Check for updates

DATA NOTE

REVISED The genome sequence of the Green Pug moth, *Pasiphila rectangulata* (Linnaeus, 1758)

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

Denise C. Wawman[1],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Edward Grey Institute, Department of Biology, University of Oxford, Oxford, England, UK

v2 **First published:** 15 May 2024, **9**:256
https://doi.org/10.12688/wellcomeopenres.21224.1
Latest published: 06 Oct 2025, **9**:256
https://doi.org/10.12688/wellcomeopenres.21224.2

Abstract

We present a genome assembly from an individual male *Pasiphila rectangulata* (the Green Pug; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 582.5 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.74 kilobases in length. Gene annotation of this assembly on Ensembl identified 17,153 protein coding genes.

Keywords

Pasiphila rectangulata, Green Pug, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ✓ ✓

	1	2	3
version 2 (revision) 06 Oct 2025			✓ view
version 1 15 May 2024	? view	✓ view	

1. **Arjen Van 't Hof**, Institute of Entomology, České Budějovice, Czech Republic
2. **Ananna Ghosh**, Atomic Energy Research Establishment, Dhaka, Bangladesh
3. **Pritha Dey**, National Centre for Biological Sciences-TIFR, Bengaluru, India

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: Wawman DC: Investigation, Resources, 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.

Copyright: © 2025 Wawman DC *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: Wawman DC, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the Green Pug moth, *Pasiphila rectangulata* (Linnaeus, 1758) [version 2; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, **9**:256 https://doi.org/10.12688/wellcomeopenres.21224.2

First published: 15 May 2024, **9**:256 https://doi.org/10.12688/wellcomeopenres.21224.1

REVISED **Amendments from Version 1**

Version 2 of this data note includes information on Chromosome painting with Merian elements and a new Figure 6, which assists in confirming the assignment of the Z chromosome.

We have also used the latest version of Merqury.FK (1.1.2) to recalculate the QV and *k*-mer completeness scores.

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

Species taxonomy

Eukaryota; Opisthokonta; 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; *Pasiphila*; *Pasiphila rectangulata* (Linnaeus, 1758) (NCBI:txid572874).

Background

The Green Pug *Pasiphila rectangulata* is one of a group of small moths in the family Geometridae that are commonly known as pugs, apparently due to their resemblance to the flat-nosed dogs (Marren, 2019). When freshly emerged, the main form is bright green and has a blackish belt around the abdomen, however, in some urban areas, such as London, the English Midlands and northern England, a uniformly dark brown form, f. *anthrax*, predominates (Skinner & Wilson, 2009; Waring *et al.*, 2017).

Pasiphila rectangulata is a native species in the Palaearctic Region, but has been accidentally introduced to North America, where it was first detected in 1970, and possibly to Japan where it was first described in 1957 (Maier, 2005). Eggs overwinter in cracks on trees and the distinctive greenish larvae, with a mid-dorsal stripe, emerge to feed on the buds, flowers and leaves of a range of trees and shrubs, including apples and crab-apples *Malus* spp., pear *Pyrus* spp. cherries *Prunus* spp., hawthorn *Crataegus monogyna*, blackthorn *Prunus spinosa*, and common Juneberry *Amelanchier canadensis* (Maier, 2005; Skinner & Wilson, 2009; Waring *et al.*, 2017). The larvae can cause significant damage in orchards (Maier, 2005). *P. rectangulata* is single brooded, with the adults flying from June until July or August, and can be attracted to light (Skinner & Wilson, 2009; Waring *et al.*, 2017).

Pasiphila rectangulata has previously been partially sequenced – one mitochondrial and seven nuclear markers (Lee *et al.*, 2018) – and RNA sequencing has been used to detect novel viruses in the families Nyamiviridae and Bunyaviridae from specimens caught in a suburban area of Seattle, Washington (Makhsous *et al.*, 2017).

We present a chromosomally complete genome sequence for *Pasiphila rectangulata*, based on one specimen of the green morphotype, collected using a mercury vapour light trap in a rural garden in the hamlet of Bratton, near Minehead, in Somerset, as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from one male *Pasiphila rectangulata* (Figure 1) collected from Bratton, Somerset, UK (51.20, –3.51). A total of 43-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 12 missing joins or mis-joins and removed 3 haplotypic duplications, reducing the scaffold number by 3.70%.

The final assembly has a total length of 582.5 Mb in 51 sequence scaffolds with a scaffold N50 of 20.9 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, 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. Most (99.78%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). Chromosome Z was assigned based on the presence of conserved genes (synteny) with Z chromosomes of *Perizoma affinitatum* (GCA_961405105.1) and *Camptogramma bilineatum* (GCA_958496255.1).

