



DATA NOTE

The genome sequence of the Lunar-spotted Pinion, *Cosmia pyralina* (Denis & Schiffermüller, 1775) [version 1; peer review: 2 approved]

Douglas Boyes^{1†}, Inez Januszczak^{id}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

¹UK Centre for Ecology & Hydrology, Wallingford, England, UK

²Natural History Museum, London, England, UK

[†] Deceased author

V1 First published: 23 Nov 2023, 8:545
<https://doi.org/10.12688/wellcomeopenres.20148.1>
Latest published: 23 Nov 2023, 8:545
<https://doi.org/10.12688/wellcomeopenres.20148.1>

Abstract

We present a genome assembly from an individual male *Cosmia pyralina* (the Lunar-spotted Pinion; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 803.3 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.39 kilobases in length. Gene annotation of this assembly on Ensembl identified 19,901 protein coding genes.

Keywords

Cosmia pyralina, Lunar-spotted Pinion, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status

	1	2
version 1 23 Nov 2023	 view	 view

- Ningning Wu** , China Agricultural University Education Foundation, Beijing, China
China Agricultural University College of Plant Protection, Beijing, China
- Shaolong Qiu**, China Agricultural University College of Plant Protection, Beijing, China
- Jose Martinez** , University of Florida, Florida, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: **Boyes D:** Investigation, Resources; **Januszczak I:** 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.

Copyright: © 2023 Boyes D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Boyes D, Januszczak I, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Lunar-spotted Pinion, *Cosmia pyralina* (Denis & Schiffmüller, 1775) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:545 <https://doi.org/10.12688/wellcomeopenres.20148.1>

First published: 23 Nov 2023, 8:545 <https://doi.org/10.12688/wellcomeopenres.20148.1>

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; Noctuoidea; Noctuidae; Ipimorphinae; *Cosmia*; *Cosmia pyralina* (Denis & Schiffermüller, 1775) (NCBI:txid987909).

Background

The Lunar-spotted Pinion (*Cosmia pyralina*) is a noctuid moth, commonly found in central Europe, but its territory extends all the way through the Palaearctic to Korea and Japan (Fauna Europea, 2023). In Britain this is a species of central and southern England and (albeit more scarcely) Wales, particularly the south-east.

With a wingspan of 28–32 mm, the Lunar-spotted Pinion could be confused with the Lesser-spotted Pinion, *Cosmia affinis* (Linnaeus, 1767), but has paler hindwings and broader forewings (Waring *et al.*, 2017). *Cosmia pyralina* has a variety of colour morphs, with dull or bright red-brown forewings, with a contrasting white edge in the outer line, that eventually joins the white streak before the submarginal line. Similarly, larvae are pale green with white lines. The larvae are polyphagous and feed on deciduous trees, such as blackthorn (*Prunus spinosa*) and hawthorn (*Crataegus monogyna*) as well as English elm (*Ulmus procera*) and wych elm (*U. glabra*) (Waring *et al.*, 2017).

The species is declining in Britain (Conrad *et al.*, 2006), most likely due to Dutch elm disease. Caused by the fungus *Ophiostoma novo-ulmi*, Dutch elm disease is one of the most severe tree diseases in the world, causing foliage and tip dieback in all of Britain's major native elms (Potter *et al.*, 2011). Ten beetle species from the genus *Scotylus* (a genus of bark beetle) are known to feed on elms and a key insect vector for the disease to spread. The beetles frequently visit or inflict wounds in healthy elms; thus, depositing any spores directly into the damaged tree (Peacock *et al.*, 1981). Other *Cosmia* species which have suffered due to this disease include the White-spotted Pinion, *Cosmia diffinis*, which is now listed as 'Nationally Scarce' (Fox *et al.*, 2019). The fourth British *Cosmia* species, *C. trapezina* (the Dun-bar), still has a healthy population despite the described challenges, perhaps due to its trait of eating the larvae of other moth species as well as cannibalising its own (Turčáni & Patočka, 2011; Ward, 1903).

We present a chromosomally complete genome sequence for the Lunar-spotted Pinion (*Cosmia pyralina*) as part of the Darwin Tree of Life Project. This project is a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago, encompassing Britain and Ireland.

