DATA NOTE

The genome sequence of the Lilac Beauty, *Apeira syringaria* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes[1+], Owen T. Lewis[2],
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

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK

[+] Deceased author

v1 **First published:** 12 Apr 2023, **8**:159
https://doi.org/10.12688/wellcomeopenres.19208.1
Latest published: 12 Apr 2023, **8**:159
https://doi.org/10.12688/wellcomeopenres.19208.1

Abstract

We present a genome assembly from an individual male *Apeira syringaria* (the Lilac Beauty; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 544.4 megabases in span. Most of the assembly is scaffolded into 30 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 18,426 protein coding genes.

Keywords

Apeira syringaria, Lilac Beauty, genome sequence, chromosomal, Lepidoptera

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 12 Apr 2023	✓ view	✓ view

1. **Jurate De Prins**, Royal Belgian Institute of Natural Sciences, Brussels, Belgium

2. **Violaine Llaurens**, Museum National d'Histoire Naturelle, Paris, France

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; **Lewis OT**: 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, 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.

How to cite this article: Boyes D, Lewis OT, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Lilac Beauty, *Apeira syringaria* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:159 https://doi.org/10.12688/wellcomeopenres.19208.1

First published: 12 Apr 2023, **8**:159 https://doi.org/10.12688/wellcomeopenres.19208.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Geometroidea; Geometridae; Ennominae; *Apeira*; *Apeira syringaria* (Linnaeus, 1758) (NCBI:txid934915).

Background

The Lilac Beauty, *Apeira syringaria* (Linnaeus, 1758) is a moth in the family Geometridae, from the 'thorn' subfamily, Ennominae. Adult moths of this species have an unusual resting posture, with the forewings slightly raised, and the leading edge slightly folded (Waring *et al.*, 2017), increasing their resemblance to a crumpled dead leaf. Males are smaller and more brightly coloured than females (South, 1961).

Apeira syringaria has a local distribution in Britain and Ireland, occurring mostly in south and central areas. It was not recorded from Scotland in the early part of the twentieth century (South, 1961), but has extended its distribution there in recent decades (Randle *et al.*, 2019). At monitored sites, the abundance of this species has declined greatly since 1970 (Randle *et al.*, 2019). Internationally, the distribution of *A. syringaria* extends across Europe and temperate Asia (GBIF Secretariat, 2022).

The main larval foodplants include honeysuckle (*Lonicera* spp.), privet (*Ligustrum* spp.), ash (*Fraxinus* spp.) and lilac (*Syringa vulgaris*) among other trees and shrubs (Henwood *et al.*, 2020).

A genome sequence for *Apeira syringaria* will contribute to a growing data set of resources for understanding Lepidopteran biology. The genome of *Apeira syringaria* 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 *Apeira syringaria,* based on one female specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one female *Apeira syringaria* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.34). A total of 49-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data.

The final assembly has a total length of 544.4 Mb in 51 sequence scaffolds with a scaffold N50 of 21.0 Mb (Table 1). Most (99.96%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 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). The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.4% (single

Figure 1. Photograph of the *Apeira syringaria* (ilApeSyri1) specimen used for genome sequencing.

97.7%, duplicated 0.7%), using the lepidoptera_odb10 reference set. 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 *Apeira syringaria* genome assembly GCA_934044485.1 (ilApeSyri1.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; Ensembl accession number GCA_934044485.1). The resulting annotation includes 18,426 protein-coding and 18,577 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A female *Apeira syringaria* specimen (ilApeSyri1) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire) (latitude 51.77, longitude –1.34) on 13 June 2020. The specimen was taken from woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The specimen was identified by Douglas Boyes using field ID and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilApeSyri1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Head and thorax 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. 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

Table 1. Genome data for *Apeira syringaria*, ilApeSyri1.1.

