DATA NOTE

REVISED The genome sequence of the acorn piercer, *Pammene fasciana* (Linnaeus, 1761) [version 2; peer review: 2 approved]

Douglas Boyes[1],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Thomas Lewin[2],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]University of Oxford, Oxford, Oxfordshire, UK

v2 **First published:** 13 Oct 2022, **7**:258
https://doi.org/10.12688/wellcomeopenres.18114.1
Latest published: 17 Oct 2024, **7**:258
https://doi.org/10.12688/wellcomeopenres.18114.2

Abstract

We present a genome assembly from an individual male *Pammene fasciana* (acorn piercer; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence is 564 megabases in span. Most of the assembly (99.94%) is scaffolded into 28 chromosomal pseudomolecules with the Z sex chromosome assembled. The complete mitochondrial genome was also assembled and is 16.4 kilobases in length. Gene annotation of this assembly on Ensembl identified 21,224 protein-coding genes.

Keywords

Pammene fasciana, acorn piercer, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 17 Oct 2024		✓ view
version 1 13 Oct 2022	✓ view	? view

1. **Chang-Bae Kim**, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea

2. **Ljiljana Šašić Zorić**, BioSense Institute, Dr Zorana Đinđića, Novi Sad, Serbia

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; **Lewin T**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) 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.

How to cite this article: Boyes D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the acorn piercer, *Pammene fasciana* (Linnaeus, 1761) [version 2; peer review: 2 approved]** Wellcome Open Research 2024, **7**:258 https://doi.org/10.12688/wellcomeopenres.18114.2

First published: 13 Oct 2022, **7**:258 https://doi.org/10.12688/wellcomeopenres.18114.1

REVISED **Amendments from Version 1**

The corrections requested by the reviewers have been implemented, including details of the longest scaffold length, the percentage of the sequence assembled onto chromosomes, the rationale for sequencing for the Darwin Tree of Life project, Figure 2 and the figure captions. We have included details of the collection of the sample used for Hi-C sequencing. In addition, version 2 of the data note includes details of genome annotation of the *Pammene fasciana* assembly.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Tortricoidea; Tortricidae; Olethreutinae; Grapholitini; *Pammene*; *Pammene fasciana* (Linnaeus, 1761) (NCBI: txid1101027).

Background

Pammene fasciana, commonly known as the acorn piercer or the chestnut leafroller, is a moth of the family Tortricidae. *P. fasciana* is univoltine, and is found on the wing in the UK from June to August (Bland *et al.,* 2015). Adults have a wingspan of 13 to 17 mm, with dark brown forewings, chocolate brown hindwings and a white or ochreous dorsal patch that forms a thick, curved fascia. It is distributed across Europe and into Russia, and mainly inhabits oak woodlands (GBIF, 2022; Bland *et al.,* 2015). In the British Isles, it is most common in southern England, Wales and the south of the Republic of Ireland, but is also found more sporadically in northern England, Scotland and Northern Ireland (Bland *et al.,* 2015).

Several tortricid species are among the most economically important pests in Europe and beyond (Bradley *et al.,* 1979; Kadoić Balaško *et al.,* 2020; Suckling & Brockerhoff, 2010; van der Geest & Evenhuis, 1991), and *P. fasciana* is a notable pest of oak (*Quercus* spp.) and sweet chestnut (*Castanea sativa*), especially in mainland Europe (Avtzis, 2012; Clausi *et al.,* 2016; Speranza, 1998). Eggs are laid on these species' leaves in the summer; carpophagous (fruit-feeding) larvae hatch approximately two weeks later, and burrow into acorns or chestnuts and consume the kernel, causing premature dropping of fruits, reduced produce quality and subsequently reduced yields and economic losses (Bland *et al.,* 2015; Pedrazzoli *et al.,* 2012; Speranza, 1998).

The genome of the acorn piercer, *Pammene fasciana*, 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.

Genome sequence report

The genome was sequenced from a single male *P. fasciana* collected from Wytham Woods, Berkshire, UK (Figure 1). A total of 36-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 69-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 11 missing/misjoins, reducing the scaffold number by 19.51%, and increasing the scaffold N50 by 1.19%.

