




DATA NOTE

# The genome sequence of the early grey, *Xylocampa areola* (Esper, 1789) [version 1; peer review: 2 approved]

David Lees <sup>1</sup>, Douglas Boyes<sup>2+</sup>,  
 Natural History Museum Genome Acquisition Lab,  
 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

<sup>1</sup>Natural History Museum, London, UK

<sup>2</sup>UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK

+ Deceased author

**V1** First published: 23 Dec 2022, 7:312  
<https://doi.org/10.12688/wellcomeopenres.18663.1>  
 Latest published: 23 Dec 2022, 7:312  
<https://doi.org/10.12688/wellcomeopenres.18663.1>

## Abstract

We present a genome assembly from an individual *Xylocampa areola* (the early grey; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 565 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the assembled Z sex chromosome. The mitochondrial genome has also been assembled and is 15.5 kilobases in length. Gene annotation of this assembly on Ensembl identified 18,869 protein coding genes.

## Keywords



*Xylocampa areola*, the early grey, genome sequence, chromosomal, Lepidoptera





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

## Open Peer Review

Approval Status  

	1	2
<b>version 1</b>		
23 Dec 2022	<a href="#">view</a>	<a href="#">view</a>

1. **Lars Höök** , Uppsala University, Uppsala, Sweden

2. **Fernando Seixas** , Harvard University, Cambridge, Massachusetts, USA

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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **Lees D:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Boyes D:** Investigation, Resources;

**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:** © 2022 Lees 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:** Lees D, Boyes D, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the early grey, *Xylocampa areola* (Esper, 1789) [version 1; peer review: 2 approved]** Wellcome Open Research 2022, 7:312 <https://doi.org/10.12688/wellcomeopenres.18663.1>

**First published:** 23 Dec 2022, 7:312 <https://doi.org/10.12688/wellcomeopenres.18663.1>

## Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea; Noctuidae; Amphipyriinae; *Xylocampa*; *Xylocampa areola* (Esper, 1789) (NCBI:txid1870430)

## Background

The early grey, *Xylocampa areola*, is a medium sized noctuid moth, intricately grey mottled with conjoined stigmata on the forewing, that starts to emerge early in the temperate moth season (usually February in the UK). The early grey is found in deciduous woodland, hedges and a wide range of other habitats, including heathland, fens, and gardens, wherever its larval foodplant, species of *Lonicera* (Caprifoliaceae) and especially *L. xylosteum*, is found. Despite its larval preference for honeysuckle, *X. areola* is not generally considered to be an important garden pest.

*Xylocampa areola* is generally common and widespread in the western Palaearctic only, from southern Scandinavia to the northern Mediterranean; but there are relatively few records of its presence in eastern Europe (GBIF Secretariat, 2021). Populations in the UK appear to be stable or even increasing (Conrad *et al.*, 2006; Randle *et al.*, 2019). The adult flies from February, protracted until around the end of May, with a peak in April in the UK (Wheeler, 2022) and up to mid-June in Europe (Kettner, 2019). Adults seek nectar at night, in the Spring particularly willow (*Salicaceae*) blossoms. In the Mediterranean, *X. areola* has been considered a very important Lepidoptera pollinator in the community there studied, especially of *Arbutus unedo* L., Ericaceae (Ribas-Marquès *et al.*, 2022).