Chromosome painting with Merian elements (Figure 6; Wright *et al.*, 2024) illustrates the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups.

While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to an alternate haplotype have also been deposited.

Figure 1. Photograph of the *Pasiphila rectangulata* (ilPasRect1) specimen used for genome sequencing.

Table 1. **Genome data for *Pasiphila rectangulata*, ilPasRect1.1.**

Project accession data		
Assembly identifier	ilPasRect1.1	
Species	*Pasiphila rectangulata*	
Specimen	ilPasRect1	
NCBI taxonomy ID	572874	
BioProject	PRJEB63440	
BioSample ID (source individual)	SAMEA112226468	
Isolate information	ilPasRect1: whole organism (DNA and Hi-C sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	Primary: 66.2; alternate: 66.6; combined: 66.5	*≥ 40*
k-mer completeness	Primary: 70.17%; alternate: 69.86%; combined: 99.12%	*≥ 95%*
BUSCO**	C:98.2%[S:97.7%,D:0.5%], F:0.5%,M:1.3%,n:5,286	*S > 90%; D < 5%*
Percentage of assembly assigned to chromosomes	99.78%	*≥ 90%*
Sex chromosomes	Z	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.74 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11593805	
Hi-C Illumina	ERR11606320	
Genome assembly		
Assembly accession	GCA_963082625.1	
Accession of alternate haplotype	GCA_963082775.1	
Span (Mb)	582.5	
Number of contigs	107	
Contig N50 length (Mb)	12.6	
Number of scaffolds	51	
Scaffold N50 length (Mb)	20.9	
Longest scaffold (Mb)	33.74	
Genome annotation		
Number of protein-coding genes	17,153	
Number of gene transcripts	17,368	

* 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 version 5.3.2. 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/CAUJBF01.1/dataset/CAUJBF01.1/busco.

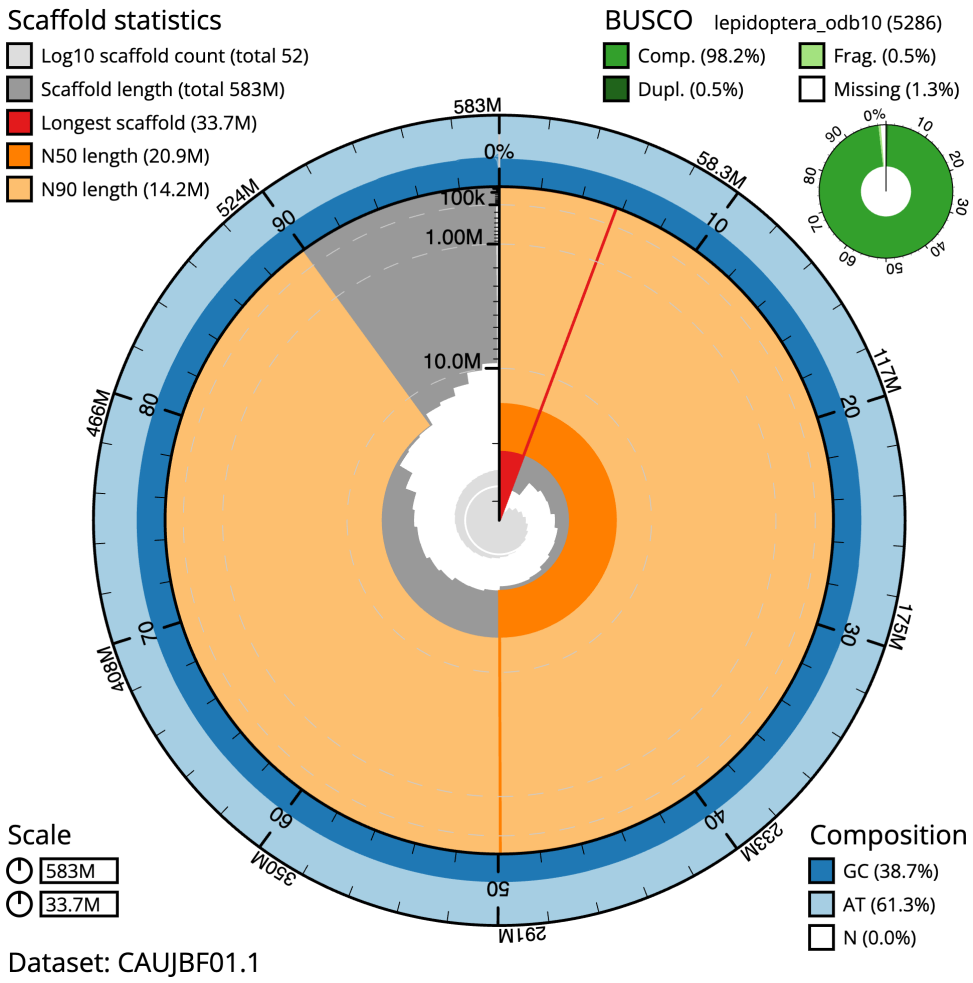

Figure 2. Genome assembly of *Pasiphila rectangulata*, ilPasRect1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 582,513,096 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (33,741,088 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (20,901,828 and 14,179,607 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/CAUJBF01.1/dataset/CAUJBF01.1/snail.

