DATA NOTE

REVISED The genome sequence of the Barred Straw, *Gandaritis pyraliata* (Denis & Schiffermüller, 1775)

[version 2; peer review: 1 approved, 3 approved with reservations]

Douglas Boyes[1+], Peter W.H. Holland[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

v2 **First published:** 10 Oct 2023, **8**:435
https://doi.org/10.12688/wellcomeopenres.19526.1
Latest published: 15 Oct 2025, **8**:435
https://doi.org/10.12688/wellcomeopenres.19526.2

Abstract

We present a genome assembly from an individual female *Gandaritis pyraliata* (the Barred Straw; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 295.6 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the W and Z sex chromosomes. The mitochondrial genome has also been assembled and is 15.74 kilobases in length. Gene annotation of this assembly on Ensembl identified 15,805 protein coding genes.

Keywords

Gandaritis pyraliata, the Barred Straw, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ? ? ?

	1	2	3	4
version 2 (revision) 15 Oct 2025	✓ view		? view	? view
version 1 10 Oct 2023	? view	? view		

1. **Alessandro Grapputo**, Universita degli Studi di Padova, Padua, Italy
2. **Fahad Alqahtani**, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia
3. **Yu-Feng Huang**, Yuan Ze University (Ringgold ID: 34895), Zhongli District, Taiwan
4. **Olli-Pekka Smolander**, University of Helsinki, Helsinki, Finland

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; **Holland PWH**: 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, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Barred Straw, *Gandaritis pyraliata* (Denis & Schiffermüller, 1775) [version 2; peer review: 1 approved, 3 approved with reservations]** Wellcome Open Research 2025, **8**:435 https://doi.org/10.12688/wellcomeopenres.19526.2

First published: 10 Oct 2023, **8**:435 https://doi.org/10.12688/wellcomeopenres.19526.1

REVISED **Amendments from Version 1**

In Version 2 of this data note, we added a citation to the first paragraph of the Background for the resting position of moths as an evolutionary adaption.

We have added more information to the assembly method to specify settings used. We also added a link to the Tree of Life production code in the Data Availability section.

A new version of the the Merqury.FK tool was used to calculate the consensus quality and *k*-mer completeness, and these values have been updated in the data note.

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

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Geometroidea; Geometridae; Larentiinae; *Gandaritis*; *Gandaritis pyraliata* (Denis & Schiffermüller, 1775) (NCBI:txid934938).

Background

The resting position of an insect can be as important as shape and colour for effective camouflage or mimicry (Kang *et al.*, 2012). Most moths of the family Geometridae rest with their wings flat against the substrate, often enabling wing colour and pattern to blend with surroundings. The Barred Straw *Gandaritis pyraliata* (synonym *Eulithis pyraliata*) is an exception. *G. pyraliata* invariably rests with its wings outstretched and held aloft at an angle above horizontal; the hindwings are also rotated forward under the forewings, with the trailing edge curled. Only one pair of wings is visible, therefore, stretched out and partially raised. The adaptive significance of this unusual resting position is unclear. One possibility is that when settled in herbaceous vegetation, rather than on tree trunks, the shape may appear less 'moth-like' to visual predators.

As the common name suggests, *Gandaritis pyraliata* is a pale straw-coloured moth with a jagged, black-edged band across each forewing. The species is widely distributed across Europe and Asia, and in Britain it is found as far north as Orkney (GBIF Secretariat, 2022; Randle *et al.*, 2019). The larvae feed primarily on *Galium aparine* (cleavers), a plant avoided by many insects due to the presence of deterrent chemicals including alkaloids, glycosides and phenolic compounds (Morimoto *et al.*, 2005). In southern Britain, the adult moth is on the wing from June to August, with eggs overwintering and larvae developing from April to June (South, 1961; Waring *et al.*, 2017).

The complete genome sequence of *Gandaritis pyraliata* was determined as part of the Darwin Tree of Life project. The assembled genome will facilitate research into the evolutionary arms race between insect and plant biochemistry, and contribute to the growing set of resources for studying lepidopteran ecology and evolution.