Genome sequence report

The genome was sequenced from one male *Cosmia pyralina* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 22-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary



Figure 1. Photograph of the *Cosmia pyralina* (ilCosPyra2) specimen used for genome sequencing.

assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 26 missing joins or mis-joins and removed 1 haplotypic duplication, reducing the scaffold number by 18.01% and increasing the scaffold N50 by 1.63%.

The final assembly has a total length of 803.3 Mb in 50 sequence scaffolds with a scaffold N50 of 27.4 Mb (Table 1). Most (99.89%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. A summary of the assembly statistics is shown in Figure 2, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). Comparators *Cosmia trapezina* (GCA_905163495.2) (Boyes *et al.*, 2022) and *Amphipoea oculea* (GCA_945859645.1) (Boyes *et al.*, 2023) were used for Z chromosome identification. 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 and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 61.5 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 99.0% (single = 98.1%, duplicated = 0.9%), 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/987909>.

Genome annotation report

The *Cosmia pyralina* genome assembly (GCA_946251885.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Cosmia_pyralina_GCA_946251885.1/Info/Index). The resulting annotation includes 20,077 transcribed mRNAs from 19,901 protein-coding genes.

Table 1. Genome data for *Cosmia pyralina*, ilCosPyra2.1.

Project accession data		
Assembly identifier	ilCosPyra2.1	
Species	<i>Cosmia pyralina</i>	
Specimen	ilCosPyra2	
NCBI taxonomy ID	987909	
BioProject	PRJEB54090	
BioSample ID	SAMEA10979084	
Isolate information	ilCosPyra2, male: head and thorax (DNA sequencing and Hi-C scaffolding), abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	61.5	≥ 50
k-mer completeness	100%	≥ 95%
BUSCO**	C:99.0%[S:98.1%,D:0.9%],F:0.2%,M:0.8%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.89%	≥ 95%
Sex chromosomes	Z chromosomes	<i>localised homologous pairs</i>
Organelles	Mitochondrial genome assembled	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR9924618	
Hi-C Illumina	ERR9930693	
PolyA RNA-Seq Illumina	ERR10890695	
Genome assembly		
Assembly accession	GCA_946251885.1	
Accession of alternate haplotype	GCA_946251865.1	
Span (Mb)	803.3	
Number of contigs	185	
Contig N50 length (Mb)	9.4	
Number of scaffolds	50	
Scaffold N50 length (Mb)	27.4	
Longest scaffold (Mb)	39.3	
Genome annotation		
Number of protein-coding genes	19,901	
Number of gene transcripts	20,077	

* 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/ilCosPyra2.1/dataset/CAMIUC01/busco>.