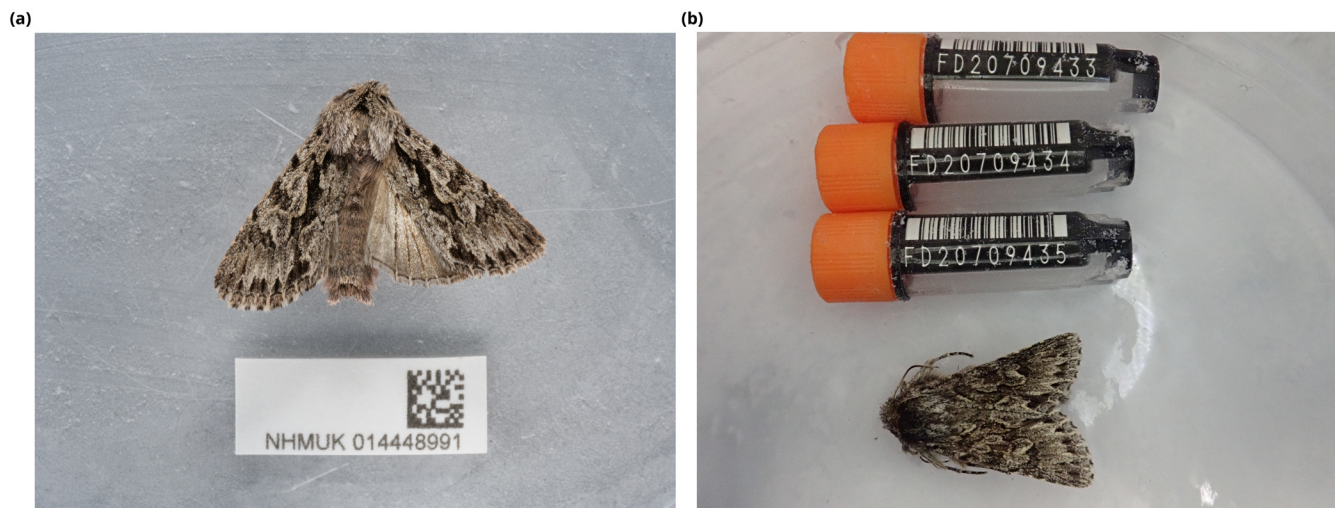
*Xylocampa* appears not to have been used in molecular phylogenies, and it would be interesting to determine its closest relatives. The sister group of the genus is apparently not known.

It is currently placed in the tribe Psaphidini. The genome sequence should not only be useful in phylogeny, but in studies of potentially cryptic species. There are two DNA barcode clusters (BOLD SYSTEMS, no date), i.e., the BINs BOLD:AAE4200 and BOLD:ABZ8138 (the latter recorded from Norway, Spain and France: Corsica, the first and more common of which already DNA barcoded for the UK), and these two discrete clusters are 1.6% divergent. There is only one other species classified in *Xylocampa*: *X. mustapha* (Oberthür, 1910) (BOLD:AEI5117), which is mainly from southern Italy, the Aegean Islands, north-west Africa, and the Middle East. The Italian population is only 2–3% divergent from BOLD:ABZ8138, and about 3.4–3.6% divergent from BOLD:AAE4200.

## Genome sequence report

The genome was sequenced from an individual *X. areola* (Figure 1) collected from High Wycombe, Buckinghamshire (NHMUK014448991, latitude 51.63, longitude -0.74). A total of 58-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 8 missing/misjoins and removed 2 haplotypic duplications, reducing the scaffold number by 4.76%.

The final assembly has a total length of 565 Mb in 40 sequence scaffolds with a scaffold N50 of 20 Mb (Table 1). Most (99.96%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–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.