Figure 1. Image of the *Pammene fasciana* specimen (ToLID ilPamFasc1) taken prior to preservation and processing.

The final assembly has a total length of 564 Mb in 33 sequence scaffolds with a scaffold N50 of 20.7 Mb (Table 1). Most of the assembly sequence (99.94%) was assigned to 28 chromosomal-level scaffolds, representing 27 autosomes (numbered by sequence length) and the Z sex chromosome (Figure 2–Figure 5; Table 2).

Genome annotation report

The Pammene fasciana genome assembly (GCA_911728535.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 21,492 transcribed mRNAs from 21,224 protein-coding genes (Table 1; https://rapid.ensembl.org/Pammene_fasciana_GCA_911728535.1/Info/Index). The average transcript length is 7,240.24, and there are 5.34 exons per transcript.

The assembly has a BUSCO v5.3.2 (Manni *et al.,* 2021) completeness of 98.2% (single 97.3%, duplicated 0.9%) using the lepidoptera_odb10 reference set (n=5,286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition and nucleic acid extraction

A single male *P. fasciana* specimen (ilPamFasc1) was collected using a light trap from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude –1.338) by Douglas Boyes (University of Oxford). Specimen ilPamFasc1 was used for DNA sequencing. The specimen used for Hi-C sequencing (ToLID ilPamFasc2) was collected on the same occasion. The specimens were identified by Douglas Boyes, based on morphological examination, and then snap-frozen on dry ice.

Table 1. Genome data for *Pammene fasciana*, ilPamFasc1.1.

Project accession data	
Assembly identifier	ilPamFasc1.1
Species	*Pammene fasciana*
Specimen	ilPamFasc1 (genome assembly); ilPamFasc2 (Hi-C)
NCBI taxonomy ID	1101027
BioProject	PRJEB45670
BioSample ID	SAMEA7701530
Isolate information	Male, whole organism (ilPamFasc1); whole organism (ilPamFasc2)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6939220
10X Genomics Illumina	ERR6363300-ERR6363303
Hi-C Illumina	ERR6363304
Genome assembly	
Assembly accession	GCA_911728535.1
Accession of alternate haplotype	GCA_911728525.2
Span (Mb)	564
Number of contigs	71
Contig N50 length (Mb)	16.01
Number of scaffolds	34 (including the mitochondrial sequence)
Scaffold N50 length (Mb)	20.7
Longest scaffold (Mb)	46.7
BUSCO* genome score	C:98.2%[S:97.3%,D:0.9%], F:0.5%,M:1.3%,n:5,286
Genome annotation of assembly GCA_911728535.1 at Ensembl	
Number of protein-coding genes	21,224
Number of gene transcripts	21,492

*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/ilPamFasc1.1/dataset/CAJVRX01/busco.

DNA was extracted at the Tree of Life Laboratory, Wellcome Sanger Institute. The ilPamFasc1 sample was weighed and dissected on dry ice. Whole organism tissue was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was