The mitochondrial genome was also assembled (length 15.74 kb, OY720265.1). 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 66.5. The k-mer completeness is 70.17% for the primary assembly, 69.86% for the alternate haplotype, and 99.12% for the combined assemblies (Figure 4). assembly has a BUSCO v5.3.2 completeness of 98.2% (single = 97.7%, duplicated = 0.5%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/572874.

Genome annotation report

The *Pasiphila rectangulata* genome assembly (GCA_963082625.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 17,368 transcribed mRNAs from 17,153 protein-coding and genes (Table 1; https://rapid.ensembl.org/Pasiphila_rectangulata_GCA_963082625.1/Info/Index).

Methods

Sample acquisition and nucleic acid extraction

A male *Pasiphila rectangulata* (specimen ID Ox002242, ToLID ilPasRect1) was collected from Bratton, Somerset, UK (latitude 51.20, longitude –3.51) on 2022-06-20, using a light trap. The specimen was collected and identified by Denise Wawman (University of Oxford) and then preserved on dry ice.

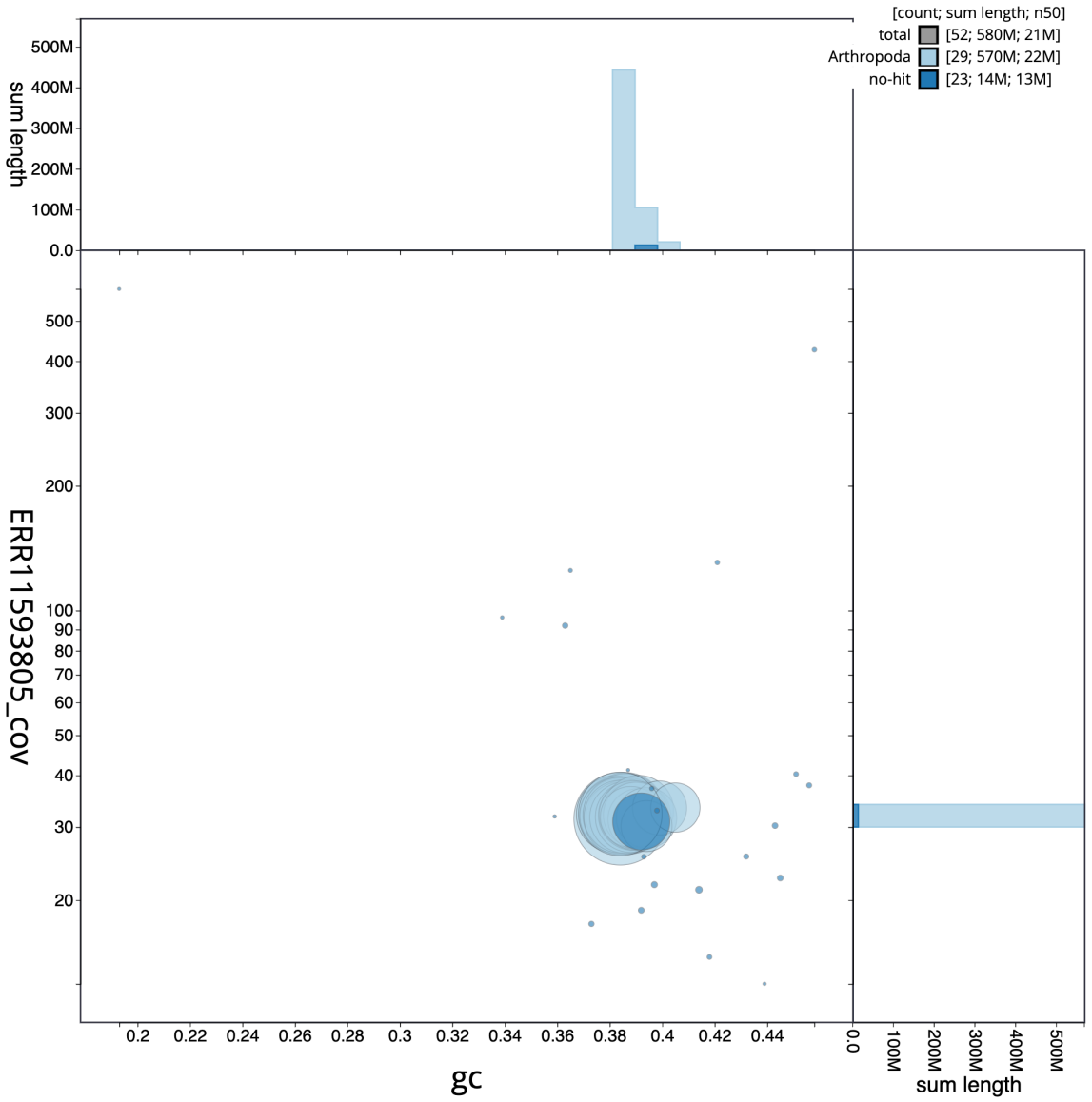

Figure 3. Genome assembly of *Pasiphila rectangulata*, ilPasRect1.1: BlobToolKit Blob plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAUJBF01.1/dataset/CAUJBF01.1/blob.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. The sample was prepared at the WSI Tree of Life Core Laboratory: the ilPasRect1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023), and tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

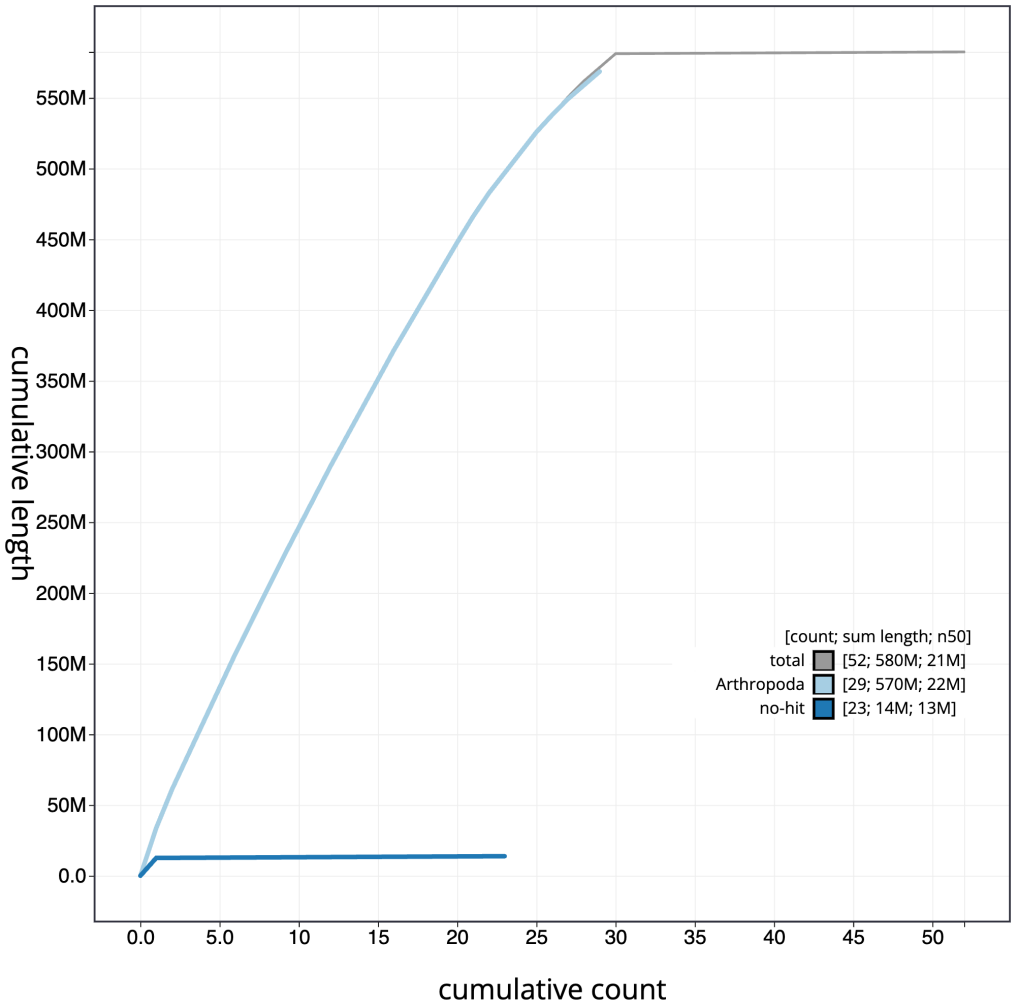

Figure 4. Genome assembly of *Pasiphila rectangulata*, ilPasRect1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAUJBF01.1/dataset/CAUJBF01.1/cumulative.