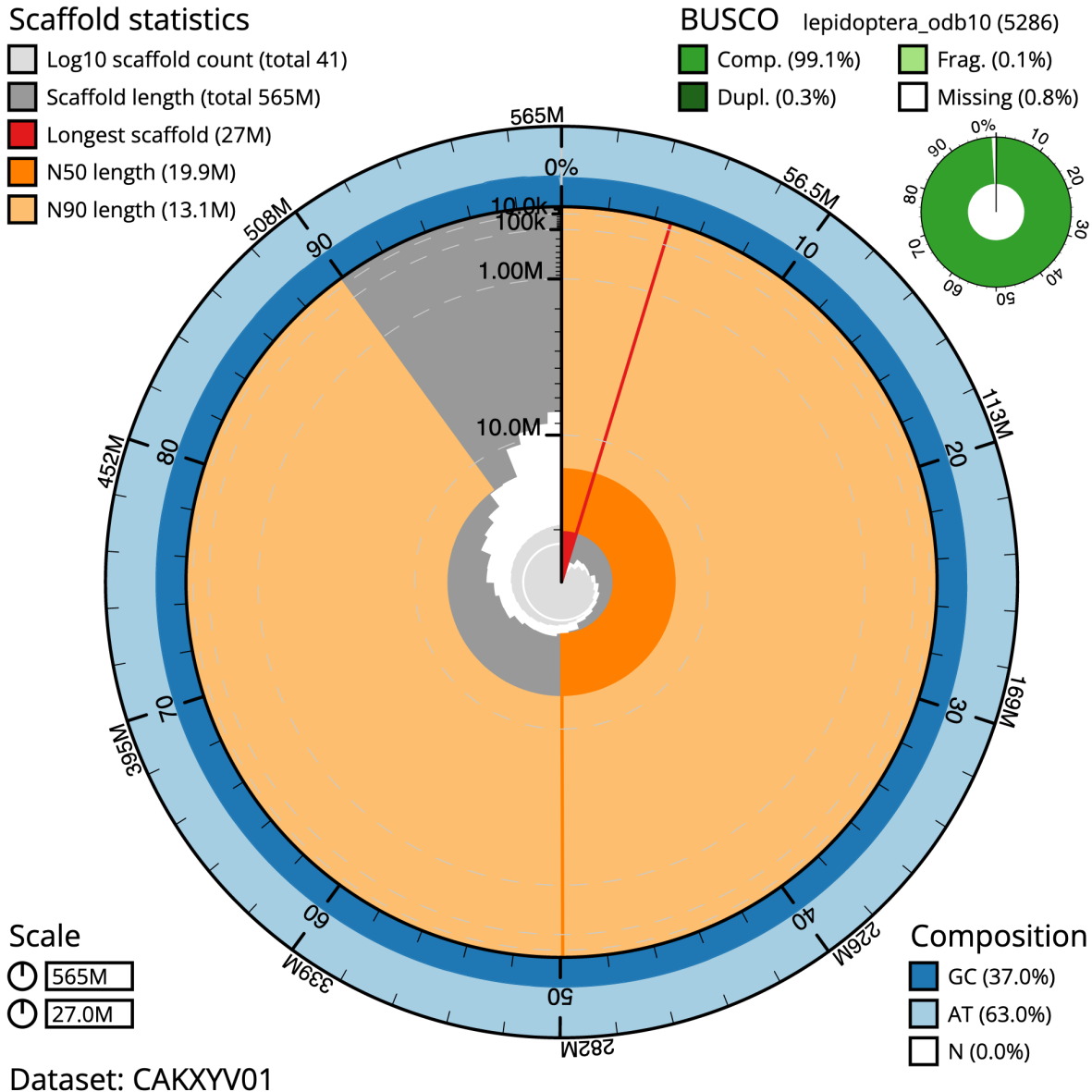
**Figure 1.** **a)** Photograph of the *X. areola* (ilXylAreo1) specimen used for genome sequencing. **b)** Photograph of the *X. areola* (ilXylAreo2) specimen used for RNA sequencing.