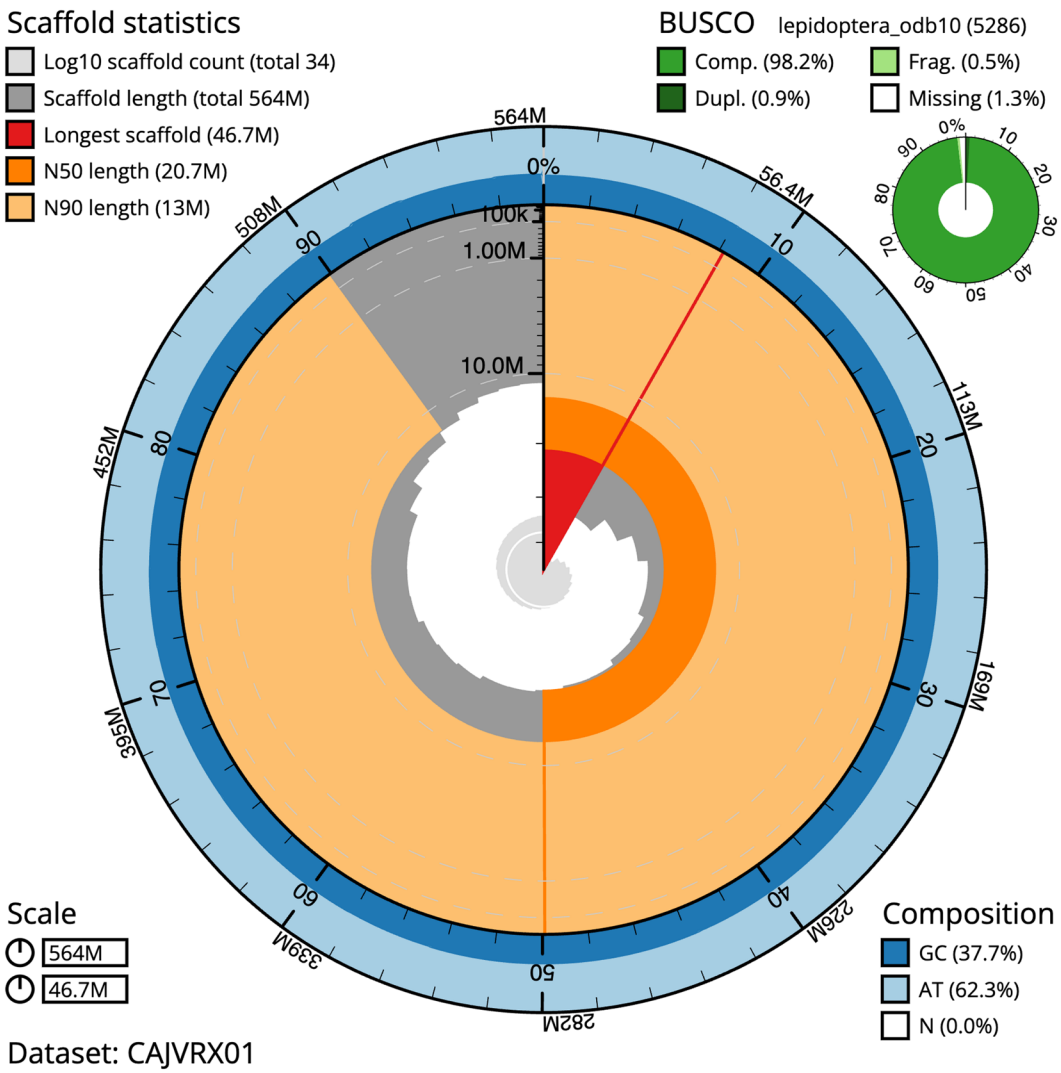

Figure 2. Genome assembly of *Pammene fasciana*, ilPamFasc1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 564,437,882 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (46,694,440 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (20,742,756 and 12,967,311 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/ilPamFasc1.1/dataset/CAJVRX01/snail.

performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were generated in the Tree of Life Laboratory from whole organism tissue of ilPamFasc2 using the Arima v2 kit and sequenced on a NovaSeq 6000 instrument.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021); haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes

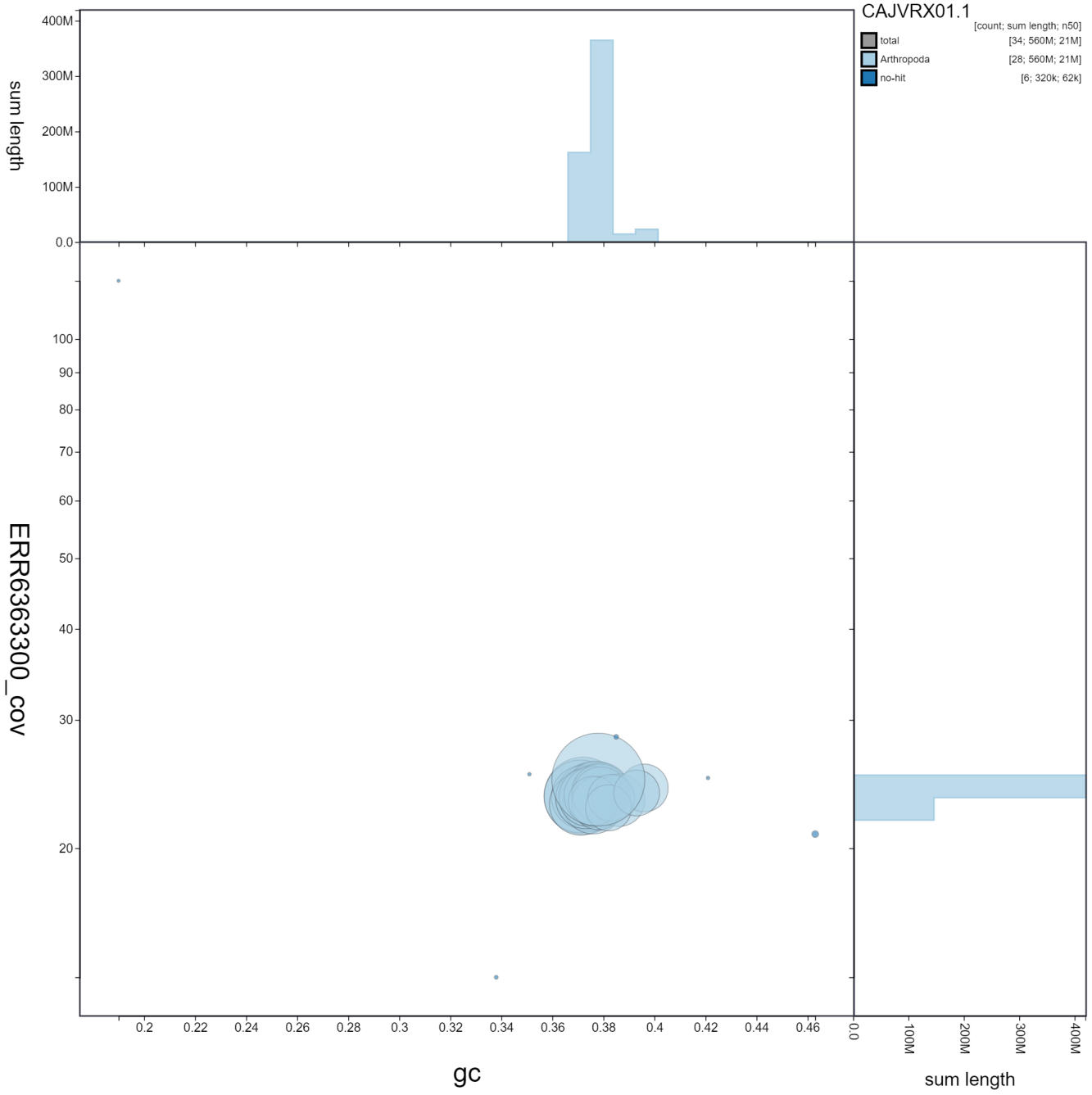

Figure 3. Genome assembly of *Pammene fasciana*, ilPamFasc1.1: BlobToolKit GC-coverage plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilPamFasc1.1/dataset/CAJVRX01.1/blob.