Project accession data		
Assembly identifier	ilApeSyri1.1	
Species	*Apeira syringaria*	
Specimen	ilApeSyri1	
NCBI taxonomy ID	934915	
BioProject	PRJEB50739	
BioSample ID	SAMEA7520685	
Isolate information	ilApeSyri1; female, head and thorax (PacBio and Hi-C)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	64.8	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.4%[S:97.7%,D:0.7%], F:0.5%,M:1.0%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.96%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR8575375	
Hi-C Illumina	ERR8571657	
Genome assembly		
Assembly accession	GCA_934044485.1	
Accession of alternate haplotype	GCA_934045895.1	
Span (Mb)	544.4	
Number of contigs	51	
Contig N50 length (Mb)	21.0	
Number of scaffolds	51	
Scaffold N50 length (Mb)	21.0	
Longest scaffold (Mb)	37.7	
Genome annotation		
Number of protein-coding genes	18,426	
Number of non-coding genes	18,577	

* 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/ilApeSyri1.1/dataset/CAKOGW01/busco.

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'

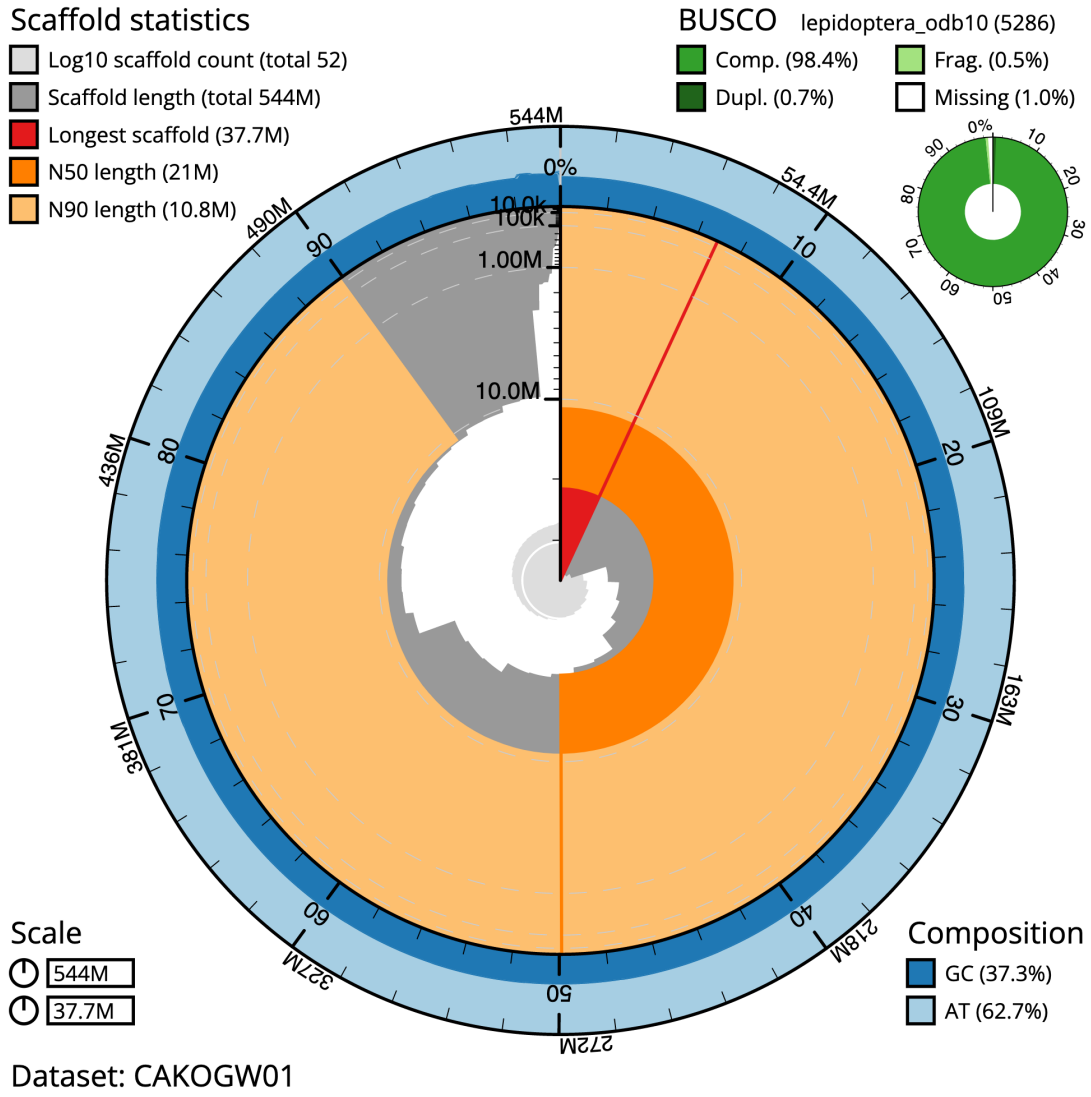

Figure 2. Genome assembly of *Apeira syringaria*, ilApeSyri1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 544,443,574 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (37,666,467 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (20,999,187 and 10,846,250 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/ilApeSyri1.1/dataset/CAKOGW01/snail.

instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL II (HiFi) instruments. Hi-C data were also generated from tissue of ilApeSyri1 using the Arima v2 kit and sequenced on the HiSeq X Ten 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 then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination 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.*, 2022), which performed annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within