**Table 1. Genome data for *X. areola*, iXylAreo1.1.**

<b>Project accession data</b>	
Assembly identifier	iXylAreo1.1
Species	<i>Xylocampa areola</i>
Specimen	iXylAreo1 (DNA sequencing, Hi-C), iXylAreo2 (RNA sequencing)
NCBI taxonomy ID	1870430
BioProject	PRJEB50748
BioSample ID	SAMEA9359445
Isolate information	Thorax tissue (PacBio), Head tissue (Hi-C), Abdomen (RNA-Seq)
<b>Assembly metrics*</b>	
Base pair QV	63.8 (Benchmark: $\geq 50$ )
<i>k</i> -mer completeness	100% (Benchmark: $\geq 95\%$ )
BUSCO**	C:99.1%[S:98.7%,D:0.3%],F:0.1%,M:0.8%,n:5,286 (Benchmark: C $\geq 95\%$ )
Percentage of assembly mapped to chromosomes	99.96% (Benchmark: $\geq 95\%$ )
Sex chromosomes	Z chromosome assembled (Benchmark: localised homologous pairs)
Organelles	Assembled mitochondrion (Benchmark: complete single alleles)
<b>Raw data accessions</b>	
PacificBiosciences SEQUEL II	ERR8575387, ERR8575388
Hi-C Illumina	ERR8571674
PolyA RNA-Seq Illumina	ERR10123671
<b>Genome assembly</b>	
Assembly accession	GCA_935421205.1
<i>Accession of alternate haplotype</i>	GCA_935412865.1
Span (Mb)	565
Number of contigs	46
Contig N50 length (Mb)	19.6
Number of scaffolds	40
