




DATA NOTE

The genome sequence of the clay, *Mythimna ferrago* (Fabricius, 1787) [version 1; peer review: 2 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, Melanie Gibbs¹,
Darwin Tree of Life Consortium

¹UK Centre for Ecology and Hydrology, Wallingford, Oxford, UK

+ Deceased author

V1 First published: 04 Jul 2022, 7:177
<https://doi.org/10.12688/wellcomeopenres.17923.1>
Latest published: 04 Jul 2022, 7:177
<https://doi.org/10.12688/wellcomeopenres.17923.1>

Abstract

We present a genome assembly from an individual female *Mythimna ferrago* (the clay; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 861 megabases in span. The majority of the assembly (99.98%) is scaffolded into 32 chromosomal pseudomolecules, with the W and Z chromosomes assembled. The complete mitochondrial genome was also assembled and is 15.3 kilobases in length. Gene annotation of this assembly on Ensembl has identified 14,075 protein coding genes.

Keywords



Mythimna ferrago, the clay, genome sequence, chromosomal, Arthropoda




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

Open Peer Review

Approval Status

	1	2
version 1 04 Jul 2022	 view	 view

1. **Steven M Van Belleghem** , University of Puerto Rico, San Juan, Puerto Rico
KU Leuven, Leuven, Belgium
2. **Niclas Backström** , Uppsala University, Uppsala, Sweden

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; **Gibbs 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, [href=https://doi.org/10.35802/218328](https://doi.org/10.35802/218328)><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: © 2022 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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the clay, *Mythimna ferrago* (Fabricius, 1787) [version 1; peer review: 2 approved]** Wellcome Open Research 2022, 7:177 <https://doi.org/10.12688/wellcomeopenres.17923.1>

First published: 04 Jul 2022, 7:177 <https://doi.org/10.12688/wellcomeopenres.17923.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Hadeninae; Mythimna; *Mythimna ferrago* (Fabricius, 1787) (NCBI:txid997540).

Background

The Clay, *Mythimna ferrago* (Fabricius, 1787) is a common, nocturnal, non-pest, macro-moth species that occurs across the Palearctic. In Great Britain, *M. ferrago* has been assessed against the International Union for Conservation of Nature (IUCN) Red List criteria, and categorised as a resident species of Least Concern (Fox *et al.*, 2021). The larvae feed on grasses (Robinson *et al.*, 2010), and overwinter as small larvae. The adult flight period is July and August. *Mythimna ferrago* can be found in a range of open habitats, including woodland, grassland, scrub, heathland, gardens and farmland.

Moths are important indicators of land-use and climate change (Wagner *et al.*, 2021), used to study site-specific stressors such as light pollution (Boyes *et al.*, 2021), pesticide use, and the effectiveness of agri-environment management schemes (Botham *et al.*, 2015; Merckx *et al.*, 2009; Staley *et al.*, 2016). Within the genus *Mythimna*, *M. ferrago* is unusual in that, on the basis of DNA barcoding data, it forms two distinct clusters across its entire range that do not correspond with a geographical pattern (Huemer *et al.*, 2019). Analyses at the genomic level could contribute to elucidating what drives their intra-specific variation and key genes under selection.

Genome sequence report

The genome was sequenced from a single female *M. ferrago*, collected from Wytham Woods, Berkshire, UK (Figure 1). A total of 31-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 51-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded



Figure 1. Image of the *Mythimna ferrago* specimen taken prior to preservation and processing.

with chromosome conformation Hi-C data. Manual assembly curation corrected 15 missing/misjoins and removed 1 haplotypic duplications, reducing the assembly size by 0.99% and the scaffold number by 18.52%, and increasing the scaffold N50 by 1.95%.

The final assembly has a total length of 861 Mb in 44 sequence scaffolds with a scaffold N50 of 27.9 Mb (Table 1). The majority, 99.98%, of the assembly sequence 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).