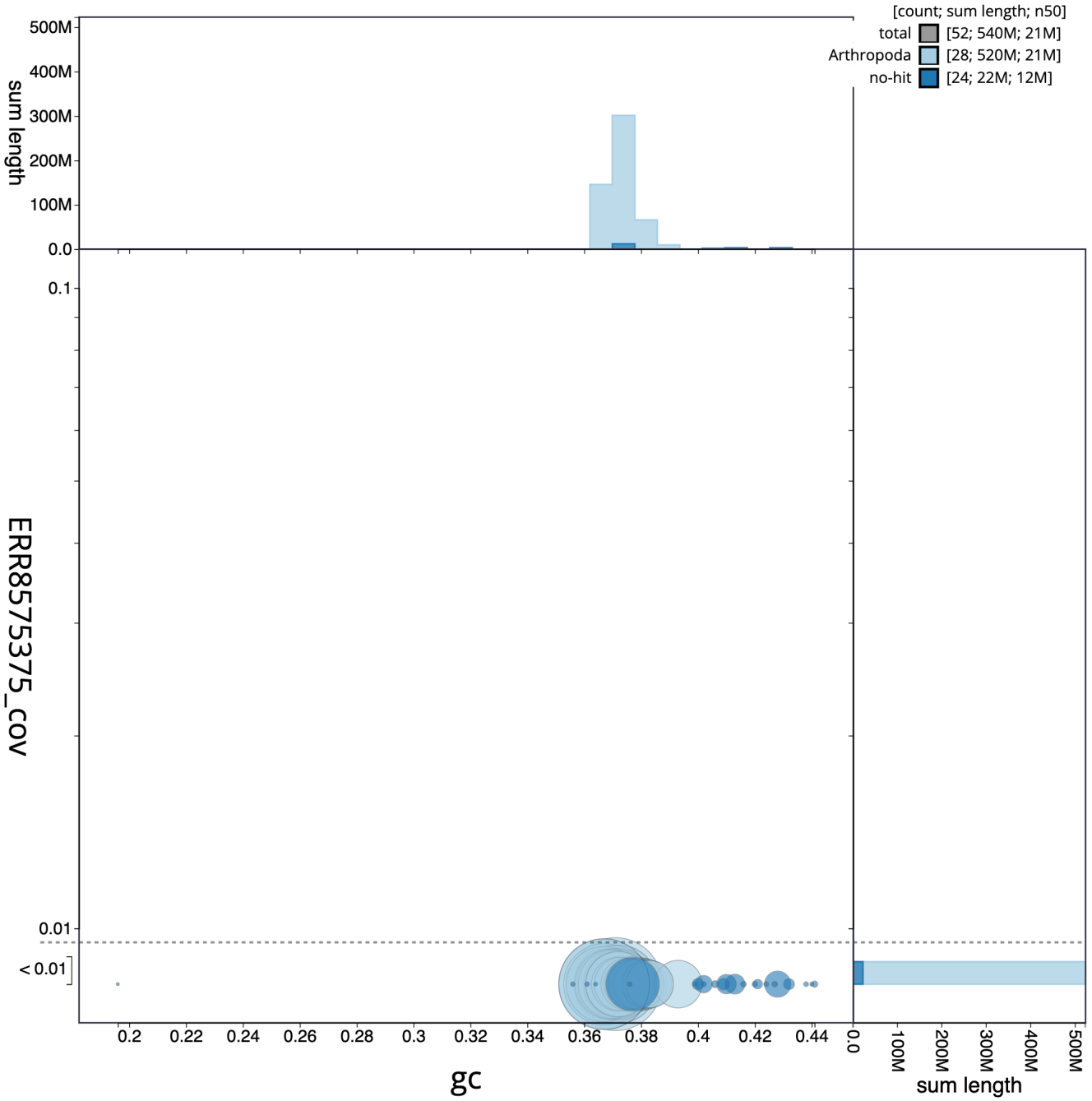

Figure 3. Genome assembly of *Apeira syringaria*, ilApeSyri1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilApeSyri1.1/dataset/CAKOGW01/blob.