Scaffold N50 length (Mb)	19.9
Longest scaffold (Mb)	23.9
<b>Genome annotation</b>	
Number of protein-coding genes	18,869
Number of gene transcripts	19,064

\* 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 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/iXylAreo1.1/dataset/CAKXYV01/busco>.



**Figure 2. Genome assembly of *X. areola*, iIXylAreo1.1: 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 564,568,605 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 (27,003,092 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (19,946,087 and 13,085,905 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/iIXylAreo1.1/dataset/CAKXYV01/snail>.

The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 99.1% (single, 98.7%, duplicated 0.3%) using the lepidoptera\_odb10 reference set.

### Genome annotation report

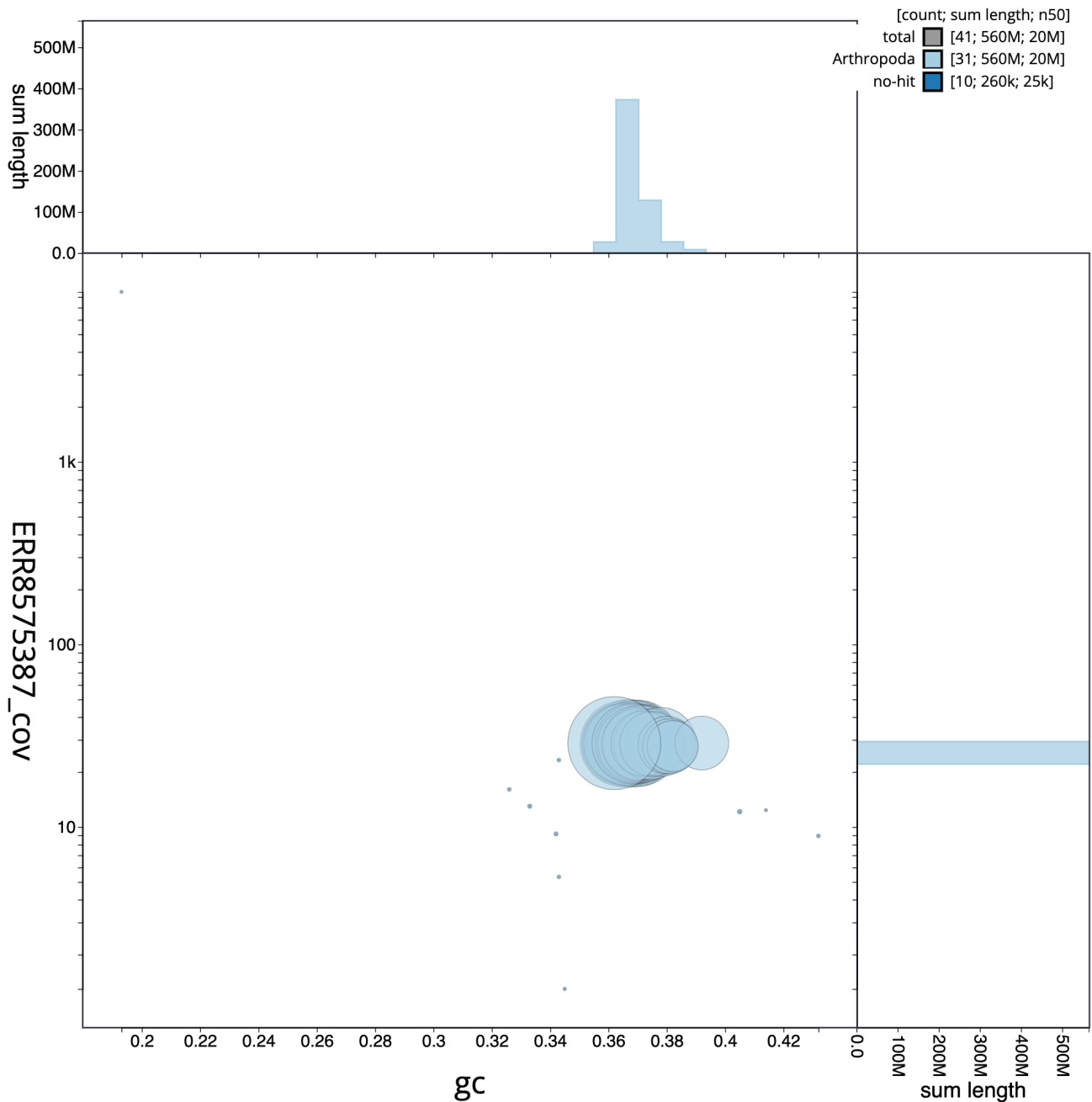
The *Xylocampa areola* genome assembly (GCA\_935421205.1) was annotated using BRAKER2 (Table 1; [https://rapid.ensembl.org/Xylocampa\\_areola\\_GCA\\_935421205.1/Info/Index](https://rapid.ensembl.org/Xylocampa_areola_GCA_935421205.1/Info/Index)). The resulting

