 1758), GBIF Backbone Taxonomy.** 2023; [Accessed 31 December 2023].
[Reference Source](#)
- Guan D, McCarthy SA, Wood J, et al.: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harry E: **PretextView (Paired Read TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
[Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Manni M, Berkeley MR, Seppely M, et al.: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Palmer C, Palmer S: ***Athrips mouffetella*, Gelechiid Recording Scheme.** 2023; [Accessed 31 December 2023].
[Reference Source](#)
- Rao SSP, Huntley MH, Durand NC, et al.: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, McCarthy SA, Fedrigo O, et al.: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rhie A, Walenz BP, Koren S, et al.: **Mercury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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.** Gillingham, Dorset: British Wildlife Publishing, 2012.
[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.
[Reference Source](#)
- 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, Krashenninnikova K, et al.: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vasimuddin 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](#)
- Wellcome Sanger Institute: **The genome sequence of the Dotted Grey Groundling, *Athrips mouffetella* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB55448, 2022.
- 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 15 July 2024

<https://doi.org/10.21956/wellcomeopenres.23059.r82148>

© 2024 Mireji P. 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.



Paul Mireji 

Biotechnology Research Institute, Kenya Agricultural and Livestock Research Organization, Kikuyu, Kenya

In this manuscript, the authors sequenced and annotated the genome of Dotted Grey Groundling, from UK part of the part of the Darwin Tree of Life Project focused on sequencing all named eukaryotic species in Britain and Ireland. The authors employed standard/established procedures and protocols in this excise. The findings/conclusions are acceptable and supported by the data. It appears that the focus of the project, at this point on generating generic/automatic annotations that can be refined later thorough community annotations, probably by end users. This might account for absence of in-depth information on the annotations, such as gene families identified, which should have come out clearly. That being the case, I have the following minor concerns.

1. The authors should provide more information on the Darwin Tree of Life Project and probably what informed the choice of this species for preferential sequencing/annotations beyond the goal of the project and feeding on honeysuckle and snowberry.
2. Better description of identification/validation of the specimen e.g. taxonomic key other than who and where this was done [identified by Douglas Boyes (University of Oxford)] to facilitate reproducibility.

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

Partly

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: Medical Entomology and genomics


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 15 July 2024

<https://doi.org/10.21956/wellcomeopenres.23059.r87328>

© 2024 Clark M. 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.



Melody Clark 

Natural Environment Research Council, Cambridge, UK

This article describes the genome sequencing of the Dotted Grey Groundling, a moth common to the southern two thirds of England. The sequenced specimen was sourced from Wytham Woods. This is one of the most researched areas of woodland in the world and includes a Wytham Genome Project which aims to sequence all species in this wood. The Wytham Genome Project is a key site for the Darwin Tree of Life Project. The protocols underpinning the genome sequencing and assembly are fully described and referenced. The bioinformatic tools used to assemble, curate and annotate this genome are all freely available. The genome assembly is of very high quality with most data scaffolded to 31 chromosomal pseudomolecules, including the sex chromosomes. These data are freely available for reuse.

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, transcriptomics, organisms responses to change

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.