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

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Apeira*

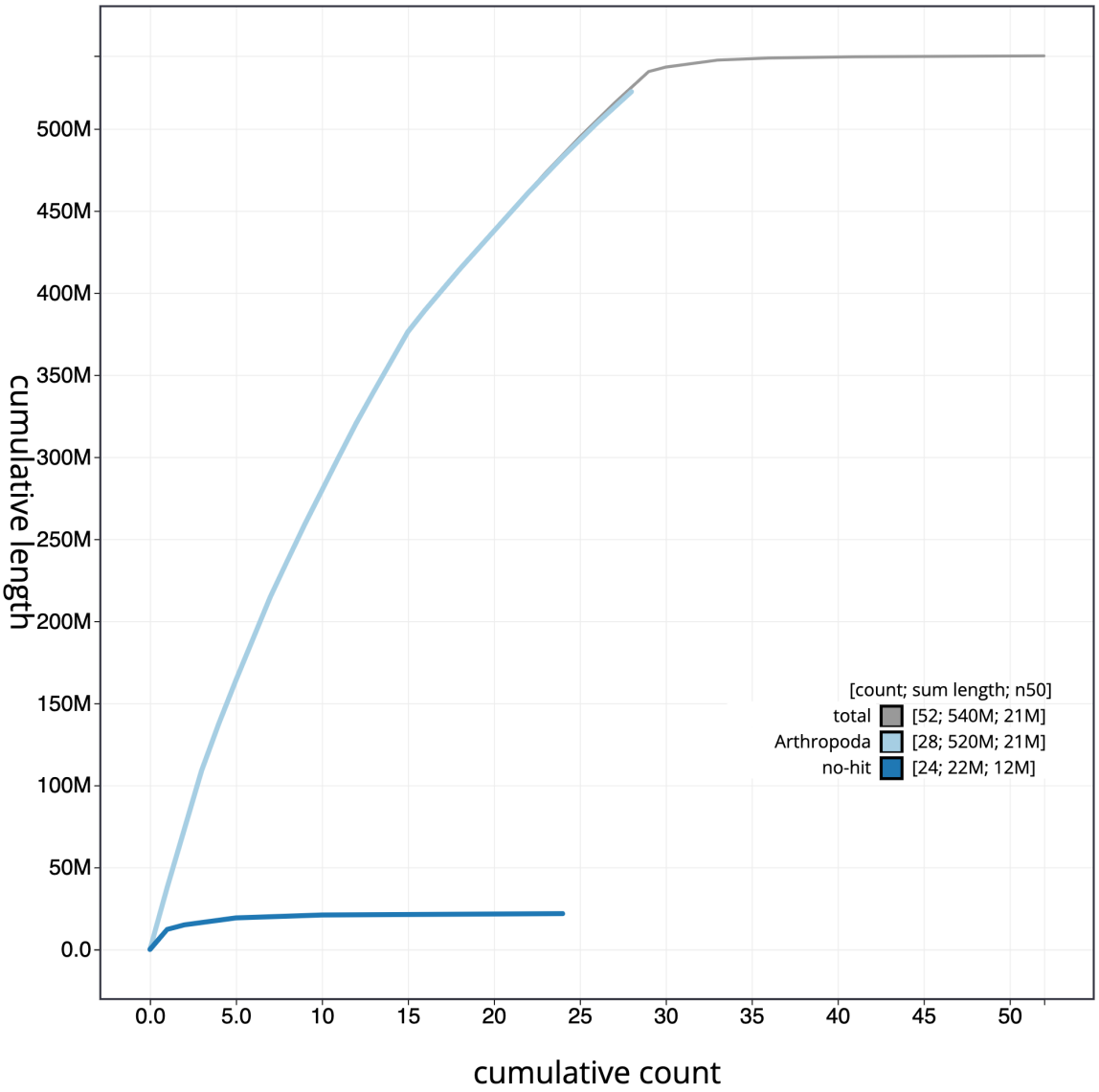

Figure 4. Genome assembly of *Apeira syringaria*, ilApeSyri1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilApeSyri1.1/dataset/CAKOGW01/cumulative.

syringaria assembly (GCA_934044485.1) in Ensembl Rapid Release.

Ethics and 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 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. All