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

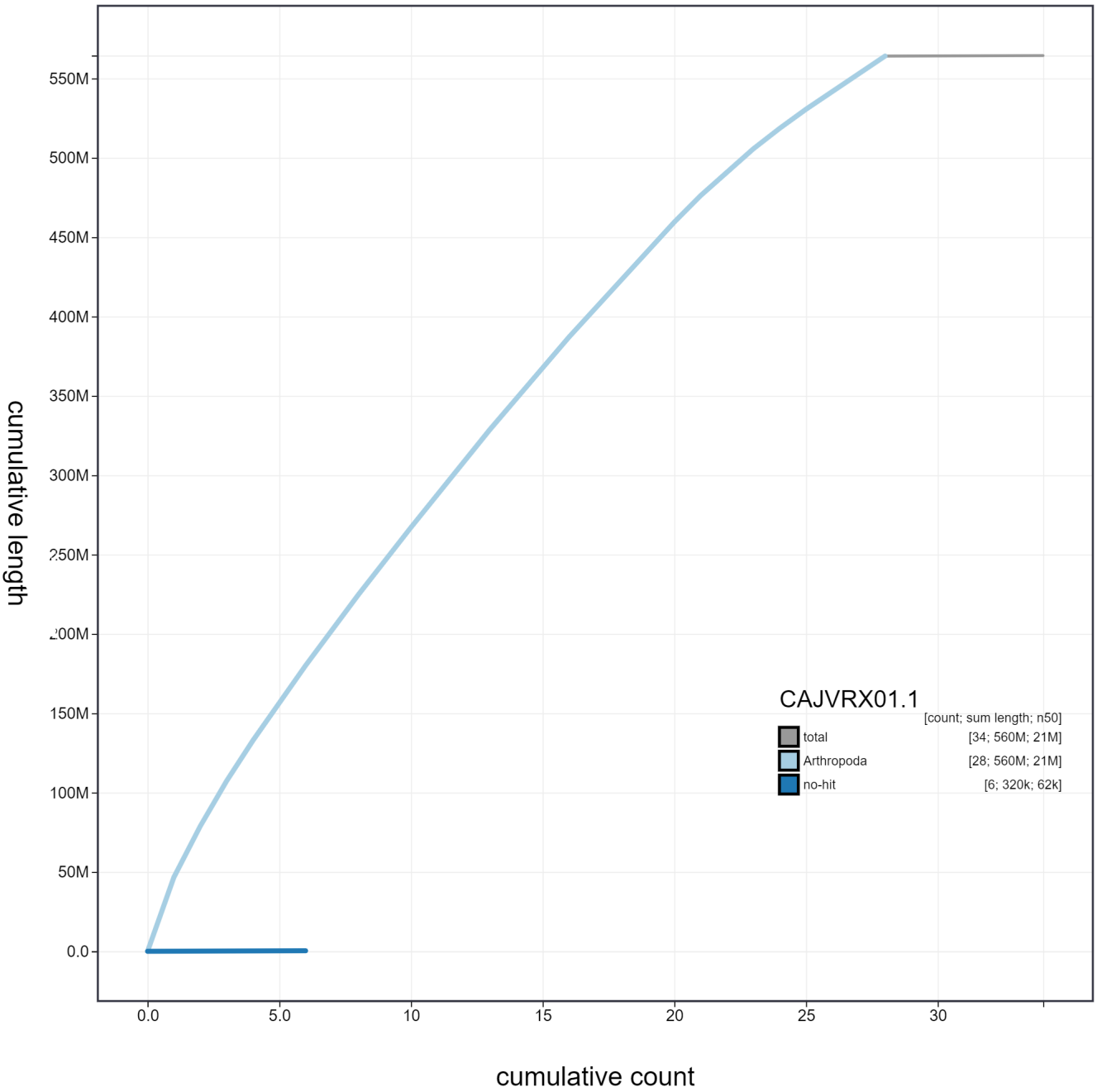

Figure 4. Genome assembly of *Pammene fasciana*, ilPamFasc1.1: cumulative sequence length. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilPamFasc1.1/dataset/CAJVRX01/cumulative.