annotation includes 19,064 transcribed mRNAs from 18,869 protein-coding genes.

### Methods

#### Sample acquisition and nucleic acid extraction

An unsexed individual *X. areola* (iIXylAreo1) was collected and identified by David Lees (Natural History Museum, London) from High Wycombe, Buckinghamshire, UK (latitude

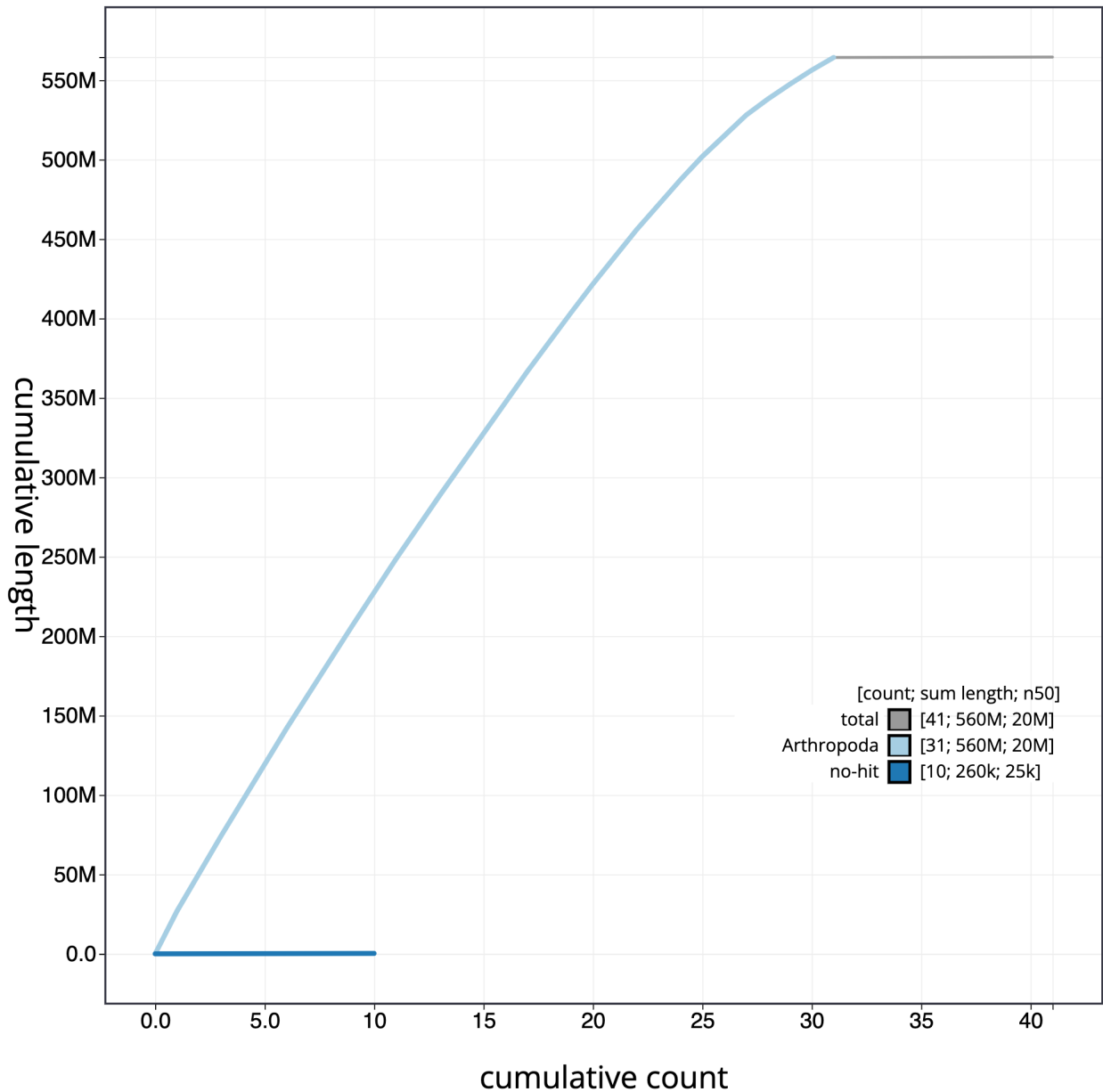


**Figure 3. Genome assembly of *X. areola*, ilXylAreo1.1: GC coverage.** BlobToolKit GC-coverage plot. Chromosomes are coloured by phylum. Circles are sized in proportion to chromosome length. Histograms show the distribution of chromosome length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilXylAreo1.1/dataset/CAKXYV01/blob>.

51.63, longitude  $-0.74$ ). The sample was preserved by dry freezing at  $-80^{\circ}\text{C}$ .

A second unsexed *X. areola* (ilXylAreo2) was collected in Wytham Woods, Berkshire, UK (latitude 51.77, longitude  $-1.34$ ) 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 ilXylAreo1 sample was weighed and dissected on dry ice with head tissue set aside for Hi-C sequencing. The whole body tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular