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 a Pacific Biosciences SEQUEL II instrument. Hi-C data were also generated from remaining tissue of ilPasRect1 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 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.*, 2023), 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 also analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021) were calculated.

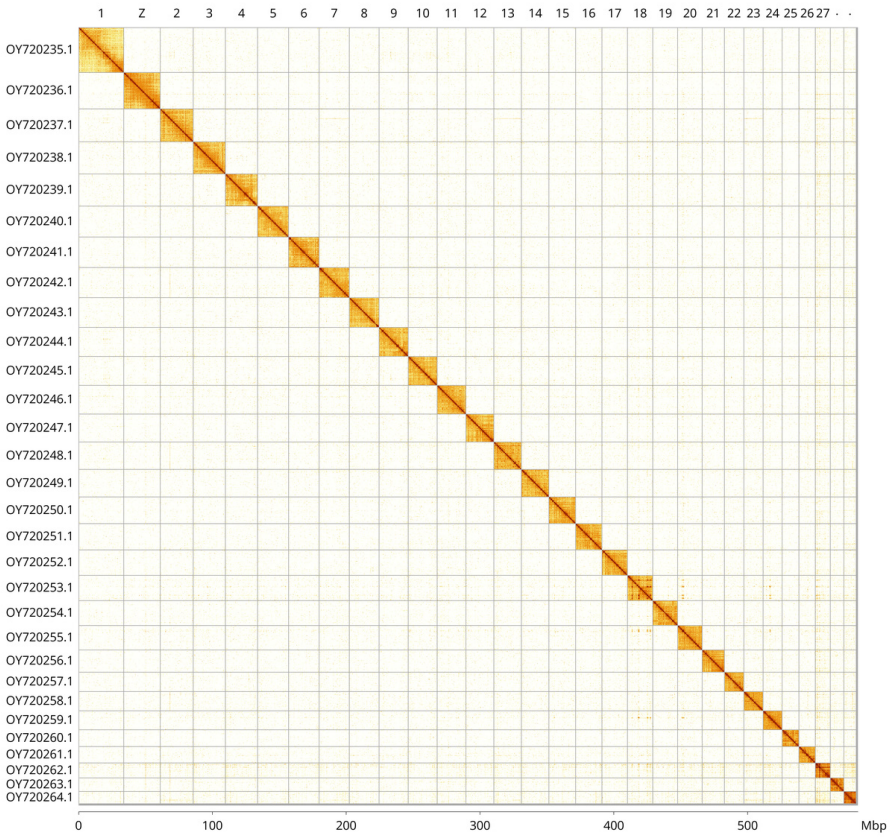

Figure 5. Genome assembly of *Pasiphila rectangulata*, ilPasRect1.1: Hi-C contact map of the ilPasRect1.1 assembly, visualised using PretextView and PretextSnapshot. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. An interactive version of the Hi-C contact map in HiGlass may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=P5lT0Q1CS3aPR2gv7Qh76A.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Pasiphila rectangulata*, ilPasRect1.

INSDC accession	Chromosome	Length (Mb)	GC%
OY720235.1	1	33.74	38.5
OY720237.1	2	24.61	38.5
OY720238.1	3	24.08	38.5
OY720239.1	4	23.95	38.5
OY720240.1	5	23.28	38.0
OY720241.1	6	22.73	38.5
OY720242.1	7	22.62	38.5
OY720243.1	8	22.26	38.5
OY720244.1	9	21.76	38.5
OY720245.1	10	21.57	38.5
OY720246.1	11	21.57	38.5
OY720247.1	12	20.9	38.5
OY720248.1	13	20.59	38.5
OY720249.1	14	20.55	38.5
OY720250.1	15	19.94	38.5
OY720251.1	16	19.59	39.0
OY720252.1	17	19.07	39.0
OY720253.1	18	19.03	39.0
OY720254.1	19	18.59	39.0
OY720255.1	20	18.25	39.0
OY720256.1	21	16.6	39.0
OY720257.1	22	14.58	39.0
OY720258.1	23	14.39	39.0
OY720259.1	24	14.18	39.5
OY720260.1	25	12.61	39.0
OY720261.1	26	12.22	39.0
OY720262.1	27	11.28	40.0
OY720263.1	28	10.02	39.5
OY720264.1	29	9.33	40.5
OY720236.1	Z	27.37	38.5
OY720265.1	MT	0.02	19.5

