



DATA NOTE

The genome sequence of the broad-bordered yellow underwing, *Noctua fimbriata* (Schreber, 1759) [version 1; peer review: 1 approved, 1 approved with reservations]

Peter W.H. Holland 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, Darwin Tree of Life Consortium

¹Department of Zoology, University of Oxford, Oxford, UK

V1 First published: 15 Dec 2021, 6:345
<https://doi.org/10.12688/wellcomeopenres.17490.1>
Latest published: 15 Dec 2021, 6:345
<https://doi.org/10.12688/wellcomeopenres.17490.1>

Abstract

We present a genome assembly from an individual female *Noctua fimbriata* (the broad-bordered yellow underwing; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 574 megabases in span. The complete assembly is scaffolded into 32 chromosomal pseudomolecules, with the W and Z sex chromosomes assembled.

Keywords

Noctua fimbriata, broad-bordered yellow underwing, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1		
15 Dec 2021	view	view

- Joost van den Heuvel**, Wageningen University & Research, Wageningen, The Netherlands
- Craig J. Anderson** , University of Edinburgh, Edinburgh, UK

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: Holland PWH: Investigation, Resources, Supervision, 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) and the Darwin Tree of Life Discretionary Award (218328).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Holland PWH *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: Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the broad-bordered yellow underwing, *Noctua fimbriata* (Schreber, 1759) [version 1; peer review: 1 approved, 1 approved with reservations]** Wellcome Open Research 2021, 6:345
<https://doi.org/10.12688/wellcomeopenres.17490.1>

First published: 15 Dec 2021, 6:345 <https://doi.org/10.12688/wellcomeopenres.17490.1>

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Noctuinae; Noctuini; Noctua; *Noctua fimbriata* (Schreber, 1759) (NCBI:txid753202).

Background

Noctua fimbriata (broad-bordered yellow underwing) is a common noctuid moth with marked sexual dimorphism in its wing colouration: females are orange/buff coloured, whereas males are darker brown. There is also variation between individuals in wing colour and pattern, but only on the upper surface of the forewing; (Owen & Whiteley, 1989) have pointed out that restriction to the visual surface is consistent with polymorphism maintained by frequency-dependent selection by predators. The species is found across Europe and western parts of Asia; it occurs throughout the UK, but is less common in the north of England and in Scotland. Larvae of *Noctua fimbriata* are polyphagous, feeding on many species of herbaceous plant as well as low-growing trees and shrubs; the species is common in woodlands and also often recorded in gardens. *N. fimbriata* has an unusual flight period in the UK with adults emerging in July, then undergoing a summer aestivation

period before a second flight period in late August and September (Randle *et al.*, 2019); the adaptive significance of the aestivation period is unclear.

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

Genome sequence report

The genome was sequenced from a single female *N. fimbriata* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338). A total of 20-fold coverage in Pacific Biosciences single-molecule long reads (N50 16 kb) and 73-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 222 missing/misjoins and removed 9 haplotypic duplications, reducing the assembly length by 0.37% and the scaffold number by 69.09%, and increasing the scaffold N50 by 41.75%.



Figure 1. Image of the ilNocFimb1 specimen taken during preservation and processing.

The final assembly has a total length of 574 Mb in 51 sequence scaffolds with a scaffold N50 of 19.0 Mb (Table 1). Of the assembly sequence, 100% was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes (numbered by sequence length), and the W and Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO (Manni *et al.*, 2021) completeness of 99.0% using thelepidoptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition, DNA extraction and sequencing

A single female *N. fimbriata* (ilNocFimb1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.338) by Peter Holland, University of Oxford, and identified

by the same individual. The specimen was found alive in a rain puddle in daytime and preserved on dry ice prior to transfer to the Wellcome Sanger Institute.

DNA was extracted from thorax/abdomen tissue at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. Pacific Biosciences HiFi circular consensus and 10X Genomics 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 and Illumina HiSeq X instruments. Hi-C data were generated from remaining thorax/abdomen tissue using the Arima Hi-C+ kit in the Sanger Tree of Life core laboratory and sequenced on HiSeq X.

Table 1. Genome data for *Noctua fimbriata*, ilNocFimb1.1.