Genome sequence report

The genome was sequenced from one female *Gandaritis pyraliata* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.76, –1.34). A total of 31-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 17 missing joins or mis-joins and removed 3 haplotypic duplications, reducing the scaffold number by 25.93%.

The final assembly has a total length of 295.6 Mb in 39 sequence scaffolds with a scaffold N50 of 10.6 Mb (Table 1). Most (99.94%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the W and Z sex chromosomes. 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 mitochondrial genome was also assembled (length 15.74 kb, OX401881.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

The combined primary and alternate assemblies achieve an estimated QV of 66.9. The *k*-mer completeness is 80.37% for the primary assembly, 75.68% for the alternate haplotype, and 99.30% for the combined assemblies. The primary assembly has a BUSCO v5.3.2 completeness of 98.0% (single = 97.4%, duplicated = 0.5%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/934938.

Genome annotation report

The *Gandaritis pyraliata* genome assembly (GCA_947859175.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI) using BRAKER2. This annotation includes

Figure 1. Photograph of the *Gandaritis pyraliata* (ilGanPyra1) specimen used for genome sequencing.

Table 1. **Genome data for *Gandaritis pyraliata*, ilGanPyra1.1.**

Project accession data		
Assembly identifier	ilGanPyra1.1	
Species	*Gandaritis pyraliata*	
Specimen	ilGanPyra1	
NCBI taxonomy ID	934938	
BioProject	PRJEB56798	
BioSample ID	SAMEA10978764	
Isolate information	ilGanPyra1, head and thorax (DNA sequencing and Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	Primary: 65.5; alternate: 67.8; combined: 66.9	*≥ 40*
k-mer completeness	Primary: 80.37%; alternate: 75.68%; combined: 99.30%	*≥ 95%*
BUSCO**	C:98.0%[S:97.4%,D:0.5%], F:0.5%,M:1.5%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.94%	*≥ 95%*
Sex chromosomes	W and Z chromosomes	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10395976	
Hi-C Illumina	ERR10395982	
Genome assembly		
Assembly accession	GCA_947859175.1	
Accession of alternate haplotype	GCA_947859165.1	
Span (Mb)	295.6	
Number of contigs	79	
Contig N50 length (Mb)	8.3	
Number of scaffolds	39	
Scaffold N50 length (Mb)	10.6	
Longest scaffold (Mb)	13.3	
Genome annotation		
Number of protein-coding genes	15,805	
Number of gene transcripts	15,998	

* 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/ilGanPyra1.1/dataset/CANUEG01/busco.