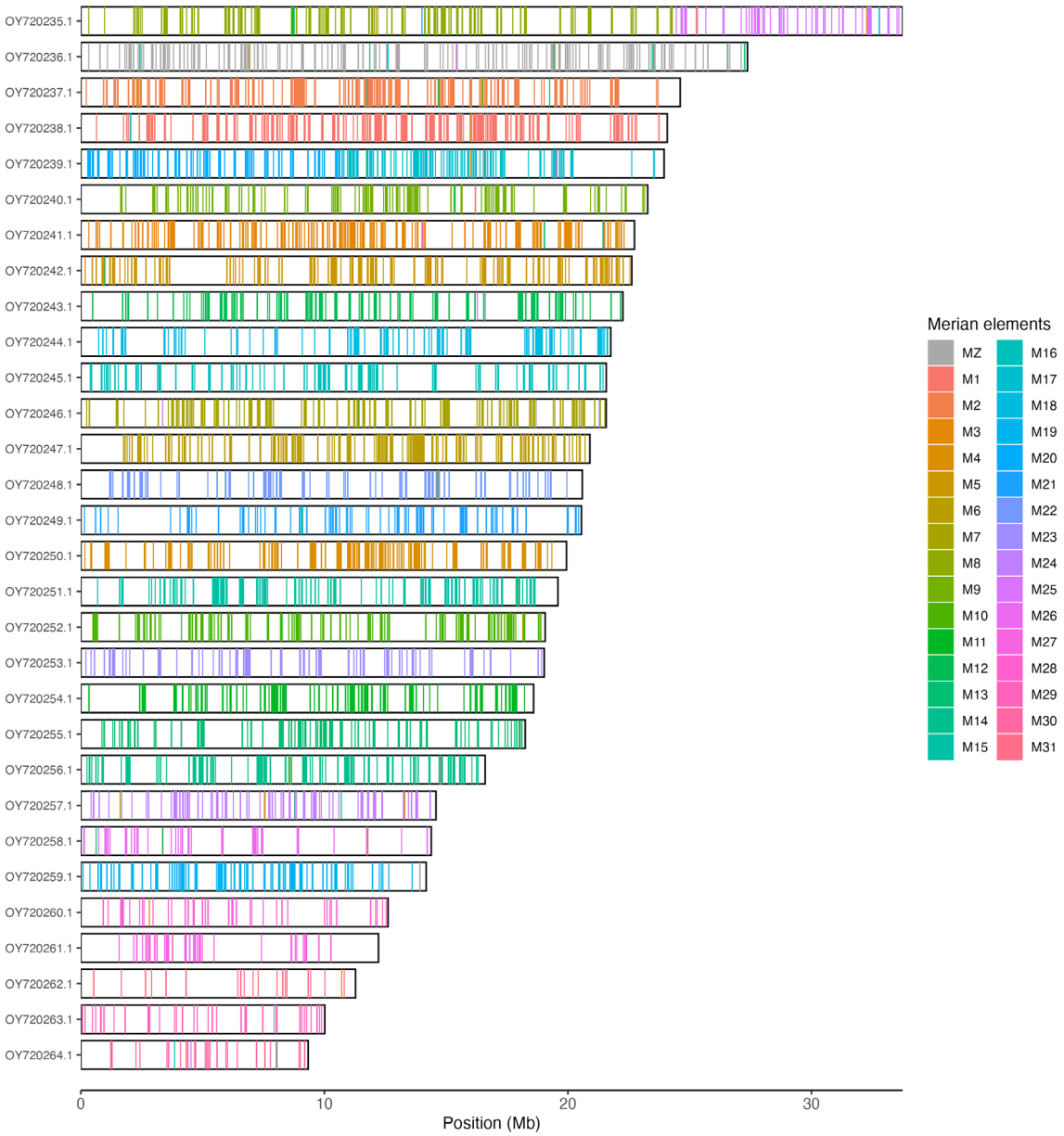

Figure 6. Merian elements painted across chromosomes in the ilPasRect1.1 assembly. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Chromosomal painting was performed using lep_busco_painter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths. Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

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 *Pasiphila rectangulata* assembly (GCA_963082625.1) in Ensembl Rapid Release at the EBI. For further information about the annotation, please refer to the annotation page on Ensembl.

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

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	https://github.com/blobtoolkit/blobtoolkit
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
lep_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter
Merqury.FK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2.5	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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.

Data availability

European Nucleotide Archive: *Pasiphila rectangulata*. Accession number PRJEB63440; https://identifiers.org/ena.embl/PRJEB63440. The genome sequence is released openly for reuse. The *Pasiphila rectangulata* 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.7125292.