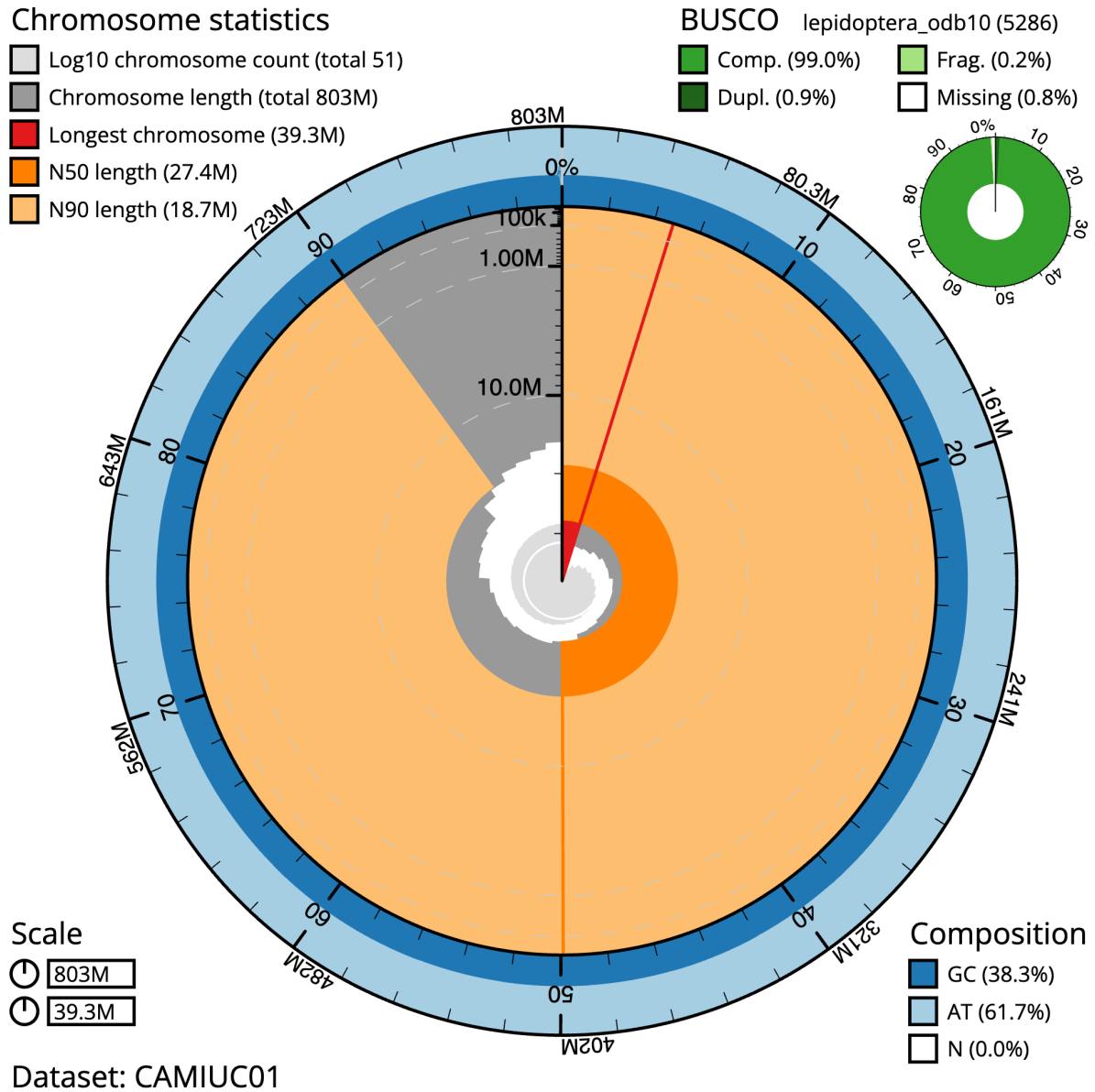


Figure 2. Genome assembly of *Cosmia pyralina*, ilCosPyra2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 803,276,861 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (39,281,260 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (27,366,473 and 18,734,863 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/ilCosPyra2.1/dataset/CAMIUC01/snail>.

Methods

Sample acquisition and nucleic acid extraction

A male *Cosmia pyralina* (specimen ID Ox001826, individual ilCosPyra2) was collected in a light trap from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude -1.34) on 2021-07-24. The specimen

was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilCosPyra2 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing.

