DATA NOTE

The genome sequence of the Sycamore Piercer, *Pammene aurita* (Razowski, 1991) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], James Hammond[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**:160
https://doi.org/10.12688/wellcomeopenres.19243.1
Latest published: 12 Apr 2023, **8**:160
https://doi.org/10.12688/wellcomeopenres.19243.1

Abstract

We present a genome assembly from an individual male *Pammene aurita* (the Sycamore Piercer; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence is 1,041.8 megabases in span. Most of the assembly is scaffolded into 28 chromosomal pseudomolecules, including the assembled Z sex chromosome. The mitochondrial genome has also been assembled and is 16.7 kilobases in length.

Keywords

Pammene aurita, Sycamore Piercer, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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; **Hammond J**: 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, Hammond J, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Sycamore Piercer, *Pammene aurita* (Razowski, 1991) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:160 https://doi.org/10.12688/wellcomeopenres.19243.1

First published: 12 Apr 2023, **8**:160 https://doi.org/10.12688/wellcomeopenres.19243.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Tortricoidea; Tortricidae; Olethreutinae; Grapholitini; *Pammene; Pammene aurita* (Razowski, 1991) (NCBI:txid1870148).

Background

The Sycamore Piercer *Pammene aurita* (Razowski, 1992) is a moth of the Tortricidae family. The adults exhibit attractive light orange markings on the forewings that fade to a dull brown towards the head. There is little variation in these markings (Bradley *et al.*, 1979). The English vernacular name of this species is derived from the feeding habits of the larvae, which feed internally in the seeds of Sycamore (*Acer pseudoplatanus),* 'piercing' the seed and leaving a neat round hole when the larva exits the seed to pupate (Bradley *et al.*, 1979; Elliott *et al.*, 2018; Hancock *et al.*, 2015). Larvae feed between August and September, and the moth overwinters as a pupa, choosing dead wood or bark as a pupation site in captivity (Bradley *et al.*, 1979; Elliott *et al.*, 2018; Hancock *et al.*, 2015). Adults can be found between June and August, flying around Sycamores in the afternoon, and coming to light after dark (Elliott *et al.*, 2018).

Globally, *P. aurita* is confined to northern and central Europe (GBIF Secretariat, 2022; Hancock *et al.*, 2015). This species is a relatively recent addition to the British and Irish fauna, apparently colonising south-eastern England in the mid-20th century (Bradley *et al.*, 1979; Hancock *et al.*, 2015). Since then, the moth has spread across England and Wales, and has been recorded in eastern Ireland (Elliott *et al.*, 2018).

The genome of *Pammene aurita* 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 *Pammene aurita*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Pammene aurita* specimen (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.34). A total of 24-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 76 missing joins or mis-joins and removed 19 haplotypic duplications, reducing the assembly length by 1.13% and the scaffold number by 32.47%.

The final assembly has a total length of 1,041.8 Mb in 52 sequence scaffolds with a scaffold N50 of 37.5 Mb (Table 1). Most (99.81%) of the assembly sequence was assigned to 28 chromosomal-level scaffolds, representing 27 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.

The estimated Quality Value (QV) of the final assembly is 62.5 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.1% (single 97.3%, duplicated 0.8%) using the lepidoptera_odb10 reference set (n = 5,286).

Methods

Sample acquisition and nucleic acid extraction

A male *Pammene aurita* specimen (ilPamAuri1) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude –1.34) on 24 July 2021. The specimen was taken from woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The specimen was identified by the collector and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilPamAuri1 sample was weighed

Figure 1. Photograph of the *Pammene aurita* (ilPamAuri1) specimen used for genome sequencing.

Table 1. Genome data for *Pammene aurita*, ilPamAuri1.1.

Project accession data		
Assembly identifier	ilPamAuri1.1	
Species	*Pammene aurita*	
Specimen	ilPamAuri1	
NCBI taxonomy ID	1870148	
BioProject	PRJEB55029	
BioSample ID	SAMEA10979080	
Isolate information	ilPamAuri1, male, whole organism (genome sequencing and Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	62.5	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.1%[S:97.3%,D:0.8%], F:0.5%,M:1.5%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.81%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10008907	
Hi-C Illumina	ERR10015064	
Genome assembly		
Assembly accession	GCA_947086415.1	
Accession of alternate haplotype	GCA_947086445.1	
Span (Mb)	1,041.8	
Number of contigs	245	
Contig N50 length (Mb)	9.2	
Number of scaffolds	51	
Scaffold N50 length (Mb)	37.5	
Longest scaffold (Mb)	72.7	