The assembly has a BUSCO v5.1.2 (Manni *et al.*, 2021) completeness of 98.9% (single 98.1%, duplicated 0.8%) using the lepidoptera_odb10 reference set (n=954). 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 ilMytFerr1.2 genome has been annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Mythimna_ferrago_GCA_910589285.1/Info/Index). The resulting annotation includes 25,713 transcribed mRNAs from 14,075 protein-coding and 3,455 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A single adult female *M. ferrago* specimen (ilMytFerr1) was collected using a light trap from Wytham Woods, Berkshire, UK (latitude 51.772, longitude -1.338) by Douglas Boyes (University of Oxford). The specimen was identified by Douglas Boyes and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The ilMytFerr1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Abdomen tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. 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 between 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 sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Table 1. Genome data for *Mythimna ferrago*, ilMytFerr1.2.

Project accession data	
Assembly identifier	ilMytFerr1.2
Species	<i>Mythimna ferrago</i>
Specimen	ilMytFerr1 (genome assembly; Hi-C)
NCBI taxonomy ID	1789172
BioProject	PRJEB45178
BioSample ID	SAMEA7701536
Isolate information	Adult female. Abdomen (genome sequencing); head/thorax (Hi-C)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6436381
10X Genomics Illumina	ERR6054871-ERR6054874
Hi-C Illumina	ERR6054870
Genome assembly	
Assembly accession	GCA_910589285.1
Accession of alternate haplotype	GCA_910589535.1
Span (Mb)	861
Number of contigs	60
Contig N50 length (Mb)	27.2
Number of scaffolds	44
Scaffold N50 length (Mb)	27.9
Longest scaffold (Mb)	41.4
BUSCO* genome score	C:98.9%[S:98.1%,D:0.8%],F:0.3%,M:0.8%,n:5286
Genome annotation	
Number of protein-coding genes	14,075
Average length of coding sequence (bp)	20314.26
Average number of exons per transcript	6.64

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.1.2. 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/ilMytFerr1.2/dataset/CAJU002/busco#Filters>.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were generated in the Tree of Life laboratory from head/thorax tissue of

ilMytFerr1 using the Arima v2 kit and sequenced on a NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021); 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