efforts are undertaken to minimise the suffering of animals used for sequencing. 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.

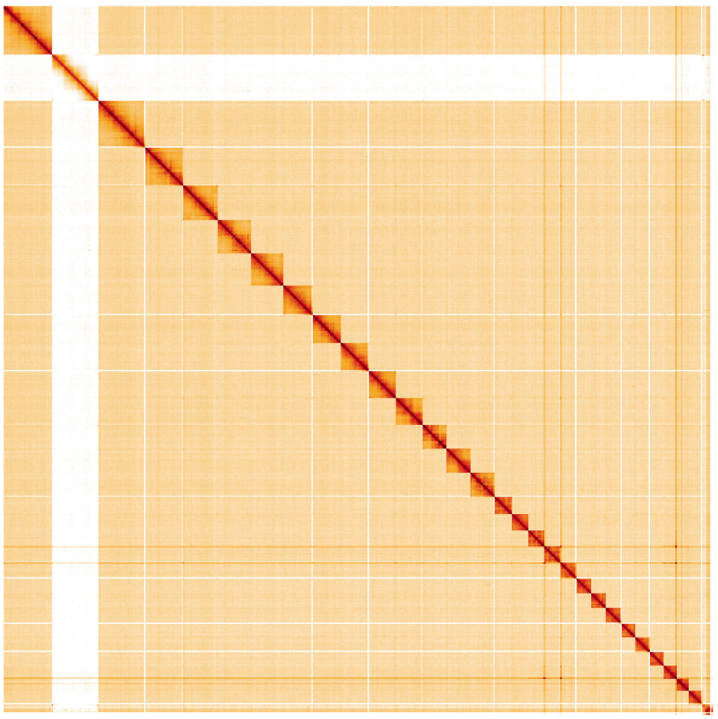

Figure 5. Genome assembly of *Apeira syringaria*, ilApeSyri1.1: Hi-C contact map. Hi-C contact map of the ilApeSyri1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=BdknbMWvQoywaKY2jyxqZA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Apeira syringaria*, ilApeSyri1.