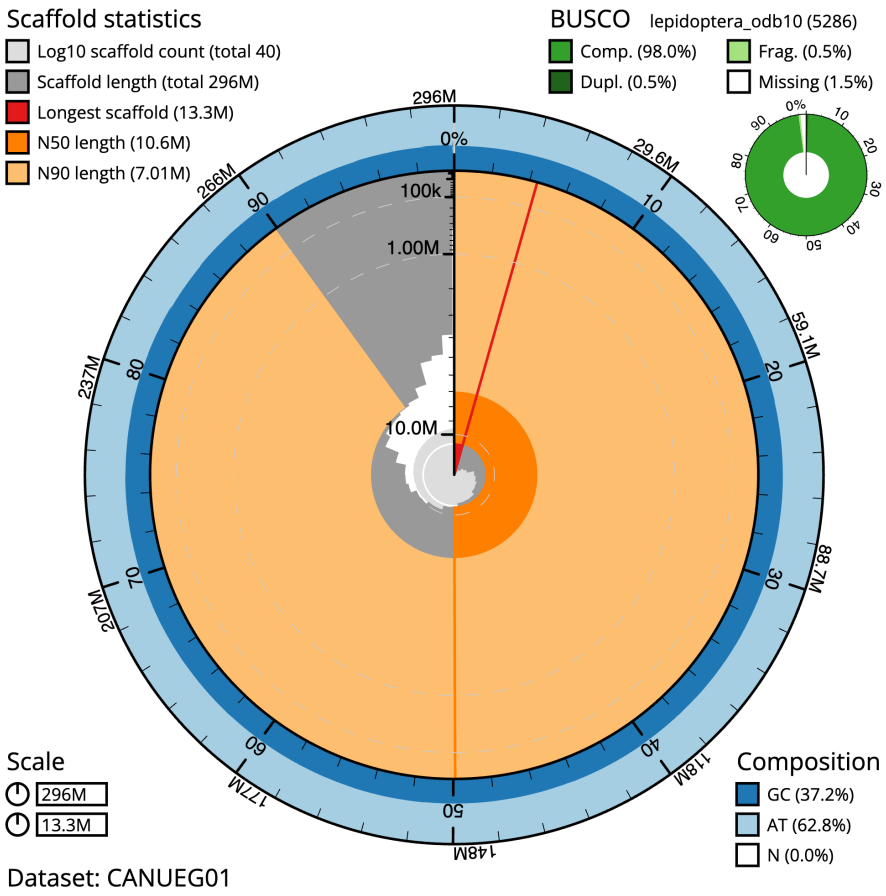

Figure 2. Genome assembly of *Gandaritis pyraliata*, ilGanPyra1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 295,653,368 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (13,285,173 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (10,572,331 and 7,009,674 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/ilGanPyra1.1/dataset/CANUEG01/snail.

15,998 transcribed mRNAs from 15,805 protein-coding genes. The average transcript length is 6 208.52 bp, with an average of 6.17 exons per transcript. For further information about the annotation, please refer to the annotation page on Ensembl.

Methods

Sample acquisition and nucleic acid extraction

The specimen selected for genome sequencing was a female *Gandaritis pyraliata* (specimen number Ox001597, ilGanPyra1) collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.76, longitude –1.34) on 2021-06-30. The specimen was taken from woodland habitat by Douglas Boyes (University of Oxford) using a light trap. The specimen was formally identified by the collector and snap-frozen on dry ice.

The ilGanPyra1 sample was prepared for DNA sequencing at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The tissue 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. DNA was extracted at the Wellcome Sanger Institute (WSI) Scientific Operations core using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions.