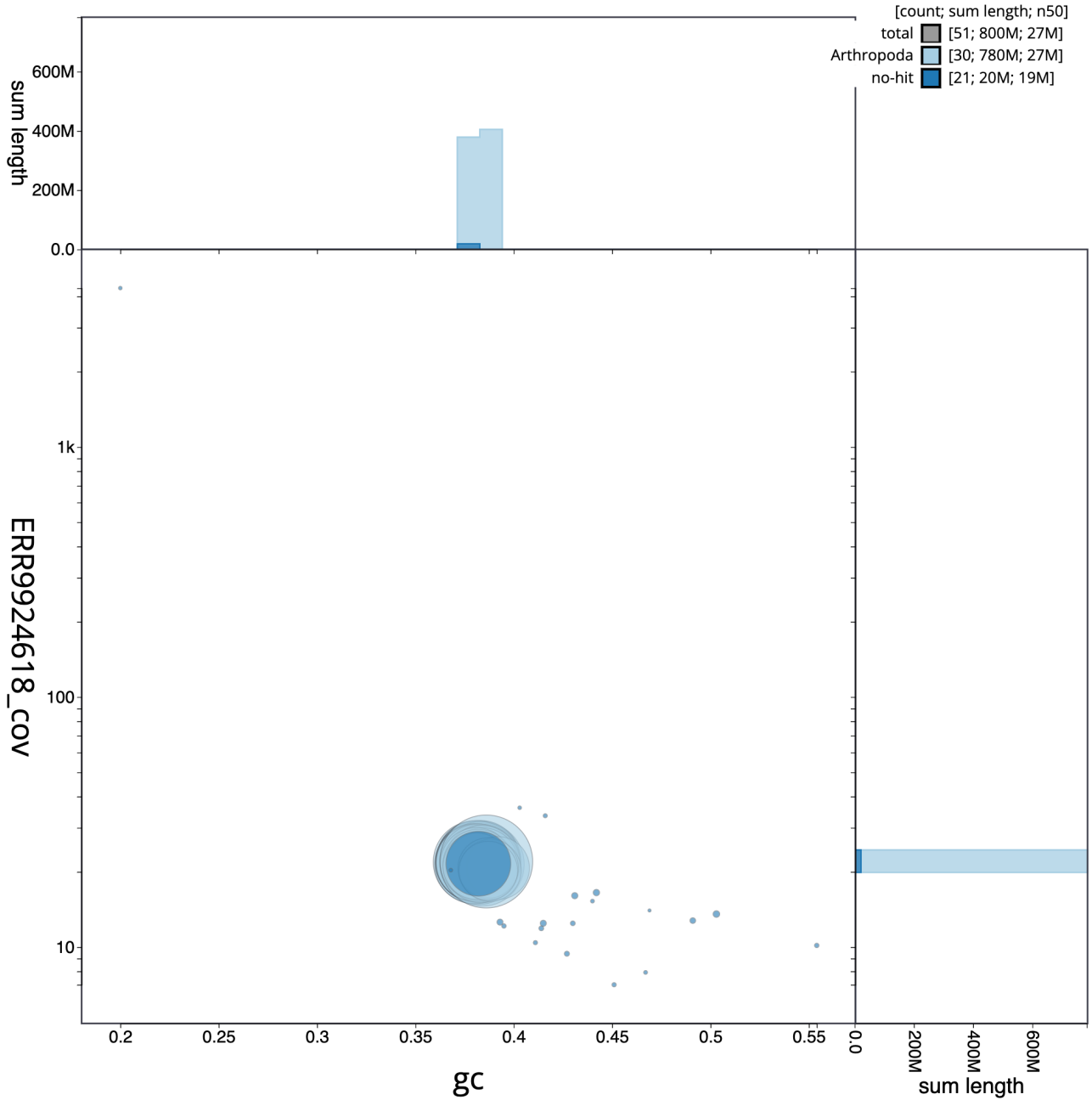


Figure 3. Genome assembly of *Cosmia pyralina*, iLCosPyra2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/iLCosPyra2.1/dataset/CAMIUC01/blob>.

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

RNA was extracted from abdomen tissue of iLCosPyra2 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer’s instructions. RNA was then eluted in