INSDC accession	Chromosome	Size (Mb)	GC%
OW203654.1	1	37.67	37.1
OW203656.1	2	35.28	36.7
OW203657.1	3	28.52	36.7
OW203658.1	4	26.6	37.2
OW203659.1	5	25.94	37
OW203660.1	6	24.87	36.7
OW203661.1	7	22.17	37
OW203662.1	8	21.85	37.2
OW203663.1	9	21	36.9
OW203664.1	10	20.5	37.2
OW203665.1	11	20.36	37
OW203666.1	12	18.76	37.2
OW203667.1	13	18.56	37.2
OW203668.1	14	18.19	37.2
OW203669.1	15	13.3	37.7
OW203670.1	16	12.4	37.2
OW203671.1	17	12.31	37.3
OW203672.1	18	12.2	37.7
OW203673.1	19	11.85	38
OW203674.1	20	11.81	37.9
OW203675.1	21	11.58	37.8
OW203676.1	22	11.43	37.5
OW203677.1	23	11.03	37.6
OW203678.1	24	10.85	37.9
OW203679.1	25	10.3	37.7
OW203680.1	26	10.27	38.2
OW203681.1	27	9.64	39.3
OW203682.1	28	9.55	38.3
OW203683.1	29	2.78	42.8
OW203655.1	Z	36.04	36.7
OW203684.1	MT	0.02	19.7
-	unplaced	6.82	41

Table 3. Software tools and versions used.

Software tool	Version	Source
BlobToolKit	4.0.7	Challis *et al.*, 2020
Hifiasm	0.16.1-r375	Cheng *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
MitoHiFi	2	Uliano-Silva *et al.*, 2022
PretextView	0.2	Harry, 2022
purge_dups	1.2.3	Guan *et al.*, 2020
YaHS	yahs-1.1.91eebc2	Zhou *et al.*, 2023

Data availability

European Nucleotide Archive: *Apeira syringaria* (lilac beauty). Accession number PRJEB50739; https://identifiers.org/ena.embl/PRJEB50739 (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Apeira syringaria* 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.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

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

GBIF Secretariat: **Apeira syringaria (Linnaeus, 1758)**. GBIF Backbone Taxonomy. 2022; (Accessed: 25 February 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

Henwood B, Sterling P, Lewington R: **Field Guide to the Caterpillars of Great Britain and Ireland**. London: Bloomsbury. 2020.
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *Gigascience.* Oxford University Press. 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

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

South R: **Moths of the British Isles**. New edition. London: Frederick Warne and Co. 1961.
Reference Source

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *bioRxiv.* [Preprint]. 2022.
Publisher Full Text

Waring P, Townsend M, Lewington R: **Field Guide to the Moths of Great Britain and Ireland.** Third Edition. Bloomsbury Wildlife Guides. 2017.
Reference Source

Wellcome Sanger Institute: **The genome sequence of the Lilac Beauty, *Apeira syringaria* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB50739, 2022.

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* Edited by C. Alkan, 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 25 May 2024

https://doi.org/10.21956/wellcomeopenres.21290.r70920

✔ **Violaine Llaurens**
Institut de Systématique, Evolution et Biodiversité, Museum National d'Histoire Naturelle, Paris, Île-de-France, France

The article carefully describes the sequencing, assembly and annotation of the genome of a lilac beauty female, following a state-of-the-art workflow. I have only two questions

1. Did the author specifically search for the W chromosome?
2. In figure 5, we can clearly see that one of the scaffold does not seem to have any contact with the others. This is quite unusual to me. Is there any biological and/or technical reason that my explain this uncommon feature of the Hi-C map?

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 ecology, population 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 09 November 2023

https://doi.org/10.21956/wellcomeopenres.21290.r69167

Jurate De Prins
Royal Belgian Institute of Natural Sciences, Brussels, Belgium

My review is only from the taxonomic point of view, for molecular details please apply to a molecular specialist. From the taxonomic point of view, this article looks ok. Just a couple of small details:

- I would suggest to explain in more detail what the authors mean by "an unusual resting posture". What is usual then in Geometridae? Geometridae moths, differently from the majority of other moths, rest with open wings.

- It would be good to mention that *Apeira syringaria* (Linnaeus, 1758) is widely distributed in Europe.

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: Systematics, biodiversity, taxonomy, bioinformatics, Lepidoptera

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.