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

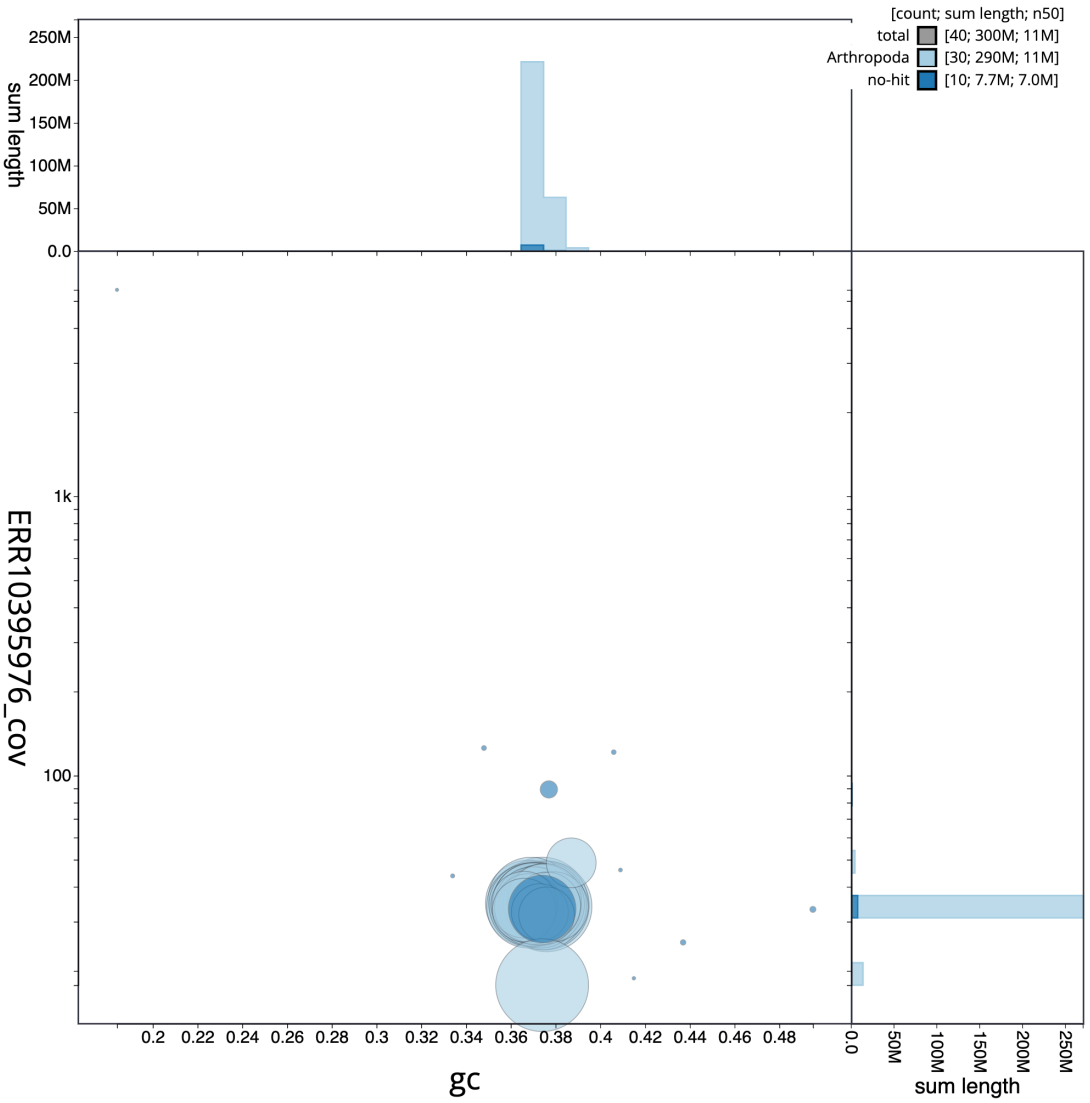

Figure 3. Genome assembly of *Gandaritis pyraliata*, ilGanPyra1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilGanPyra1.1/dataset/CANUEG01/blob.

(HiFi) instrument. Hi-C data were also generated from head and thorax tissue of ilGanPyra1 that had been set aside, using the Arimav2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021) were calculated.

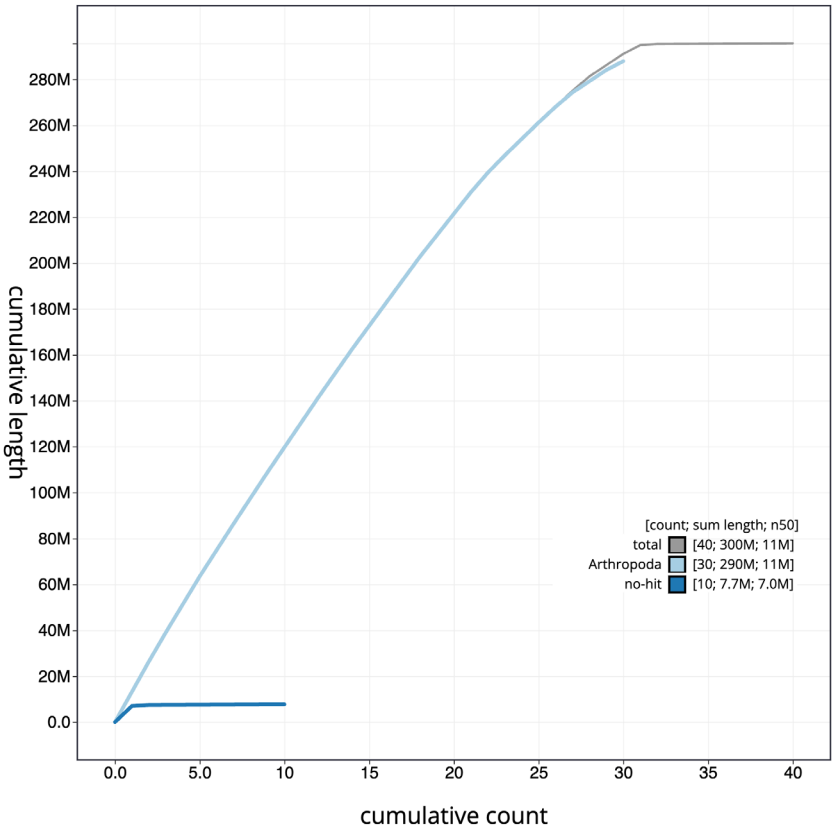

Figure 4. Genome assembly of *Gandaritis pyraliata*, ilGanPyra1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilGanPyra1.1/dataset/CANUEG01/cumulative.