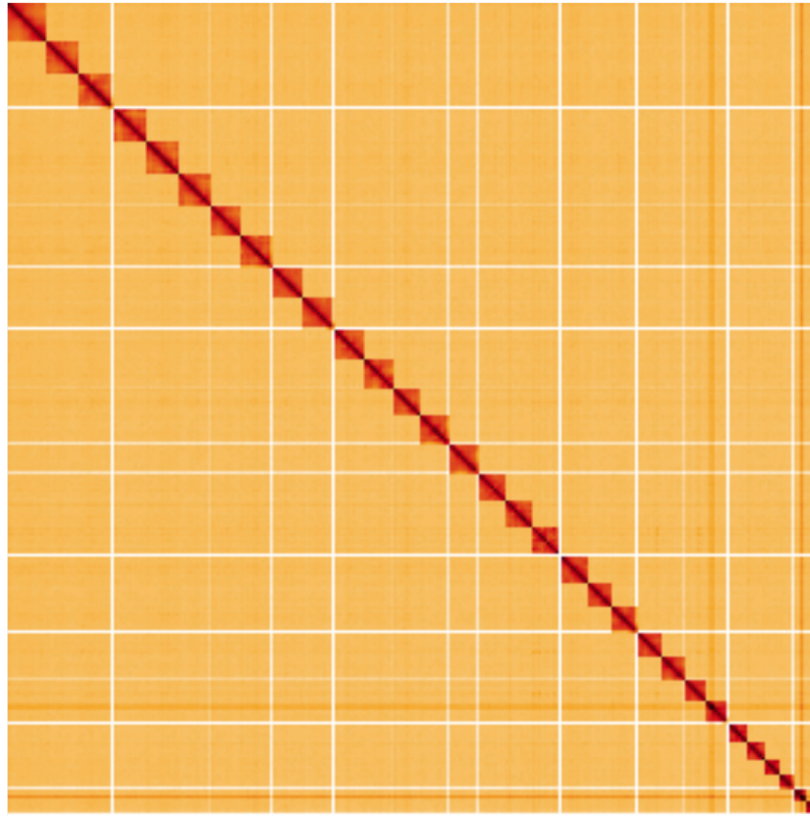
**Figure 4. Genome assembly of *X. areola*, iIXylAero1.1: cumulative sequence.** BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iIXylAero1.1/dataset/CAKXYV01/cumulative>.

weight DNA was removed from a 20 ng aliquot of extracted DNA using 0.8X AMPure XP purification kit. 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 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.

RNA was extracted from the abdomen tissue of iIXylAero2 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50  $\mu$ l RNase-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using





**Figure 5. Genome assembly of *X. areola*, ilXylAero1.1: Hi-C contact map.** Hi-C contact map of the ilXylAero1.1 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=ORLTYiKpR\\_y6hww1MMLGGQ](https://genome-note-higlass.tol.sanger.ac.uk/?d=ORLTYiKpR_y6hww1MMLGGQ).

**Table 2. Chromosomal pseudomolecules in the genome assembly of *X. areola*, ilXylAero1.**

INSDC accession	Chromosome	Size (Mb)	GC%
OW285200.1	1	23.89	36.7
OW285201.1	2	23.07	37
OW285202.1	3	22.89	36.5
OW285203.1	4	22.86	36.9
OW285204.1	5	22.2	36.7
OW285205.1	6	21.78	36.9
OW285206.1	7	21.49	36.9
OW285207.1	8	21.34	36.7
OW285208.1	9	21.08	36.5
OW285209.1	10	20.92	36.6
OW285210.1	11	20.2	36.9
OW285211.1	12	19.95	36.8
OW285212.1	13	19.6	36.5
OW285213.1	14	19.58	36.8
OW285214.1	15	19.54	36.8
OW285215.1	16	19.21	36.9

INSDC accession	Chromosome	Size (Mb)	GC%
OW285216.1	17	18.55	37.1
OW285217.1	18	18.5	37.2
OW285218.1	19	18.03	37.1
OW285219.1	20	17.17	37.1
OW285220.1	21	17.1	36.9
OW285221.1	22	15.88	37
OW285222.1	23	15.48	37.4
OW285223.1	24	14.65	37.8
OW285224.1	25	13.09	37.2
OW285225.1	26	12.97	37.5
OW285226.1	27	10.21	38
OW285227.1	28	9.24	38.1
OW285228.1	29	8.89	39.2
OW285229.1	30	7.98	38.2
OW285199.1	Z	27	36.2
OW285230.1	MT	0.02	19.4
-	-	0.24	36.6



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 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 were performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from head tissue of *iXylAreol* using the Arima v2 kit and sequenced on the 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). The assembly was scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2022). The assembly was checked for contamination as described previously (Howe *et al.*, 2021). Manual curation (Howe *et al.*, 2021) was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). 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 were generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