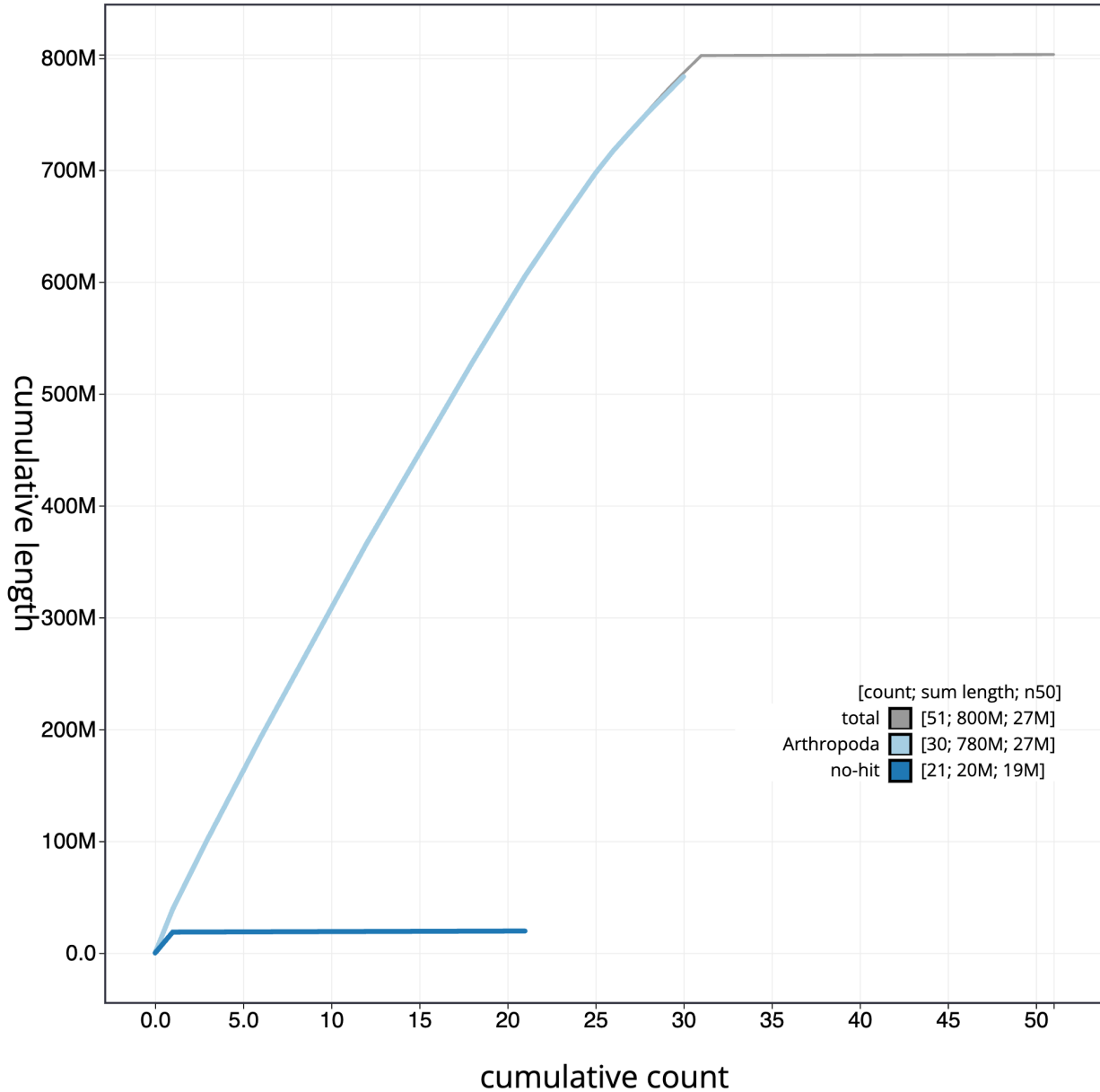


Figure 4. Genome assembly of *Cosmia pyralina*, ilCosPyra2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/ilCosPyra2.1/dataset/CAMIUC01/cumulative>.

50 µl RNase-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at

the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from remaining head and thorax tissue of ilCosPyra2 using the Arima2 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 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 and

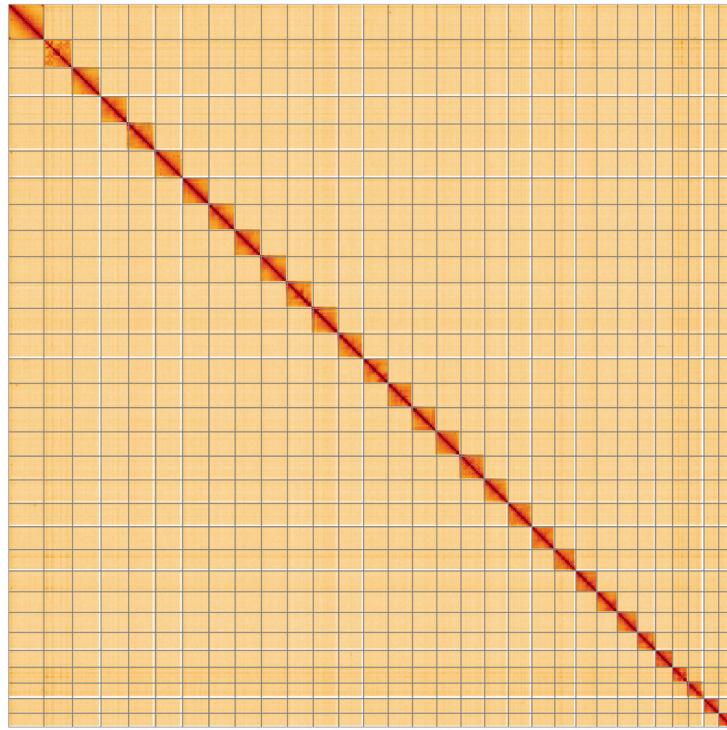


Figure 5. Genome assembly of *Cosmia pyralina*, ilCosPyra2.1: Hi-C contact map of the ilCosPyra2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/l/?d=TpGzUdJFRHyRe3gFUR058w>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Cosmia pyralina*, ilCosPyra2.