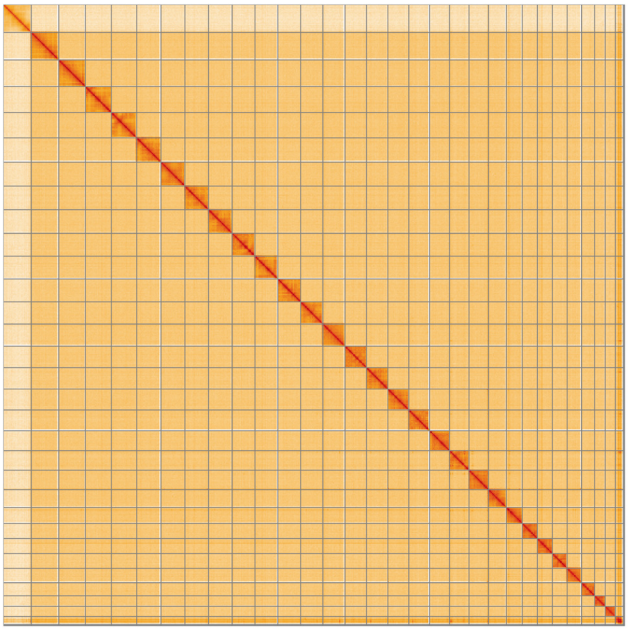

Figure 5. Genome assembly of *Gandaritis pyraliata*, ilGanPyra1.1: Hi-C contact map of the ilGanPyra1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=DNbo17mrReWbHG50fYszFg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Gandaritis pyraliata*, ilGanPyra1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX401851.1	1	13.1	37.5
OX401852.1	2	12.66	37.5
OX401853.1	3	12.31	37.5
OX401854.1	4	12.12	37.5
OX401855.1	5	11.53	37.0
OX401856.1	6	11.4	37.0
OX401857.1	7	11.25	37.0
OX401858.1	8	11.18	37.0
OX401859.1	9	10.91	37.0
OX401860.1	10	10.88	37.5
OX401861.1	11	10.85	37.0
OX401862.1	12	10.57	37.0
OX401863.1	13	10.53	37.5
OX401864.1	14	10.21	37.0
OX401865.1	15	10.18	37.0
OX401866.1	16	9.94	37.5
OX401867.1	17	9.89	37.5
OX401868.1	18	9.46	37.0
OX401869.1	19	9.3	37.5
OX401870.1	20	9.27	37.0
OX401871.1	21	8.53	37.5
OX401872.1	22	7.53	38.0
OX401873.1	23	7.19	37.0
OX401874.1	24	7.15	37.5
OX401875.1	25	7.01	37.5
OX401876.1	26	6.86	36.5
OX401877.1	27	6.26	36.5
OX401878.1	28	5.01	37.5
OX401879.1	29	4.78	37.5
OX401880.1	W	3.82	38.5
OX401850.1	Z	13.29	37.5
OX401881.1	MT	0.02	18.5

Table 3 contains a list of relevant software tool versions and sources.

Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Gandaritis pyraliata* assembly (GCA_947859175.1) in Ensembl Rapid Release.

Wellcome Sanger Institute - Legal and Governance

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'**, which can be found in full on the Darwin Tree of Life website here. 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.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.1.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury.FK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2.5	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

nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

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: *Gandaritis pyraliata* (barred straw). Accession number PRJEB56798; https://identifiers.org/ena.embl/PRJEB56798.

The genome sequence is released openly for reuse. The *Gandaritis pyraliata* 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.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol.