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

Genome annotation

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

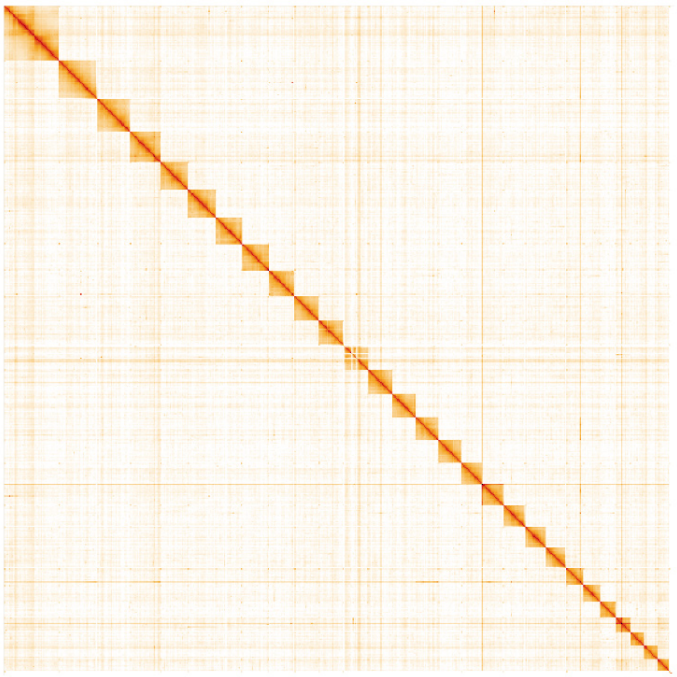

Figure 5. Genome assembly of *Pammene fasciana*, ilPamFasc1.1: Hi-C contact map. Hi-C contact map of the ilPamFasc1.1 assembly, visualised in HiGlass. Chromosomes are arranged in size order from left to right and top to bottom. The interactive Hi-C map can be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=djOZqsL5QBewlQtNZ2pkBw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Pammene fasciana*, ilPamFasc1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
OU452273.1	1	32.18	37.2
OU452274.1	2	28.49	37.1
OU452275.1	3	25.46	37.5
OU452276.1	4	23.63	37.6
OU452277.1	5	23.62	37.7
OU452278.1	6	22.45	37.6
OU452279.1	7	22.15	37.4
OU452280.1	8	21.27	37.2
OU452281.1	9	21.21	37.7
OU452282.1	10	20.74	37.6
OU452283.1	11	20.5	37.7
OU452284.1	12	20.46	37.1
OU452285.1	13	19.76	37.5
OU452286.1	14	19.38	37.6
OU452287.1	15	19.09	37.3
OU452288.1	16	18.26	37.4
OU452289.1	17	18.14	37.9
OU452290.1	18	18.09	37.6
OU452291.1	19	18.01	37.9
OU452292.1	20	16.62	37.9
OU452293.1	21	14.95	37.8
OU452294.1	22	14.52	38.6
OU452295.1	23	12.97	37.6
OU452296.1	24	12	39.6
OU452297.1	25	11.35	38.3
OU452298.1	26	11.13	38.2
OU452299.1	27	11.02	39.3
OU452272.1	Z	46.69	37.8
OU452300.1	MT	0.02	18.9
-	Unplaced	0.3	42.1

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.15.1	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
SALSA2	2.2	Ghurye *et al.*, 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	2.0	Uliano-Silva *et al.*, 2021
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
PretextView	0.2.x	https://github.com/wtsi-hpag/PretextView
BlobToolKit	3.2.6	Challis *et al.*, 2020

fasciana assembly (GCA_911728535.1) in Ensembl Rapid Release at the EBI.

Ethics/compliance issues

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the 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: Pammene fasciana (acorn piercer). Accession number PRJEB45670; https://identifiers.org/ena.embl/PRJEB45670.

The genome sequence is released openly for reuse. The *P. fasciana* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.6418202.

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

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.6866293.

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

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

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

References

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

Avtzis DN: **The distribution of *pammene fasciana* L. (lepidoptera: tortricidae) in greece: an underestimated chestnut-feeding pest.** *Int J Pest Manag.* 2012; **58**(2): 115–19.
Publisher Full Text

Bland KP, Razowski J, Hancock EF: **Tortricidae, part 2: olethreutinae.** BRILL, 2015; **5/2.**

Bradley JD, Tremewan WG, Smith A: **British tortricoid moths. tortricidae: olethreutinae.** In: *British tortricoid moths. tortricidae: olethreutinae.* The Ray Society, 1979.
Reference Source

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): 1361–74.
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

Clausi M, Leone D, Oreste M, *et al.*: **Comments on the morphology and biology of *Pammene castanicola* trematerra & clausi, 2009 (Lepidoptera: Tortricidae: Grapholitini).** *Zootaxa.* 2016; **4168**(3): 570–72.
PubMed Abstract | Publisher Full Text

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** arXiv: 1207.3907. 2012.
Publisher Full Text

GBIF: ***Pammene Fasciana* (Linnaeus, 1761).** 2022; Accessed February 25, 2022.
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**

Kadoić Balaško M, Bažok R, Mikac KM, *et al.*: **Pest management challenges and control practices in codling moth: a review.** *Insects.* 2020; **11**(1): 38.
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**

Pedrazzoli F, Salvadori C, de Cristofaro A, *et al.*: **A new strategy of environmentally safe control of chestnut tortricid moths.** *IOBC/WPRS Bulletin = Bulletin OILB/SROP.* 2012; **17**: 117–23.
Reference Source