INSDC accession	Chromosome	Length (Mb)	GC%
OX276342.1	1	31.76	38.5
OX276343.1	2	31.43	38.0
OX276344.1	3	30.65	38.0
OX276345.1	4	30.01	38.5
OX276346.1	5	29.85	38.0
OX276347.1	6	29.17	38.5
OX276348.1	7	29.05	38.0
OX276349.1	8	29.02	38.0
OX276350.1	9	28.94	38.5
OX276351.1	10	28.51	38.0
OX276352.1	11	28.33	38.5
OX276353.1	12	27.47	38.0
OX276354.1	13	27.37	38.0
OX276355.1	14	26.93	38.0
OX276356.1	15	26.92	38.5

INSDC accession	Chromosome	Length (Mb)	GC%
OX276357.1	16	26.81	38.5
OX276358.1	17	26.4	38.0
OX276359.1	18	26.1	38.0
OX276360.1	19	25.94	38.5
OX276361.1	20	25.37	38.5
OX276362.1	21	23.59	38.5
OX276363.1	22	23.24	38.0
OX276364.1	23	22.82	38.0
OX276365.1	24	22.36	38.5
OX276366.1	25	19.85	38.0
OX276367.1	26	18.73	38.0
OX276368.1	27	17.59	39.0
OX276369.1	28	17.15	39.0
OX276370.1	29	16.28	38.5
OX276371.1	30	15.44	38.5
OX276341.1	Z	39.28	38.5
OX276372.1	MT	0.02	19.5

corrected 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 runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) 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). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a) and “sanger-tol/genomenote” (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

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

Genome annotation

The BRAKER2 pipeline (Brûna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Cosmia pyralina* assembly (GCA_946251885.1) 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 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.

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
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

Data availability

European Nucleotide Archive: *Cosmia pyralina*. Accession number PRJEB54090; <https://identifiers.org/ena.embl/PRJEB54090>. (Wellcome Sanger Institute, 2022) The genome sequence is released openly for reuse. The *Cosmia pyralina* 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

- 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](#)
- Bernt M, Donath A, Jühling F, et al.: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol*. 2013; **69**(2): 313–319. [PubMed Abstract](#) | [Publisher Full Text](#)
- Boyes D, Hammond J, University of Oxford and Wytham Woods Genome Acquisition Lab: **The genome sequence of the Ear Moth, *Amphipoea ocella* (Linnaeus, 1761) [version 1; peer review: awaiting peer review].** *Wellcome Open Res*. 2023; **8**: 356. [Publisher Full Text](#)
- Boyes D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: **The genome sequence of the dun-bar pinion, *Cosmia trapezina* (Linnaeus, 1758) [version 1; peer review: 2 approved].** *Wellcome Open Res*. 2022; **7**: 189. [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](#)
- Conrad KF, Warren MS, Fox R, et al.: **Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis.** *Biological Conservation*. 2006; **132**(3): 279–291. [Publisher Full Text](#)
- Di Tommaso P, Chatzou M, Floden EW, et al.: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol*. 2017; **35**(4): 316–319. [PubMed Abstract](#) | [Publisher Full Text](#)
- Fauna Europea: ***Cosmia (Nemus) pyralina* (Denis & Schiffermüller, 1775).** <https://fauna-eu.org>. 2023; [Accessed 14 September 2023]. [Reference Source](#)
- Fox R, Parsons MS, Harrower CA: **A review of the status of the macro-moths of Great Britain.** Dorset, UK: Butterfly Conservation, 2019.
- 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): gjaa153. [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, Seppely M, et al.: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol*. 2021; **38**(10): 4647–4654. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peacock JW, Cuthbert RA, Lanier GN: **Deployment of traps in a barrier strategy to reduce populations of the European elm bark beetle, and the incidence of Dutch elm disease.** In: *Management of Insect Pests with Semiochemicals*. Boston, MA: Springer, 1981. [Publisher Full Text](#)
- Potter C, Harwood T, Knight J, et al.: **Learning from history, predicting the future: the UK Dutch elm disease outbreak in relation to contemporary tree disease threats.** *Philos Trans R Soc Lond B Biol Sci*. 2011; **366**(1573): 1966–1974. [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): 1665–1680. [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.: **Merquy: 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): 3210–3212. [PubMed Abstract](#) | [Publisher Full Text](#)
- Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo*. 2023a; [Accessed 21 July 2023]. [Publisher Full Text](#)
- Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo*. 2023b; [Accessed 21 July 2023]. [Publisher Full Text](#)
- Turčáni M, Patočka J: **Does intraguild predation of *Cosmia trapezina* L. (Lep.: Noctuidae) influence the abundance of other Lepidoptera forest pests?** *J For Sci*. 2011; **57**(11): 472–482. [Publisher Full Text](#)
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, et al.: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity**

reads. *BioRxiv*. 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](#)

Ward JJ: **Caterpillars**. *Good Words*. 1903; **44**: 431–439.