* 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/ilPamAuri1.1/dataset/CAMTYY01/busco.

and dissected on dry ice with tissue set aside for Hi-C sequencing. Whole organism 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

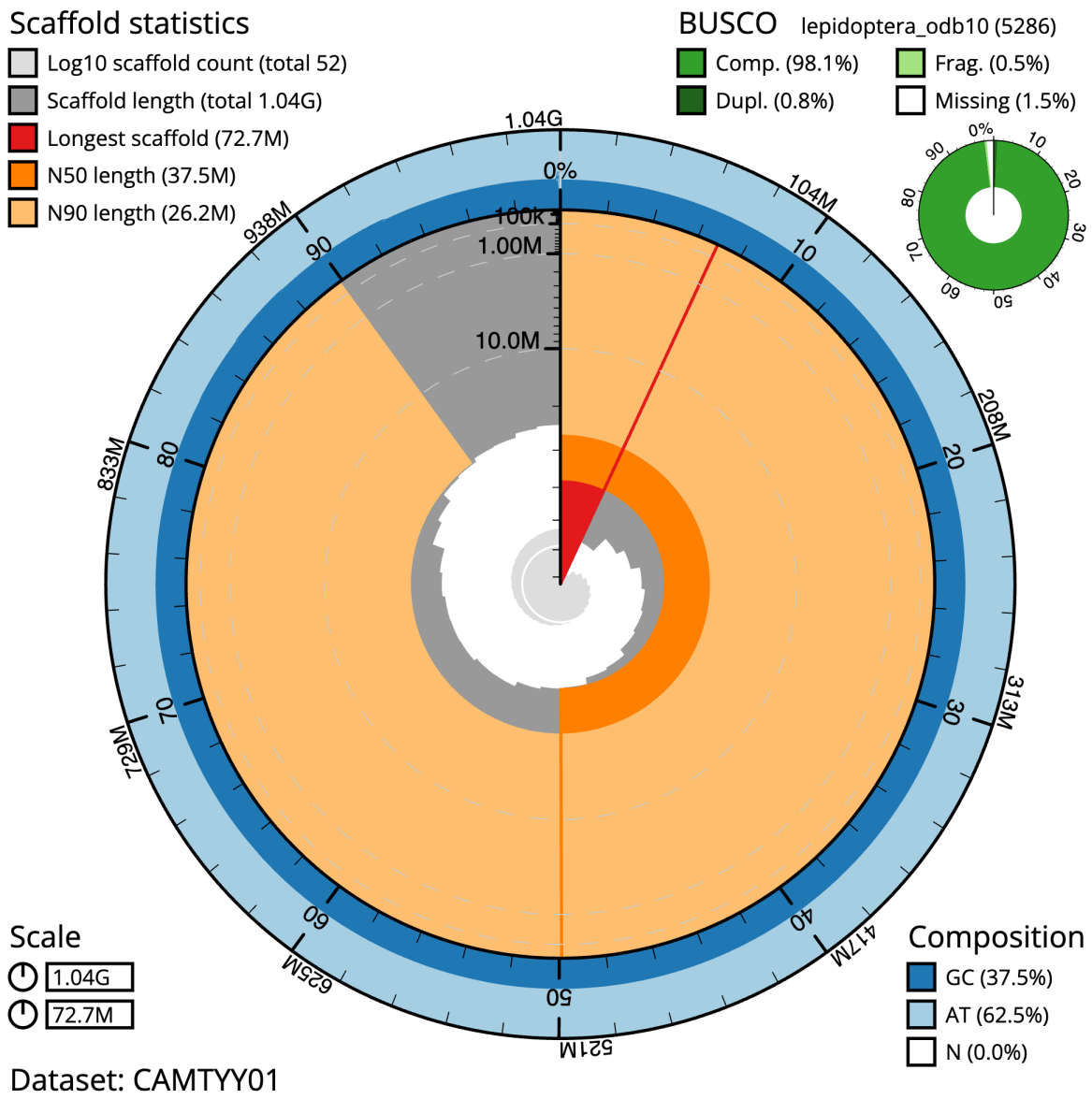

Figure 2. Genome assembly of *Pammene aurita*, ilPamAuri1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,041,812,917 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (72,712,871 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (37,501,620 and 26,227,413 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/ilPamAuri1.1/dataset/CAMTYY01/snail.

Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) instrument. Hi-C data were also generated from tissue of ilPamAuri1 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with

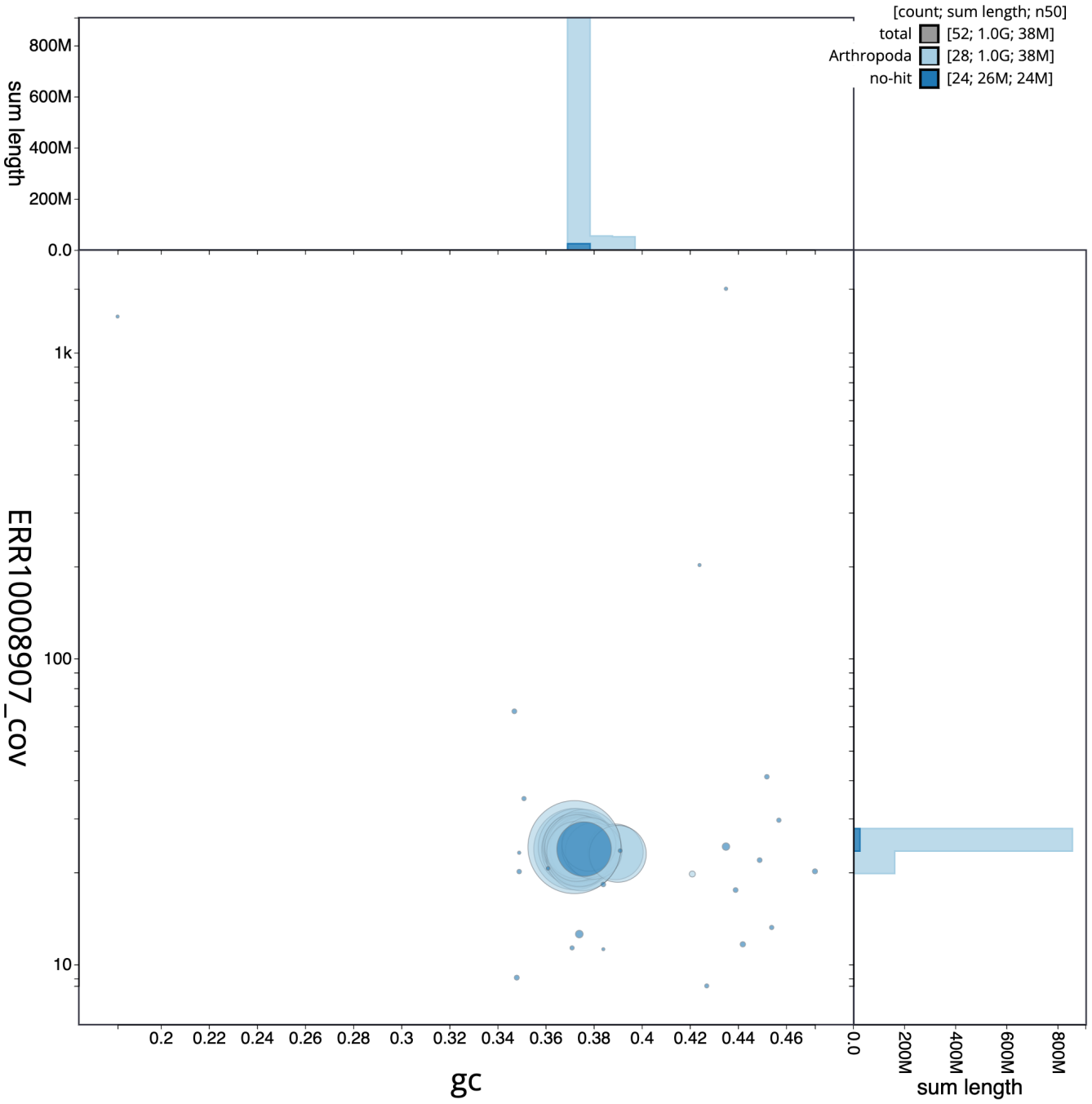

Figure 3. Genome assembly of *Pammene aurita*, ilPamAuri1.1: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilPamAuri1.1/dataset/CAMTYY01/blob.