Project accession data	
Assembly identifier	ilNocFimb1.1
Species	<i>Noctua fimbriata</i>
Specimen	ilNocFimb1
NCBI taxonomy ID	NCBI:txid753202
BioProject	PRJEB42136
BioSample ID	SAMEA7519914
Isolate information	Female, head/abdomen/thorax
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6590582
10X Genomics Illumina	ERR6002696-ERR6002699
Hi-C Illumina	ERR6002700-ERR6002702
Genome assembly	
Assembly accession	GCA_905163415.1
Accession of alternate haplotype	GCA_905163425.1
Span (Mb)	574
Number of contigs	267
Contig N50 length (Mb)	4.4
Number of scaffolds	52
Scaffold N50 length (Mb)	19.0
Longest scaffold (Mb)	21.7
BUSCO* genome score	C:98.2%[S:97.5%,D:0.7%],F:0.8%,M:1.0%,n:1658

*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/ilNocFimb1.1/dataset/CAJHZP01/busco>.

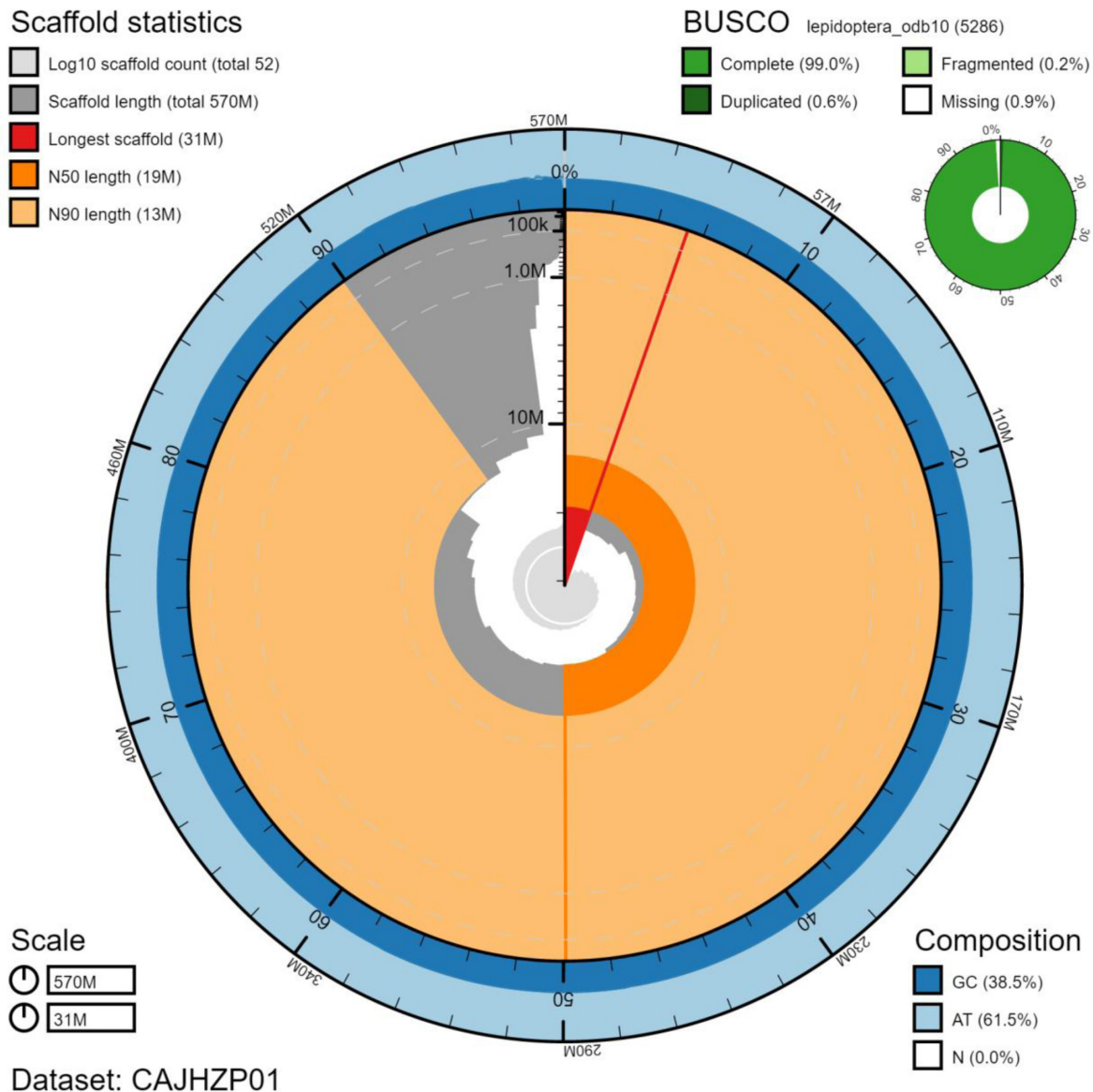


Figure 2. Genome assembly of *Noctua fimbriata*, ilNocFimb1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 573,955,380 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (30,735,469 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (18,977,511 and 13,094,214 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/ilNocFimb1.1/dataset/CAJHZP01/snail>.

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 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)