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

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

GBIF Secretariat: ***Gandaritis pyraliata* (Denis &Schiffermüller, 1775).** *GBIF Backbone Taxonomy.* 2022; (Accessed: 1 March 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

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

Kang CK, Moon JY, Lee SI, *et al.*: **Camouflage through an active choice of a resting spot and body orientation in moths.** *J Evol Biol.* 2012; **25**(9): 1695–1702.
PubMed Abstract | Publisher 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

Morimoto M, Tanimoto K, Komai K: **An insect antifeedant in catch weed (*Galium aparine* L.): a strategy for damage avoidance.** In: *Proceedings of the 4th World Congress on Allelopathy.* Centre for Rural Social Research. Charles Sturt University: Australia, 2005; 400–402. (Accessed: 19 May 2023).
Reference Source

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**

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**

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

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

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
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

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 2

Reviewer Report 28 January 2026

https://doi.org/10.21956/wellcomeopenres.27604.r145061

?

Olli-Pekka Smolander
University of Helsinki, Helsinki, Finland

The authors describe chromosomal-level genome assembly of the Barred Straw moth, *Gandaritis pyraliata*. Pacific Biosciences HiFi long-read sequencing and Illumina sequenced Hi-C data are used for contig assembly and scaffolding, respectively. The authors report that the assembly spans 295.6 Mb, with 99.94% of the sequence localized to 31 chromosomal pseudomolecules, including 29 autosomes and the W and Z sex chromosomes. Annotation was conducted using the Ensembl Rapid Release pipeline (with BRAKER2 in protein mode), identifying 15,805 protein-coding genes. BUSCO analysis against the lepidoptera_odb10 set shows high completeness at 98.0%. The result is a solid genomic resource for investigating the evolutionary biochemistry of this species and its specialised interactions with chemically unfavourable host plants.

The rationale and context (ecology and evolutionary adaptation of the species) for creating this dataset is clearly described. Furthermore, the datasets are presented in a usable and accessible format. The raw data and assembly have been deposited in appropriate international databases, and the authors provide accession numbers to ensure public access.

However, the protocols and technical soundness are only partially sufficient. While there are no questions about the quality of the assembly and the used methods, the lack of transcriptomic evidence is a significant factor affecting the annotation quality. The use BRAKER2 in default protein mode without species-specific RNA-Seq data means that the structural accuracy of gene models, particularly intron/exon boundaries, is uncertain. Furthermore, the methodological details for separating the W and Z sex chromosomes from autosomes is not described at an adequate level. To achieve complete technical clarity, it should be described if this identification was based for example on coverage analysis, homology to other *Geometridae* genomes, or specific Hi-C features. Additionally, it would be good to list all the software tools (e.g. BRAKER2 and Cooler) that have been part of this work. Similarly, for each tool, it should be stated whether default parameters were used and if not, how those were modified.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?
Partly

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: Genomics, computational biology, bioinformatics, microbiome, ML

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 16 January 2026

https://doi.org/10.21956/wellcomeopenres.27604.r144592

?

Yu-Feng Huang
Computer Science and Engineering, Yuan Ze University (Ringgold ID: 34895), Zhongli District, Taoyuan City, Taiwan

This revised Data Note reports a high-quality reference genome assembly for *Gandaritis pyraliata*, generated using PacBio HiFi and Illumina Hi-C sequencing data. The assembly metrics are robust, achieving a combined Consensus Quality (QV) of 66.9 and a Scaffold N50 of 10.6 Mb. The authors successfully assigned 99.94% of the sequence to 31 chromosomal-level scaffolds, a result that meets and exceeds Earth BioGenome Project standards. However, although the manuscript has been revised to address prior peer review comments, inconsistencies regarding scaffold counts persist between the text and figures. Furthermore, specific biological validations namely gene content verification, mitogenome circularity, and sex chromosome identification - require additional clarification to ensure the data's reliability for downstream research.

Major Comments

1. Gene Annotation Validation The manuscript now includes a genome annotation generated via the BRAKER2 pipeline in "default protein mode," which identified 15,805 protein-coding genes. This annotation appears to be derived solely from protein homology, as there is no mention of species-specific RNA-Seq data generation in the Sample Acquisition or Methods sections. Consequently, the structural accuracy of these gene models, particularly intron/exon boundaries,