Rao SS, 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**

Speranza S: **Chestnut pests in Central Italy.** *Acta Horticulturae.* 1998; **494**.
Publisher Full Text

Suckling DM, Brockerhoff EG: **Invasion biology, ecology, and management of the light brown apple moth (tortricidae).** *Annu Rev Entomol.* 2010; **55**: 285–306.
PubMed Abstract | **Publisher Full Text**

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

van der Geest LPS, Evenhuis HH: **World crop pests 5: tortricid pests: their biology, natural enemies and control.** In: *Eurasian species in forestry*, edited by H. Bogenschütz, Amsterdam; Oxford; New York: Elsevier, 1991; 673–709.
Reference Source

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 25 October 2024

https://doi.org/10.21956/wellcomeopenres.25631.r105759

✔ **Ljiljana Šašić Zorić**
BioSense Institute, Dr Zorana Đinđića, Novi Sad, Serbia

I have no further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics and molecular taxonomy

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.

Version 1

Reviewer Report 03 March 2023

https://doi.org/10.21956/wellcomeopenres.20085.r55006

? **Ljiljana Šašić Zorić**
BioSense Institute, Dr Zorana Đinđića, Novi Sad, Serbia

“The genome sequence of the acorn piercer, *Pammene fasciana* (Linnaeus, 1761)” is a technical report on the genome assembly of a moth species *Pammene fasciana*. The article is a significant contribution to further research on acorn piercer as an important pest species of oak and sweet chestnut. The sequencing approaches includes Pacific Biosciences HiFi and 10X Genomics

sequencing data, with scaffolding using Hi-C data. The applied methodology is suitable and clearly described, as well as technical aspects of genome assembly. However, there are minor inconsistencies that should be resolved prior to final publication.

Suggestions for revision:

Genome sequence report

- Please add ID (ilPamFasc1 or ilPamFasc2) for the specimen on Figure 1 since it differs from the ID shown in the figure.
- In the second paragraph, it is stated that 9.94% of the assembly sequence was assigned to 28 chromosomal-level scaffolds, while in the abstract it is 99.94%
- The number of scaffolds in Table 1 is 33 but, based on Figure 2 and results available at https://blobtoolkit.genomehubs.org/view/ilPamFasc1.1/dataset/CAJVRX01/snail, it seems that there are 34 scaffolds. The same applies for Figures 3 and 4. I suppose that one of these corresponds to mitochondrial genome, but please state that clearly.
- Figure 2 is a bit different compared to the interactive version on the link https://blobtoolkit.genomehubs.org/view/ilPamFasc1.1/dataset/CAJVRX01/snail (scale in the interactive version is set to 564 and 46.7 M, while on the figure it is 560M and 47M). The longest scaffold length is 47Mb in Figure 2, while it is 32.2 in Table 1. Please check this.
- Figures 3 and 4 need some technical corrections such as first letter capitalization in the names of axes and gc should also be capitalized.
- Figure 3 caption should be written more clearly: “Genome assembly of *Pammene fasciana*, ilPamFasc1.1: base coverage in ERR6363300 against GC proportion.”
- Please correct Figure 4 caption – “cumulative sequence length” instead of “cumulative sequence”

Methods

- Please add collection data for ilPamFasc2 specimen.
- There is no information on parameters used in data analysis. If the default values were used, please state that clearly or, in case the data analysis follows the procedure previously described in detail, please refer to the source.

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?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics and molecular taxonomy

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 15 November 2022

https://doi.org/10.21956/wellcomeopenres.20085.r52901

Chang-Bae Kim
National Genome Information Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea

The manuscript by Boyes and Lewin reports the genome sequence of the acorn piercer, *Pammene fasciana*. The genomic data is a welcome addition to the public databases and may provide raw data for future analyses. However, as a reader, I find some basic information missing from the manuscript.

- Please provide the purpose of this study in the last of the Background section.
- Please provide some details on the availability of genomic data from closely related data.
- It is not clear how the species was identified prior to sequencing.

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: Bioinformatics for environmental genome and transcriptome by using

machine learning

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.