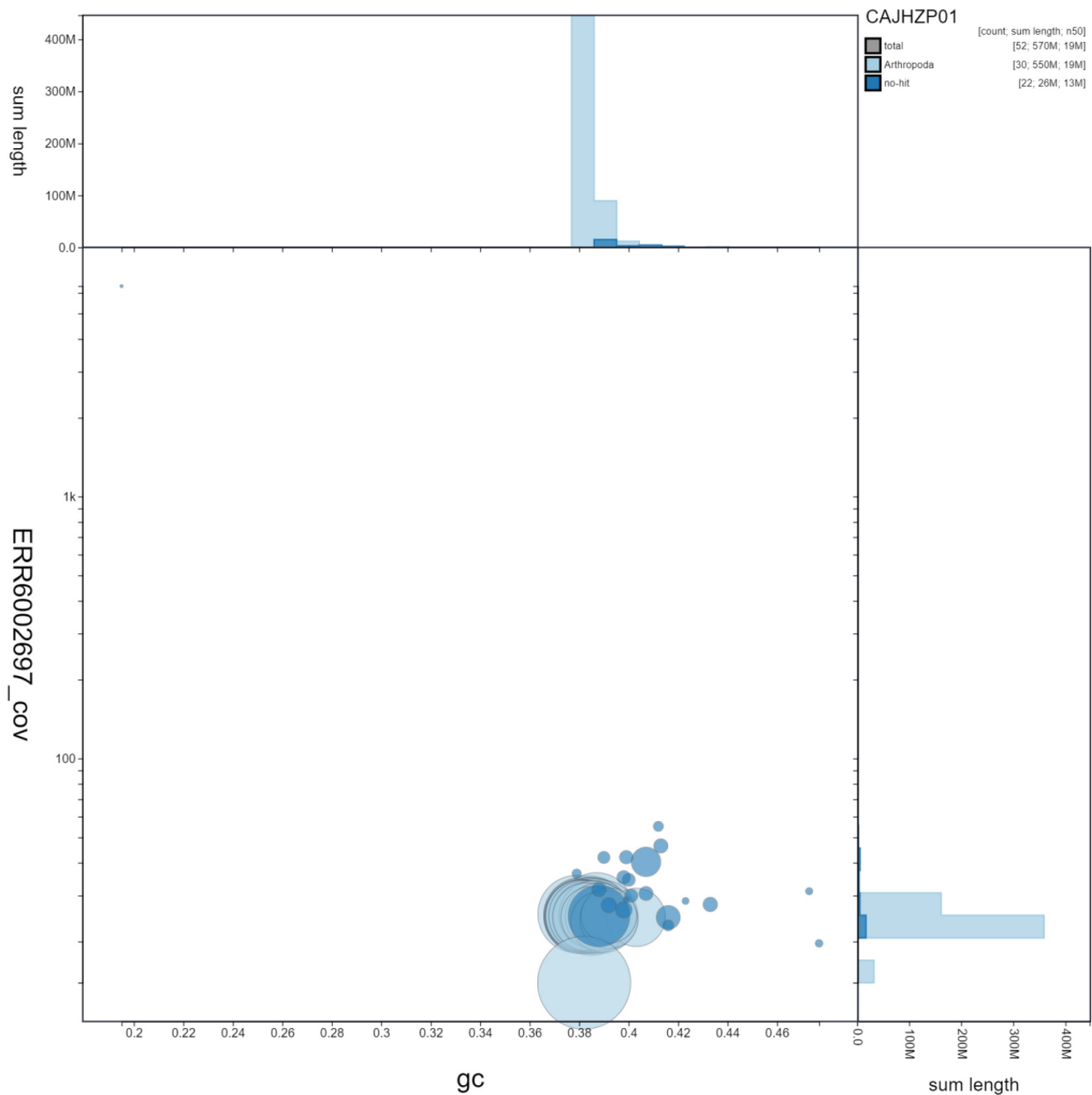


Figure 3. Genome assembly of *Noctua fimbriata*, ilNocFimb1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilNocFimb1.1/dataset/CAJHZIP01/blob>.

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 performed annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores

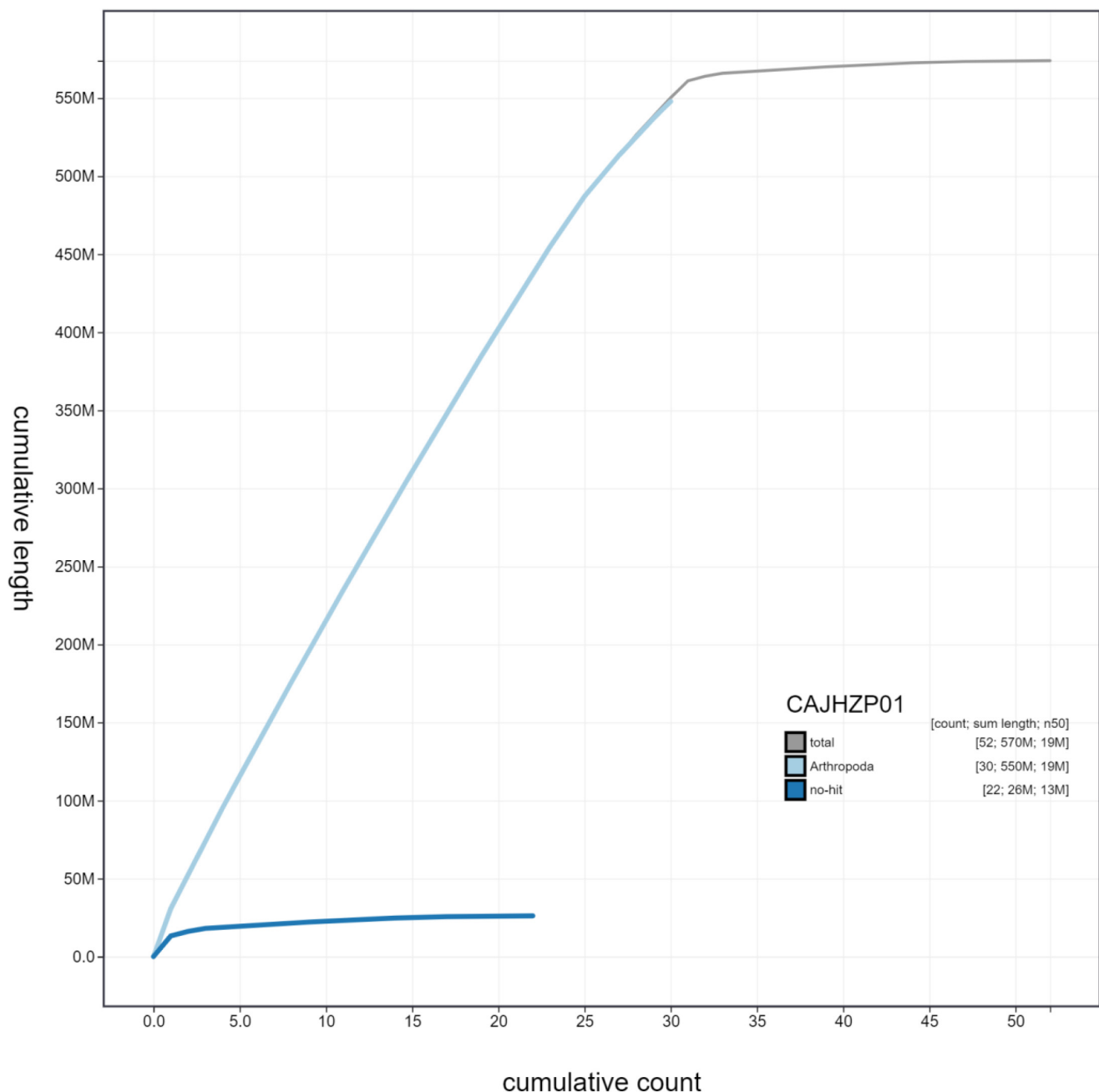


Figure 4. Genome assembly of *Noctua fimbriata*, ilNocFimb1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilNocFimb1.1/dataset/CAJHZP01/cumulative>.