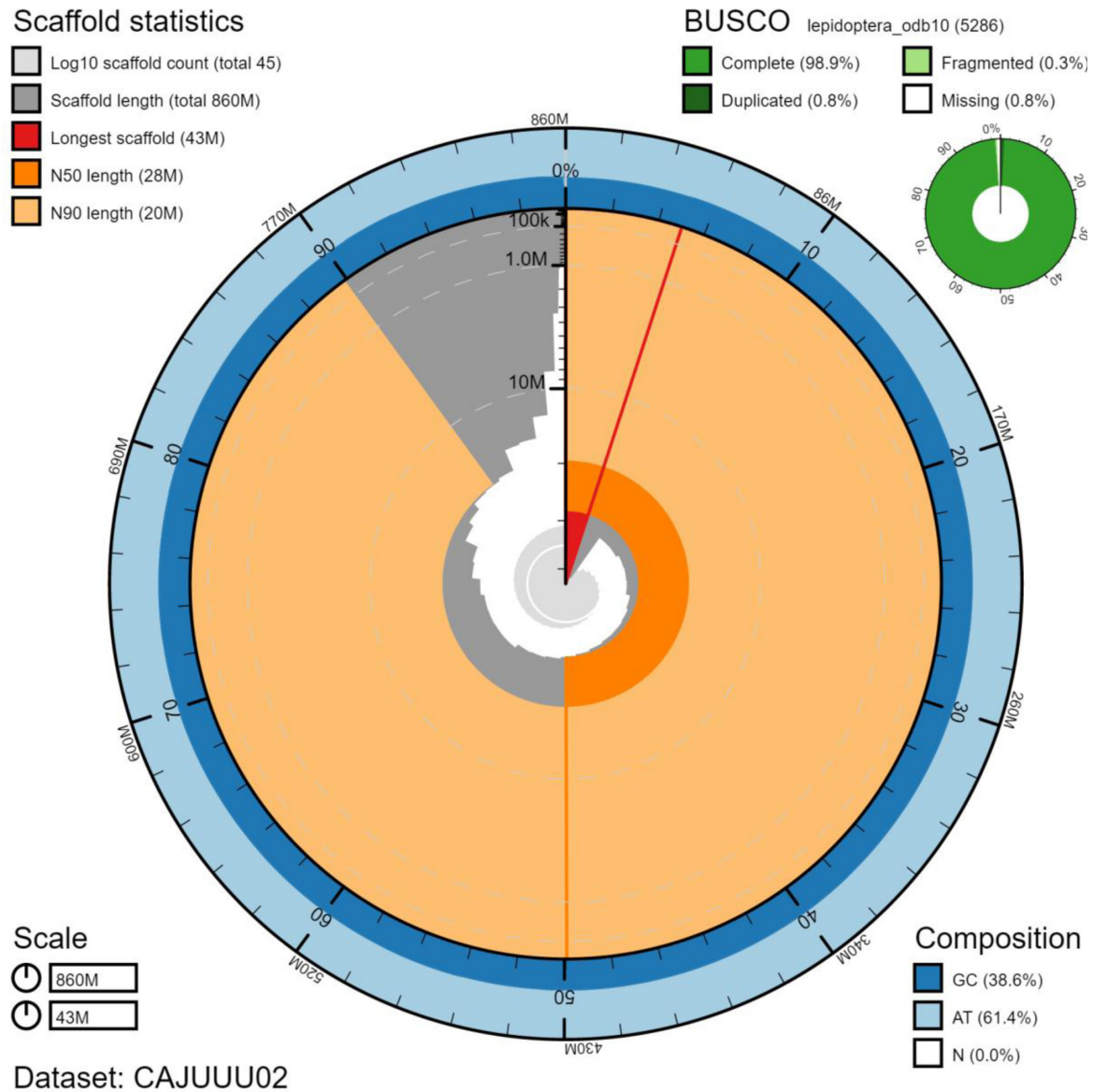


Figure 2. Genome assembly of *Mythemna ferrago*, ilMytFerr1.2: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 860,988,366 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (43,351,033 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (27,874,102 and 19,573,539 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/ilMytFerr1.2/dataset/CAJU02/snail#Filters>.

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 and corrected using the gEVAL system (Chow et al., 2016) as described previously

(Howe et al., 2021). Manual curation (Howe et al., 2021) was performed using gEVAL, HiGlass (Kerpedjiev et al., 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021), which performs annotation using MitoFinder (Allio et al., 2020). The genome was