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 Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

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

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher 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

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: powerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

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.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
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

Lee KM, Kivelä SM, Ivanov V, *et al.*: **Information dropout patterns in restriction site associated DNA phylogenomics and a comparison with multilocus Sanger data in a species-rich moth genus.** *Syst Biol.* 2018; **67**(6): 925–939.
PubMed Abstract | Publisher Full Text

Maier CT: **Hosts, pattern of seasonal emergence and flight, and new distributional records of palearctic *Pasiphila rectangulata* (Lepidoptera: Geometridae) in the northeastern United States.** *Ann Entomol Soc Am.* 2005; **98**(2): 214–218.
Publisher Full Text

Makhsous N, Shean RC, Droppers D, *et al.*: **Genome sequences of three novel bunyaviruses, two novel rhabdoviruses, and one novel nyamivirus from washington state moths.** *Genome Announc.* 2017; **5**(7): e01668–16.
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

Marren P: **Emperors, admirals, and chimney sweepers: the naming of butterflies and moths.** Toller Fratrum, Dorset: Little Toller, 2019.
Reference Source

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.
Publisher 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.*: **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

Skinner B, Wilson D: **Colour identification guide to the moths of the British Isles.** Leiden, Netherlands: Brill, 2009.
Reference Source

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free 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

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–790.
PubMed Abstract | Publisher Full Text | Free Full Text

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 2

Reviewer Report 28 October 2025

https://doi.org/10.21956/wellcomeopenres.27603.r135459

© **2025 Dey P.** 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.

✓ **Pritha Dey**
National Centre for Biological Sciences-TIFR, Bengaluru, India

This manuscript presents a high-quality genome assembly of *Pasiphila rectangulata* (Green Pug), accompanied by scaffolding to near-chromosome level, mitochondrial genome assembly, and Ensembl-based gene annotation.
The work is technically sound and meets the standards expected for genome reports.

The manuscript is clearly written, well structured, and provides a valuable genomic resource for Lepidoptera research, especially given the relative paucity of chromosome-level genomes among geometrids. I do not see any major methodological or interpretive issues.

However, I felt the figures are less intuitive, specially Figure 3. If the respective axes of the plot can be explained better for readers, it would be really helpful.

Recommendation: Accept as suitable for indexing in its current form.

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: Geometridae moth taxonomy, and evolutionary ecology

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 29 August 2024

https://doi.org/10.21956/wellcomeopenres.23471.r91233

© 2024 Ghosh A. 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.

Ananna Ghosh
[1] Scientific officer, Insect Biotechnology Division, Atomic Energy Research Establishment, Dhaka, Dhaka Division, Bangladesh
[2] Scientific officer, Insect Biotechnology Division, Atomic Energy Research Establishment, Dhaka, Dhaka Division, Bangladesh

To make the outcome more palatable, the abstract needs to be a bit more detailed. The genome annotation report and genome sequence are both excellent. A bit more work should go into the procedure section, especially the nanodrop quantification result.

Overall, the research is commendable and should be approved for indexing.

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: Entomology, Molecular biology and 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.

Reviewer Report 25 August 2024

https://doi.org/10.21956/wellcomeopenres.23471.r91231

© **2024 Hof A.** 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.

?

Arjen Van 't Hof
[1] Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic
[2] Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic

The GCA_963082625.1 assembly of the Green Pug moth is of high quality like we are used to from the Darwin Tree of Life. The assembly statistics show that everything is in order. The assembly is based on a single male according to the manuscript, but the SAMEA112226468 BioSample details show: sex NOT COLLECTED (https://www.ebi.ac.uk/ena/browser/view/SAMEA112226468)[1]. Confusingly, the sequenced male has two different BioSample codes, SAMEA112226468 and SAMEA112232922 with only the first listed in Table 1.

The identification of the Z chromosome is not explained well and based on a string of assumptions.

It is impossible to understand the following sentence without background knowledge.
“Chromosome Z was assigned based on synteny with Perizoma affinitatum (GCA_961405105.1) and Camptogramma bilineatum (GCA_958496255.1).”

What is missing here is an explanation that there is a large body of evidence that chromosomal gene content is conserved in Lepidoptera and can therefore be used to identify homologous chromosomes based on shared synteny. Some references supporting this would also be welcome. I do not object to the ‘assumption-based’ shared synteny Z-chromosome assignment as long as an ‘evidence-based’ Z chromosome is used for comparison. One might expect that the assembly used for shared synteny-based Z chromosome assignment was based on a female with a Z chromosome identified by having half the sequence read coverage of autosomes.

But following GCA_961405105 and GCA_958496255 leads to surprising trails, which do not end with evidence-based Z-chromosome assigned assemblies.