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 Lunar-spotted Pinion, *Cosmia pyralina* (Denis & Schiffermüller, 1775)**. European Nucleotide Archive. [dataset], accession number PRJEB54090, 2022.

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool**. *Bioinformatics*. 2023; **39**(1): btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 17 September 2024

<https://doi.org/10.21956/wellcomeopenres.22310.r82491>

© 2024 Martinez J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jose Martinez 

University of Florida, Florida, USA

Boyes and Januszczak present a highly complete and meticulously assembled genome of a male Lunar-spotted Pinion, *Cosmia pyralina* (Denis & Schiffermüller, 1775). The assembly, distinguished by an outstanding BUSCO score, is arranged into 31 chromosomal pseudomolecules, including the fully recovered Z sex chromosome. The comprehensive quality of this genomic resource reflects the use of cutting-edge sequencing technologies and rigorous assembly methodologies. This high-quality dataset provides an invaluable tool for evolutionary studies, offering deep insights into the species' genetic makeup and laying a strong foundation for future research on systematics, population genomics, and the broader evolutionary history of Lepidoptera.

The development of this dataset is founded on a clear, well-structured rationale, supported by concise and accessible descriptions that enhance its usability. The carefully selected protocols and methodologies are meticulously designed to ensure both accuracy and reproducibility, adhering to the highest standards of scientific rigor. The dataset is highly standardized, allowing for seamless integration into various research frameworks. Accompanied by comprehensive and detailed illustrations, it provides a visually rich resource that enhances understanding and facilitates data interpretation. Furthermore, its accessibility ensures that it can be readily utilized by researchers across multiple disciplines, making it a versatile and invaluable asset for advancing knowledge in numerous scientific fields.

Finally, the authors tackle a critical issue in natural ecosystems: "insect declines." The species in question, along with others in its genus, appears to be disappearing from Britain, likely due to the spread of Dutch elm disease. This genomic study provides valuable insights into the biology of this species and could play a pivotal role in its conservation.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary Biology, Ecology, IPM, Integrative Biology, Entomology, 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 15 August 2024

<https://doi.org/10.21956/wellcomeopenres.22310.r91944>

© 2024 Wu N et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ningning Wu 

¹ China Agricultural University Education Foundation, Beijing, Beijing, China

² College of Plant Protection, China Agricultural University College of Plant Protection, Beijing, Beijing, China

Shaolong Qiu

China Agricultural University College of Plant Protection, Beijing, Beijing, China

In this study, the authors assembled the genome of the forestry insect *Cosmia pyralina*. They utilized a single adult male specimen and integrated PacBio HiFi sequencing, Hi-C sequencing, and transcriptome sequencing to produce a chromosome-level genome and achieve comprehensive annotation of protein-coding genes. The high quality of the assembly is demonstrated by the favorable QV and BUSCO values.

The rationale for creating the dataset(s) is clearly described. However, while the protocols used are appropriate and the work is technically sound to some extent, there are areas where improvements could be made. The paper notes that the transcriptome sequencing was performed on the insect's abdomen. It might be worth considering whether gene annotation based solely on this abdominal transcriptome could fully capture all coding genes, as additional transcriptome data from other tissues might offer a more comprehensive annotation.

Furthermore, the authors did not address the annotation of repetitive sequences in the genome, which could be an important aspect to include in the analysis. The methods and materials are

detailed enough to allow for replication by others, and the datasets themselves are presented in a clear, usable, and accessible format.

References

1. Wang Y, Xu S: A high-quality genome assembly of the waterlily aphid *Rhopalosiphum nymphaeae*. *Sci Data*. 2024; **11** (1): 194 [PubMed Abstract](#) | [Publisher Full Text](#)

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: Comparative genomics; Population genetics; Toxicology

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