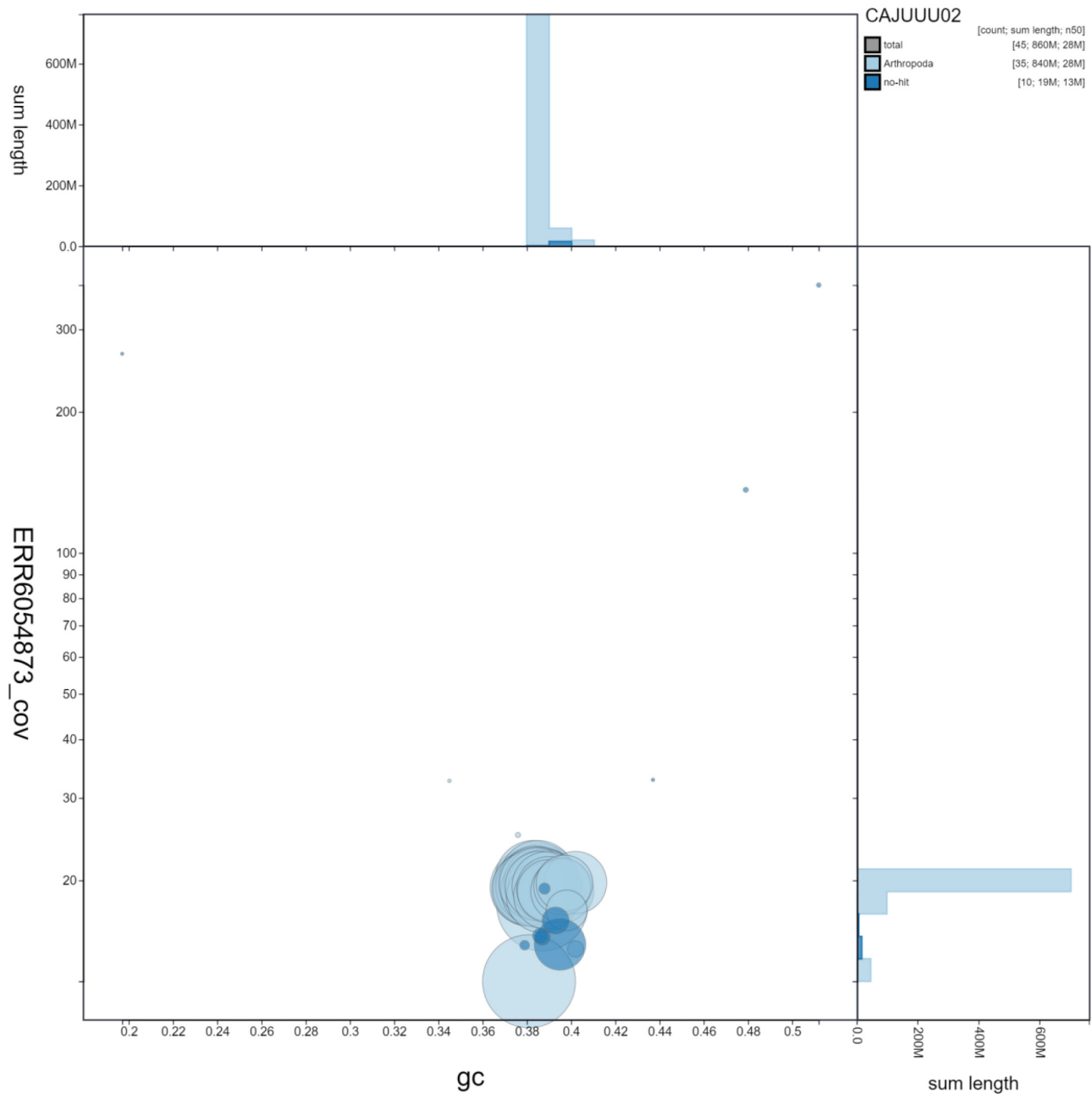


Figure 3. Genome assembly of *Mythimna ferrago*, ilMytFerr1.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/Mythimna%20ferrago/dataset/CAJU02/blob#Filters>.

analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Genome annotation

The Ensembl gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Mythimna Ferrago* assembly (GCA_910589285.1). Annotation was created primarily through

alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

Ethics/compliance issues

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

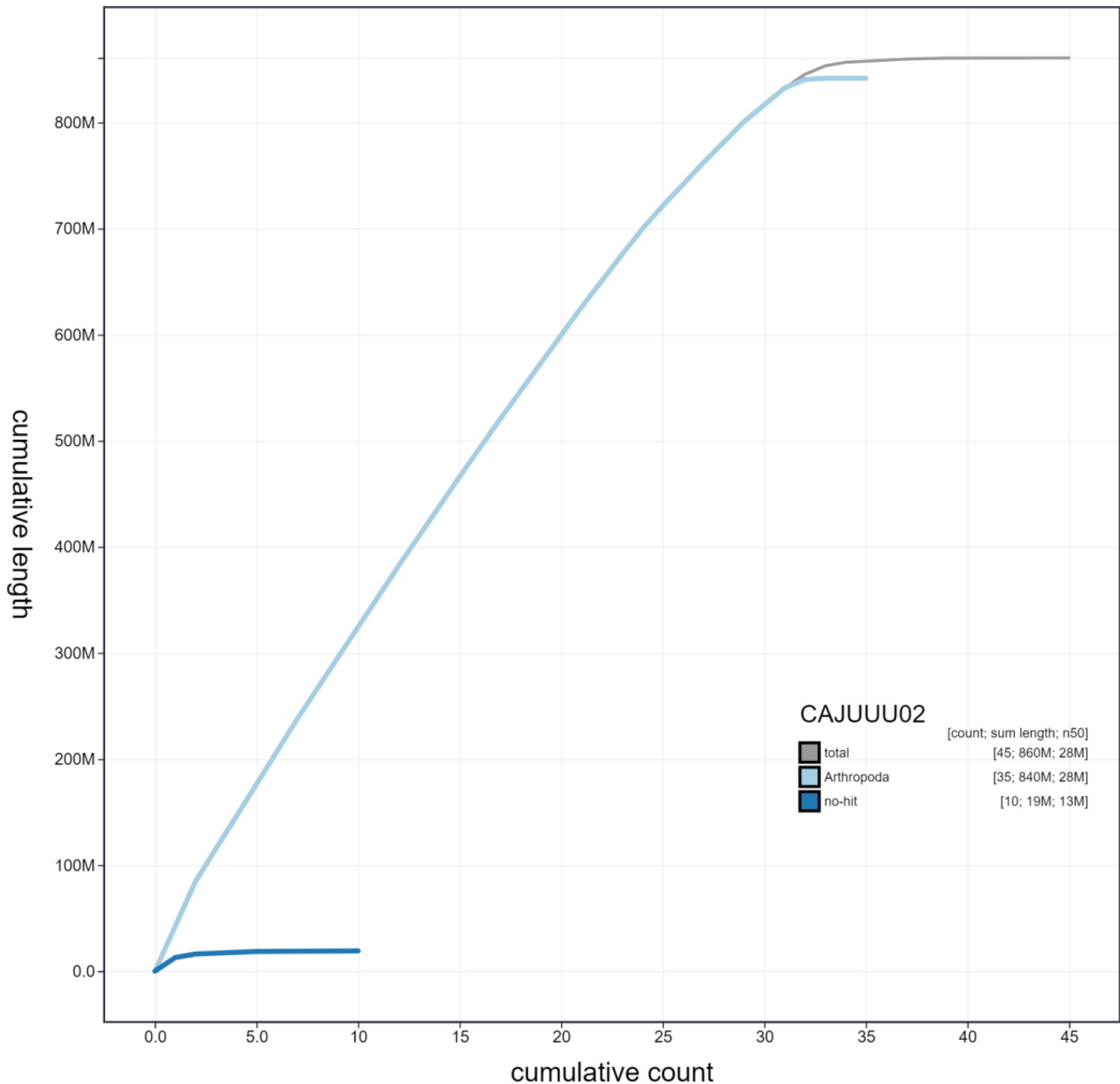


Figure 4. Genome assembly of *Mythimna ferrago*, ilMytFerr1.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/Mythimna%20ferrago/dataset/CAJU002/cumulative#Filters>.

[Darwin Tree of Life Project Sampling Code of Practice](#). By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. 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.