**Table 3. Software tools and versions used.**

Software tool	Version	Source
BlobToolKit	3.4.0	Challis <i>et al.</i> , 2020
Hifiasm	0.16.1-r375	Cheng <i>et al.</i> , 2021
HiGlass	1.11.6	Kerpedjiev <i>et al.</i> , 2018
PretextView	0.2.x	Harry, 2022
purge_dups	1.2.3	Guan <i>et al.</i> , 2020
MitoHiFi	2.0	Uliano-Silva <i>et al.</i> , 2021
YaHS	yahs-1.1.91eebc2	Zhou <i>et al.</i> , 2022

### Genome annotation

BRAKER2 (Brûna *et al.*, 2021) was used to generate annotation for the *Xylocampa areola* assembly (GCA\_935421205.1). 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: *Xylocampa areola*. Accession number PRJEB50748; <https://identifiers.org/ena.embl/PRJEB50748> (Wellcome Sanger Institute, 2022).

The genome sequence is released openly for reuse. The *Xylocampa areola* 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 Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.4790042>.

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

- 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](#)
- BOLD SYSTEMS (no date): **TAXONOMY BROWSER: *Xylocampa areola*.** (Accessed: 20 November 2022). [Reference Source](#)
- Brûna 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](#)

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

[Publisher Full Text](#)

GBIF Secretariat: **Xylocampa areola (Esper, 1789)**. 2021; (Accessed: 20 November 2022).

[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](#)

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](#)

Kettner M: **Xylocampa areola (Esper, 1789) honeysuckle owl.** 2019; (Accessed: 20 November 2022).

[Reference Source](#)

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](#)

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths.** Newbury: NatureBureau. 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): 1665–1680.

[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](#)

Ribas-Marquès E, Díaz-Calafat J, Boi M: **The role of adult noctuid moths (Lepidoptera: Noctuidae) and their food plants in a nocturnal pollen-transport network on a Mediterranean island.** *J Insect Conserv*. 2022; **26**(2): 243–255.

[Publisher Full Text](#)

Uliano-Silva M, *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](#)

Wellcome Sanger Institute: **The genome sequence of the Xylocampa areola (Esper, 1789).** European Nucleotide Archive. [Dataset]. 2022.

Wheeler J: **Early Grey (Xylocampa areola).** norfolkmoths.co.uk. 2022; (Accessed: 20 November 2022).

[Reference Source](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *bioRxiv*. [Preprint], 2022.

[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 29 November 2023

<https://doi.org/10.21956/wellcomeopenres.20696.r68071>

© 2023 Seixas F. 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.

 **Fernando Seixas** 

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

The authors assembled the genome of one *Xylocampa areola* individual, using PacBio and Hi-C data. This resulted in a complete (as attested by the BUSCO score) and highly contiguous (i.e. chromosome-level) assembly. I found the writing very clear, and the authors always provide complete information to understand their work. The one point that could be clarified is in the "Genome annotation" section, where the authors say "Annotation was created *primarily* through alignment of transcriptomic data". The *primarily* in that sentence makes me believe that other information was used (I could presume homology-based and/or de novo annotation) so it would be helpful if the authors could explain what other sources of evidence were used. Otherwise, I have no other remarks.

**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 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 27 November 2023

<https://doi.org/10.21956/wellcomeopenres.20696.r69790>

© 2023 Höök L. 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.



**Lars Höök** 

Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden

In this data note, the authors present the genome assembly of *Xylocampa areola*, the early grey moth. The assembly consist of 31 chromosome scaffolds which is common for a Lepidoptera species.

The methods are well described and follow current standards for generating high quality genome assemblies. The genome assembly, as described in the note, is of high quality and is therefore a valuable resource for studies of evolution and phylogenetics.

The relevant information describing the assembly is clearly presented and the genome data is made available at ENA.

**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 genetics

**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.**

---