remains unverified against transcriptomic evidence. Please explicitly state the limitations of this annotation method within the text. If feasible, please perform a spot-check of a well-conserved gene family to confirm that the protein-mode prediction has yielded biologically valid gene structures; alternatively, clarify that this dataset serves as a preliminary set of gene models pending transcriptomic validation.
2. Sex Chromosome Identification Methodology The assembly assigns scaffolds to "29 autosomes and the W and Z sex chromosomes". However, the manuscript lacks a methodological description explaining how the W and Z chromosomes were identified and distinguished from the autosomes. Please specify the evidentiary basis for this assignment. For instance, was the identification based on coverage analysis (e.g., observing half-coverage for the Z/W chromosomes in the female specimen), homology to other *Geometridae* genomes, or specific Hi-C contact patterns?
Minor Comments
3. Mitogenome Circularization Verification The mitochondrial genome was assembled using MitoHiFi. The mitochondrial genome size reported in the text differs from the value listed in the Table 2. While the text notes that MitoHiFi ensures "general quality," it does not explicitly confirm that the assembly was verified as circular. Please confirm whether the circular topology was validated (e.g., by checking for overlapping ends) and if the sequence was rotated to a standard start feature to ensure it represents the complete organelle genome.
4. Software Tools (Table 3) Table 3 lists the software versions utilized in this study. However, discrepancies exist between the methods text and the table. Specifically, the BRAKER2 pipeline used for annotation is not listed. Additionally, Cooler, which is mentioned in the methods regarding Hi-C map production, is absent from the table. Please audit Table 3 and update it to include these missing tools along with their respective version numbers.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

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, Cancer Bioinformatics, Mitogenomics, Genome sequencing, assembly and gene annotation.

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 25 October 2025

https://doi.org/10.21956/wellcomeopenres.27604.r136391

Alessandro Grapputo
Universita degli Studi di Padova, Padua, Veneto, Italy

The revised paper by Boyes and collaborator have addressed appropriately the concerns I have raised in the previous review report. The paper can be indexed in the present form.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics and 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.

Version 1

Reviewer Report 28 May 2024

https://doi.org/10.21956/wellcomeopenres.21629.r83549

Fahad Alqahtani
[1] King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia
[2] King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

The authors of "The genome sequence of the Barred Straw, Gandaritis pyraliata (Denis & Schiffermüller, 1775)," have achieved a highly complete genome assembly at the chromosome level for a female Gandaritis pyraliata. They reconstructed the genome assembly using two sequencing technologies: Pacific Biosciences SEQUEL II and Hi-C Illumina. BUSCO analysis was used to assess the completeness of the genome assembly, indicating a genome size of 295.6 megabases, with 98% of common genes in the lepidoptera_odb10 present.

There are a few minor comments that should be addressed:

- The number of contigs in Table 1 is 80, while it is reported in NCBI as 79.

- The number of scaffolds in Table 1 is 40, while it is reported in NCBI as 39.
- 'Pretext' in the Genome Assembly section should be corrected to 'PretextView' to match the name in Table 3.

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

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 16 May 2024

https://doi.org/10.21956/wellcomeopenres.21629.r80784

?

Alessandro Grapputo
[1] Universita degli Studi di Padova, Padua, Veneto, Italy
[2] Universita degli Studi di Padova, Padua, Veneto, Italy

The paper by Boyes and colleagues present a new genome in the tree of life project, the genome sequence and annotation of the Barred Straw, *Gandaritis pyraliata*, a lepidoptera of the family Geometridae. The Barred Straw (synonym *Eulithis pyraliata*), as the other close relatives *Eulithis* species, *E. mellinata* and *E. populate*, usually rests in a very distinctive and characteristic way with the forewings held out at 90° to the body and with the hindwings hidden behind them.

The methods for sequencing acquisition, the genome assembly pipeline and annotation with the Ensembl rapid annotation pipeline follow the Classique scheme of most Data Note papers published by Wellcome Open Research. The methods are sound and good and I do not have comments about that. However the software listed have many optional parameters and none

have been reported. Therefore to the question: Are sufficient details of methods and materials provided to allow replication by others? I would say: partly. If the default parameters have been used this should be stated for each software.

The paper is well written, however the first paragraph of the background would benefit, first of at least a reference for all the information about the species reported, and second from adding the comparison of the characteristic way of resting with other members of the genus (possibly it had evolved only once and the phylogenomic analysis of the group may shed some light on it).

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 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, however I have significant reservations, as outlined above.