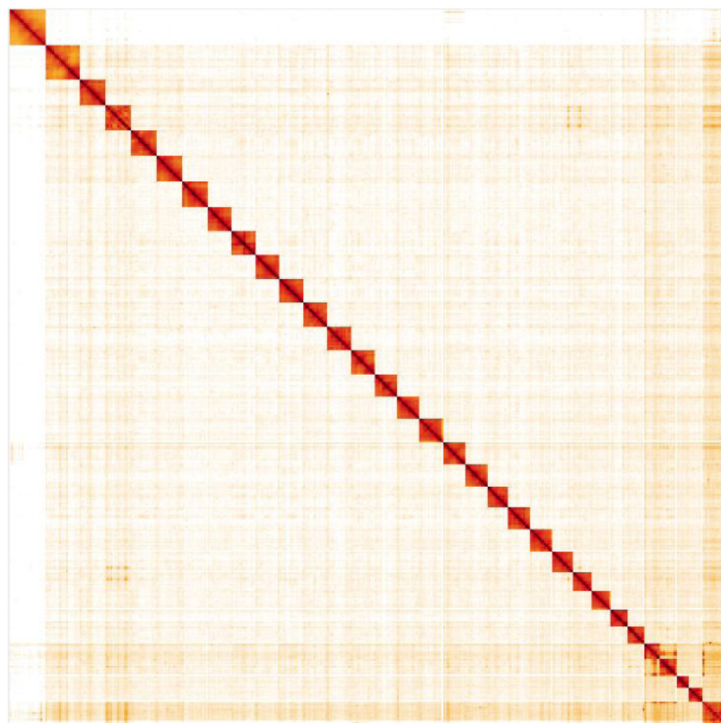


Figure 5. Genome assembly of *Mythimna ferrago*, ilMytFerr1.2: Hi-C contact map. Hi-C contact map of the ilMytFerr1.2 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=UZmmjbDIR2C13gmXjFQ-xw>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Mythimna ferrago*, ilMytFerr1.2.

INSDC accession	Chromosome	Size (Mb)	GC%
OU342674.1	1	41.45	38.7
OU342675.1	2	31.00	38.4
OU342676.1	3	30.97	38.7
OU342677.1	4	30.43	38.6
OU342678.1	5	30.31	38.5
OU342679.1	6	30.20	38.1
OU342680.1	7	29.39	38.2
OU342681.1	8	29.06	38.3
OU342682.1	9	28.89	38.2
OU342683.1	10	28.81	38.5
OU342684.1	11	28.56	38.3
OU342685.1	12	28.26	38.3
OU342686.1	13	27.87	38.4
OU342687.1	14	27.58	38.1
OU342688.1	15	27.34	38.7
OU342689.1	16	27.19	38.5

INSDC accession	Chromosome	Size (Mb)	GC%
OU342690.1	17	26.42	38.8
OU342691.1	18	26.40	38.6
OU342692.1	19	26.34	38.8
OU342693.1	20	26.30	38.4
OU342694.1	21	25.67	38.7
OU342695.1	22	24.25	38.6
OU342696.1	23	23.95	38.9
OU342697.1	24	21.68	38.9
OU342698.1	25	20.82	39.0
OU342699.1	26	19.74	39.0
OU342700.1	27	19.57	39.6
OU342701.1	28	19.18	40.2
OU342702.1	29	15.96	39.6
OU342703.1	30	15.72	39.7
OU342704.1	W	12.89	39.5
OU342673.1	Z	43.35	38.1
OU342705.1	MT	0.02	19.9
-	Unplaced	15.45	39.5

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.3	Cheng et al., 2021
purge_dups	1.2.3	Guan et al., 2020
SALSA2	2.2	Ghurye et al., 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.0	Uliano-Silva et al., 2021
HiGlass	1.11.6	Kerpedjiev et al., 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.0.5	Challis et al., 2020

Data availability

European Nucleotide Archive: *Mythimna ferrago* (the clay). Accession number [PRJEB45178](#); <https://identifiers.org/ena.embl/PRJEB45178>.

The genome sequence is released openly for reuse. The *M. ferrago* 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.6418202>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.6418156>.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: <https://doi.org/10.5281/zenodo.6418327>.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: <https://doi.org/10.5281/zenodo.5746904>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.6125046>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.6418363>.