Following the Perizoma affinitatum (GCA_961405105.1) trial:

Perizoma affinitatum (GCA_961405105.1):
Chromosome Z was assigned by synteny to Perizoma flavofasciatum (GCA_958496245.1).

Perizoma flavofasciatum (GCA_958496245.1):
Z chromosome identified based on synteny with Eulithis prunata (GCA_918843925.1).

The Eulithis prunata genome was sequenced from one male. It is not explained how the Z chromosome was identified in (https://doi.org/10.12688/wellcomeopenres.19371.1) Boyes *etal*. (2023[2]), except for "*localised homologous pairs*" mentioned in Table 1

This is where the trail starting with Perizoma affinitatum ends without any satisfactory explanation how the Z chromosome was identified in any of the species.

Following the Camptogramma bilineatum (GCA_958496255.1) trail:

Camptogramma bilineatum (GCA_958496255.1):
Chromosome Z was assigned by synteny to GCA_951804965.1 (Horisme vitalbata)

Horisme vitalbata (GCA_951804965.1):
Z chromosome was assigned based on synteny to GCA_949320045.1 (Scotopteryx bipunctaria)

Scotopteryx bipunctaria (GCA_949320045.1)
Z chromosome identified based on synteny with Gandaritis pyraliata (GCA_947859175.1)

The Gandaritis pyraliata (GCA_947859175.1) assembly paper (https://doi.org/10.12688/wellcomeopenres.19526.1) Boyes *etal*. (2023[3]) does not explain how the Z chromosome was identified. The Z chromosome is mentioned in the abstract, but besides that, all that can be found on sex chromosomes is in table 1:
Sex chromosomes, W and X chromosomes, localised homologous pairs
The presence of an X chromosome in Lepidoptera is obviously incorrect, but this review is not about the Gandaritis pyraliata assembly paper.

The Z chromosome scaffold for the Green Pug moth is probably correct despite the indirect identification. A direct comparison would have been more convincing though.

References

1. Biosample: SAMEA112226468. Reference Source
2. Boyes D, Holland PWH, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: The genome sequence of the Phoenix, Eulithis prunata (Linnaeus, 1758).*Wellcome Open Res*. 2023; **8**: 185 PubMed Abstract | Publisher Full Text
3. Boyes D, Holland P, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: The genome sequence of the Barred Straw, Gandaritis pyraliata (Denis & Schiffermüller, 1775). *Wellcome Open Research*. 2023; **8**. Publisher Full Text

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Lepidoptera genetics

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.

Author Response 28 Sep 2025

Tree of Life Team Sanger

Thank you for your careful review of this data note. We have submitted a new version of the article and our responses to your comments are given below.

1. The assembly is based on a single male according to the manuscript, but the SAMEA112226468 BioSample details show: sex NOT COLLECTED (https://www.ebi.ac.uk/ena/browser/view/SAMEA112226468). Confusingly, the sequenced male has two different BioSample codes, SAMEA112226468 and SAMEA112232922 with only the first listed in Table 1.

Response: This is a standard aspect of the structure of data at the European Nucleotide Archive BioSamples Database. One of the BioSample accessions (SAMEA112226468) identifies the individual or specimen, while the other (SAMEA112232922) identifies the part of the specimen that was used. This is useful if the specimen is dissected into multiple parts, as it can identify different parts used for different sequencing technologies. In this case, the whole organism was used.

1. The identification of the Z chromosome is not explained well and based on a string of assumptions. What is missing here is an explanation that there is a large body of evidence that chromosomal gene content is conserved in Lepidoptera and can therefore be used to identify homologous chromosomes based on shared synteny. Some references supporting this would also be welcome. I do not object to the 'assumption-based' shared synteny Z-chromosome assignment as long as an 'evidence-based' Z chromosome is used for comparison. One might expect that the assembly used for shared synteny-based Z chromosome assignment was based on a female with a Z chromosome identified by having half the sequence read coverage of autosomes.

Response: We identify Z-linked scaffolds by locating BUSCO genes known to be Z-linked across specimens, including heterogametic (ZW) individuals where the Z is unambiguous. In version 2 of the data note we have also applied Merian painting, which shows conserved Z synteny: ancestral Z-linked orthologues co-locate on this chromosome. We have provided a brief explanation and a reference for this paper in the text.

1. It is impossible to understand the following sentence without background knowledge.
"Chromosome Z was assigned based on synteny with Perizoma affinitatum

(GCA_961405105.1) and Camptogramma bilineatum (GCA_958496255.1)."
Response: We have rephrased this slightly in Version 2.

Competing Interests: No competing interests were disclosed.