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). To evaluate the assembly, MerquryFK was used to estimate consensus quality (QV) scores and *k*-mer completeness (Rhie *et al.*, 2020). The genome was analysed, and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were generated within

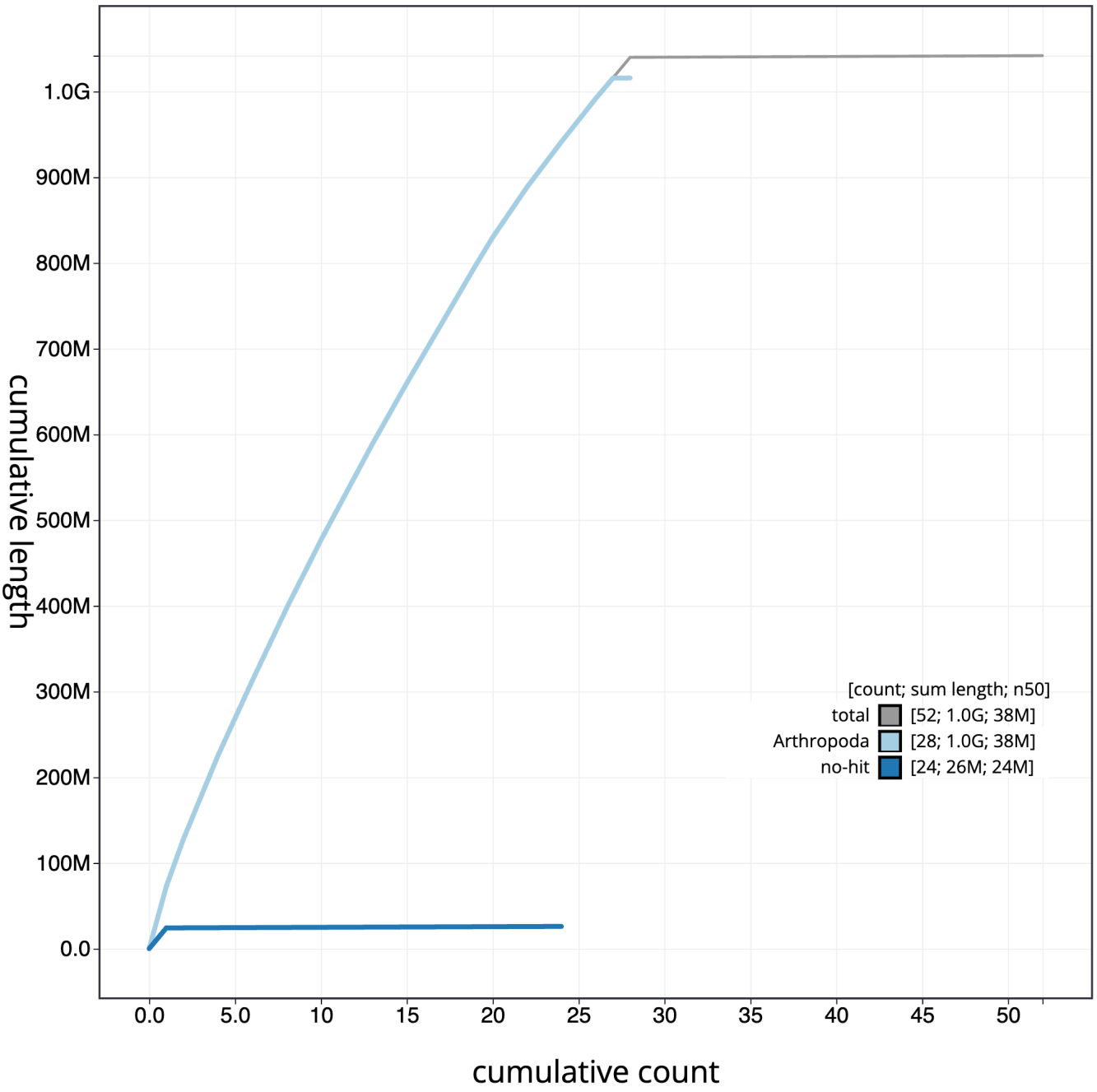

Figure 4. Genome assembly of *Pammene aurita*, ilPamAuri1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilPamAuri1.1/dataset/CAMTYY01/cumulative.

the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of software tool versions and sources.