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](#)
- Botham MS, Fernandez-Ploquin EC, Brereton T, et al.: **Lepidoptera Communities across an Agricultural Gradient: How Important Are Habitat Area and Habitat Diversity in Supporting High Diversity? J Insect Conserv.** 2015; **19**(2): 403–20.
[PubMed Abstract](#) | [Free Full Text](#)
- Boyes DH, Evans DM, Fox R, et al.: **Street Lighting Has Detrimental Impacts on Local Insect Populations. Sci Adv.** 2021; **7**(35): eabi8322.
[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): 1361–74.
[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): 170–75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chow W, Brugger K, Caccamo M, et al.: **gEVAL - a web-based browser for evaluating genome assemblies. Bioinformatics.** 2016; **32**(16): 2508–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fox R, Dennis EB, Harrower CA, et al.: **The State of Britain's Larger Moths 2021.** March, 44. 2021.
[Reference Source](#)
- Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** arXiv: 1207.3907. 2012.
[Reference Source](#)
- 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): 2896–98.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *Gigascience.* 2021; **10**(1): gaaa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Huemer P, Wieser C, Stark W, *et al.*: **DNA Barcode Library of Megadiverse Austrian Noctuoidea (Lepidoptera) - a Nearly Perfect Match of Linnean Taxonomy.** *Biodivers Data J.* 2019; **7**: e37734.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps.** *Genome Biol.* 2018; **19**(1): 125.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–54.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Merckx T, Feber RE, Dullieu RL, *et al.*: **Effect of Field Margins on Moths Depends on Species Mobility: Field-Based Evidence for Landscape-Scale Conservation.** *Agric Ecosyst Environ.* 2009; **129**(1–3): 302–9.

[Publisher 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): 1665–80.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Robinson GS, Ackery PR, Kitching IJ, *et al.*: **HOSTS-a Database of the World's Lepidopteran Hostplants.** Natural History Museum. London. 2010.

[Reference Source](#)

Staley JT, Botham MS, Chapman RE, *et al.*: **Little and Late: How Reduced Hedgerow Cutting Can Benefit Lepidoptera.** *Agric Ecosyst Environ.* 2016; **224**: 22–28.

[Publisher Full Text](#)

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0.** 2021.

[Publisher Full Text](#)

UniProt Consortium: **UniProt: A Worldwide Hub of Protein Knowledge.** *Nucleic Acids Research.* 2019; **47**(D1): D506–15.

[Publisher Full Text](#)

Wagner DL, Fox R, Salcido DM, *et al.*: **A Window to the World of Global Insect Declines: Moth Biodiversity Trends Are Complex and Heterogeneous.** *Proceedings of the National Academy of Sciences of the United States of America.* 2021; **118**(2): e2002549117.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 16 August 2022

<https://doi.org/10.21956/wellcomeopenres.19861.r51418>

© 2022 Backström N. 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.

 **Niclas Backström** 

Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden

The article describes the statistics of a genome assembly of the clay (*Mythimna ferrago*) and is written in standard format for the DToL genome notes. The paper is well written and the results are clearly presented. Accession information for all data sets is also available in the note. The quality of the assembly is impressive and the data will definitely be useful for forthcoming intra-specific population genomic studies and for comparative genomics studies in Lepidoptera in general.

Minor comments:

- 1) It's a bit unclear to me what average length of coding sequence refers to (Table 1). Is this the average length of the entire gene (introns and UTR included) or the sum of lengths of all coding exons (probably not since it is > 20 kb), or something else? Maybe a short note in the table legend can be added to make this clear.
- 2) Methods: Section 1, first sentence. "using a light trap from Wytham Woods" sounds a bit like the light trap was from WW and not that the sample site was WW?
- 3) Methods: Genome annotation paragraph. Unclear if the transcriptomic data are from the same species?
- 4) Methods: Genome annotation paragraph. Maybe specify which set of proteins from UniProt that was used for the alignments?

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: Evolutionary genomics, population genetics, molecular evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 08 July 2022

<https://doi.org/10.21956/wellcomeopenres.19861.r51417>

© 2022 Van Belleghem S. 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.



Steven M Van Belleghem 

¹ Department of Biology, Rio Piedras Campus, University of Puerto Rico, San Juan, Puerto Rico

² Department of Biology, KU Leuven, Leuven, Belgium

The note reports on the the assembly of the genome of the clay butterfly through the initiative of the Darwin Tree of Life Barcoding collective. The assembly included PacBio Hifi, 10X and Hi-C data and resulted in a high quality assembly that was further manually curated to remove minor assembly mistakes.

Comments:

Please clarify how the W and Z chromosome were identified.

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, adaptation

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.