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

Ethics/compliance issues

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

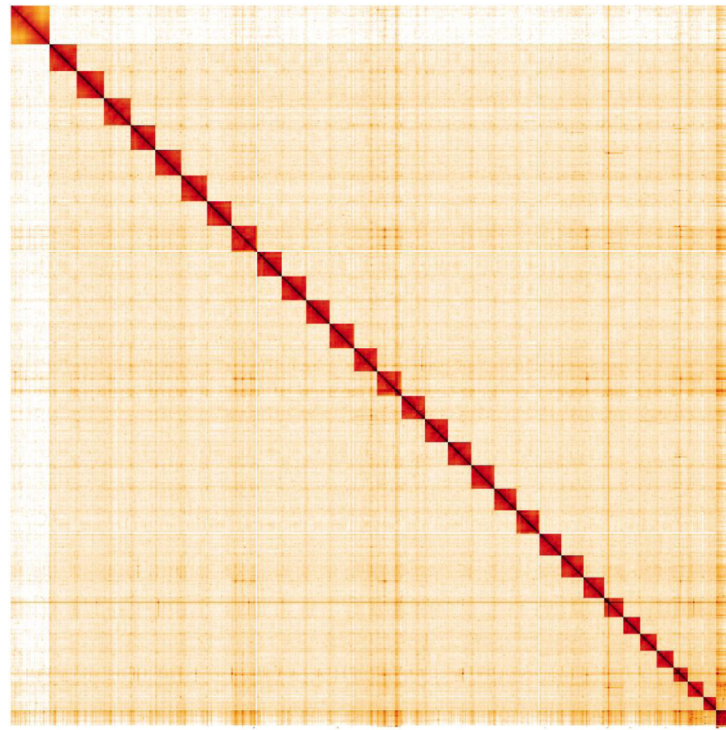


Figure 5. Genome assembly of *Noctua fimbriata*, ilNocFimb1.1: Hi-C contact map. Hi-C contact map of the ilNocFimb1.1 assembly, visualised in HiGlass. Chromosomes are shown in order of size from left to right and top to bottom.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Noctua fimbriata*, ilNocFimb1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
LR990922.1	1	21.73	38.5
LR990923.1	2	21.62	38.4
LR990924.1	3	21.23	37.9
LR990925.1	4	20.31	38.1
LR990926.1	5	20.23	38.5
LR990927.1	6	20.11	38.1
LR990928.1	7	19.97	38.6
LR990935.1	8	18.94	38.7
LR990929.1	9	19.70	38.5
LR990930.1	10	19.66	38.4
LR990931.1	11	19.52	38.1
LR990932.1	12	19.05	38.1
LR990933.1	13	18.98	38.3
LR990934.1	14	18.98	38.9
LR990936.1	15	18.69	38.6
LR990937.1	16	18.49	38.1

INSDC accession	Chromosome	Size (Mb)	GC%
LR990938.1	17	18.34	38.1
LR990939.1	18	18.31	38.6
LR990940.1	19	17.69	38.4
LR990941.1	20	17.69	38.4
LR990942.1	21	17.67	38.2
LR990943.1	22	17.25	38.6
LR990944.1	23	16.26	38.3
LR990945.1	24	15.93	38.8
LR990946.1	25	13.23	38.8
LR990947.1	26	13.18	38.5
LR990948.1	27	13.09	38.9
LR990949.1	28	11.91	40.3
LR990950.1	29	11.80	39.2
LR990951.1	30	10.79	39.2
LR990952.1	W	2.94	40.7
LR990921.1	Z	30.74	38.2
LR990953.1	MT	0.02	19.3
-	Unplaced	9.94	40.7