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

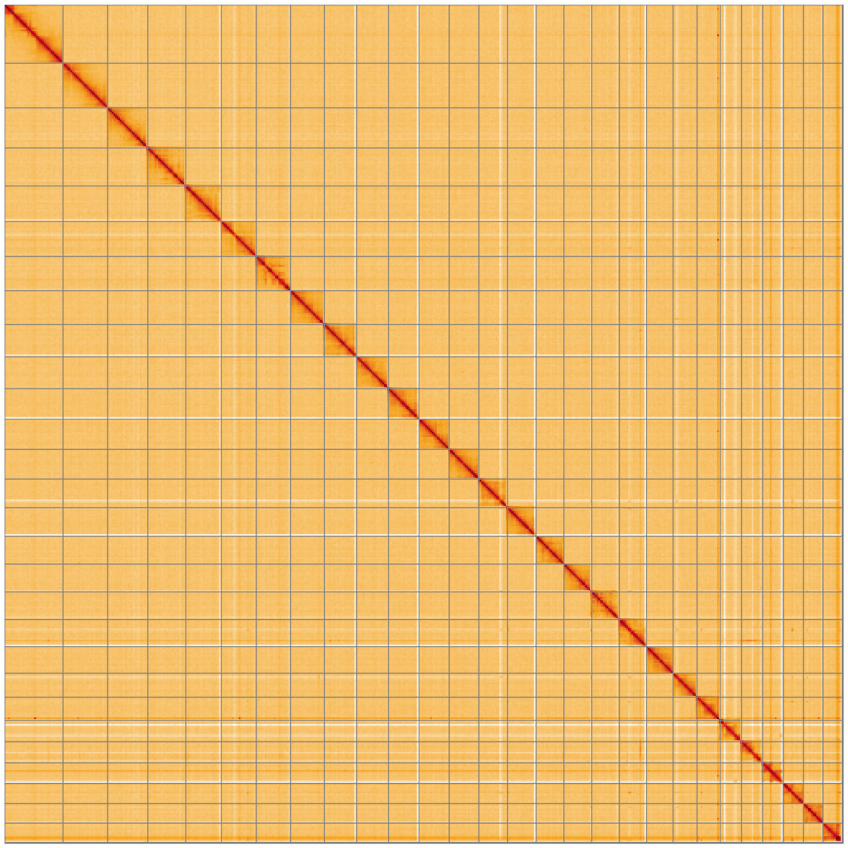

Figure 5. Genome assembly of *Pammene aurita*, ilPamAuri1.1: Hi-C contact map. Hi-C contact map of the ilPamAuri1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=XW4dzhIyThitRIyDG5Q5JQ

Table 2. Chromosomal pseudomolecules in the genome assembly of *Pammene aurita*, ilPamAuri1.

INSDC accession	Chromosome	Size (Mb)	GC%
OX352290.1	1	55.38	37.2
OX352291.1	2	49.73	37.3
OX352292.1	3	47.15	37.4
OX352296.1	4	41.96	37.4
OX352293.1	5	44.29	37.5
OX352299.1	6	37.83	37.4
OX352297.1	7	40.1	37.4
OX352298.1	8	39.56	37.3
OX352294.1	9	43.25	37.5
OX352300.1	10	37.5	37.5
OX352314.1	11	25.18	37.2
OX352301.1	12	36.88	37.3
OX352303.1	13	35.41	37.5
OX352295.1	14	42.62	37.6
OX352304.1	15	34.59	37.3
OX352305.1	16	34.48	37.4
OX352306.1	17	34.37	37.5
OX352308.1	18	32.99	37.5
OX352302.1	19	35.48	37.4
OX352307.1	20	33.65	37.3
OX352309.1	21	29.8	37.3
OX352310.1	22	28.75	37.9
OX352315.1	23	24.24	37.6
OX352313.1	24	25.35	38.9
OX352312.1	25	26.23	38
OX352316.1	26	23.78	37.8
OX352311.1	27	26.6	39
OX352289.1	Z	72.71	37.2
OX352317.1	MT	0.02	18.3

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.0.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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.

Data availability

European Nucleotide Archive: *Pammene aurita* (sycamore piercer). Accession number PRJEB55029; https://identifiers.org/ena.embl/PRJEB55029. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Pammene aurita* 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 using available RNA-Seq data 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.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

Bradley JD, Tremewan WG, Smith A: **British Tortricoid Moths - Tortricidae: Olethreutinae.** The Ray Society. 1979.

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

Elliott B, *et al.*: **Tortricidae.** In: J.R. Langmaid, S.M. Palmer, and M.R. Young (eds) *A Field Guide to the Smaller Moths of Great Britain and Ireland.* The British Entomological and Natural History Society. 2018; 279.

GBIF Secretariat: **Pammene aurita Razowski, 1992, GBIF Backbone Taxonomy. Checklist dataset.** 2022; (Accessed: 9 March 2023).
Publisher Full Text

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

Hancock F, Bland KP, Razowski J: **The Moths and Butterflies of Great Britain and Ireland.** Leiden: BRILL, 2015; **5**(2).

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): 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

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

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–2.
PubMed Abstract | Publisher Full Text

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

Wellcome Sanger Institute: **The genome sequence of the Sycamore Piercer, *Pammene aurita* (Razowski, 1991).** European Nucleotide Archive. [dataset], accession number PRJEB55029, 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