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.12	Cheng <i>et al.</i> , 2021
purge_dups	1.2.3	Guan <i>et al.</i> , 2020
SALSA2	2.2	Ghurye <i>et al.</i> , 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	1	https://github.com/marcelauliano/MitoHiFi
gEVAL	N/A	Chow <i>et al.</i> , 2016
PretextView	0.1.x	https://github.com/wtsi-hpag/PretextView
HiGlass	1.11.6	Kerpedjiev <i>et al.</i> , 2018
BlobToolKit	2.6.2	Challis <i>et al.</i> , 2020

of materials by a Darwin Tree of Life Partner is subject to the [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.

Data availability

European Nucleotide Archive: *Noctua fimbriata* (broad-bordered yellow underwing). Accession number [PRJEB42136](#): <https://www.ebi.ac.uk/ena/browser/view/PRJEB42136>.

The genome sequence is released openly for reuse. The *N. fimbriata* 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 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.5746938>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.5744972>.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: <https://doi.org/10.5281/zenodo.5744840>.

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.5743293>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.5638618>.

References

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](#)

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](#)

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): 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): 4647–54.

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

Owen DF, Whiteley D: **Evidence That Reflexive Polymorphisms Are Maintained by Visual Selection by Predators.** *Oikos*. 1989; **55**(1): 130–33.

[Publisher Full Text](#)

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain and Ireland's Larger Moths.** Pisces Publications, Newbury. 2019.

[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): 1665–80.

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

Uliano-Silva M, Ferreira Nunes JG, Krashenninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0.** 2021.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 12 August 2022

<https://doi.org/10.21956/wellcomeopenres.19340.r51888>

© 2022 Anderson C. 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.



Craig J. Anderson 

MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Holland *et al.* report the genome assembly of the Noctuid moth, *Noctua fimbriata*, as part of the Darwin Tree of Life project. The authors provide key insights into the likely generalisability of this reference with respect to the wider distribution of the species, before going on to precisely describe the origin of the specimen and assembly quality, which is of very high quality, containing no ambiguous nucleotides.

All putative chromosomes are completely assembled and include the W and Z chromosomes, as well as a mitochondrial genome. The metrics provided to highlight the quality of the assembly are all appropriate and at the standard for a user to confidently align to this reference.

I do have a couple of very minor recommendations that essentially update the record of resources available, and once they're addressed, I see no reason not to accept this work.

- The manuscript says that the genome will be annotated, though the annotation is currently available. Looking at similar data notes from the DToL project, all that would be required is to mention the pipeline used in the methods and a few bits of info with regard to the numbers of transcribed mRNAs and protein-coding and non-coding genes. Similarly, masked versions of the genome are available at the reported accession - it would be good if the authors could report this and note how it was done and is a useful resource that readers may not be aware is available.
- The authors state that there are 222 filled gaps, though on the iINocFimb1.1 ENA accession, 215 are reported. Regardless of what the number actually is, it would be useful if a bed file of these locations were made available.
- Finally, the authors attribute the identity of the W and Z chromosomes - it would be good if they could report how these were assigned. I suspect it's to do with coverage and subsequently size, but if any other checks were performed, it would be good to know.

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?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Currently cancer genomics, previously invertebrate genomics and population genetics. I contributed towards the *Helicoverpa armigera* genome project and was responsible for a chromosome scale assembly using linkage mapping that contributed to whole-genome sequencing populations.

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 04 May 2022

<https://doi.org/10.21956/wellcomeopenres.19340.r49845>

© 2022 van den Heuvel J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Joost van den Heuvel

Laboratory of Genetics, Wageningen University & Research, Wageningen, The Netherlands

Authors present a short genome report on the broad-bordered yellow underwing which assembled to chromosome level. This assembly is an impressive piece of work and a marked leap of genomic knowledge since before this study only genetic information was present for the mitochondrion. The current assembly allows for comparison with the closely related species *Noctua pronuba*, as well as other recently published related species of the Noctuid family.

The data availability section mentions an annotation by the Ensembl pipeline. With that this resource will prove to be valuable to people that want to study this species and work in fields ranging from development, genetics up to ecology.

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: population genomics, bioinformatics, evolution, Lepidoptera, Diptera

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.
