DATA NOTE

REVISED The genome sequence of the Buff Arches, *Habrosyne pyritoides* (Hufnagel, 1766)

[version 2; peer review: 2 approved, 1 approved with reservations, 1 not approved]

Douglas Boyes[1+], Clare Boyes[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]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

v2 **First published:** 20 Oct 2023, **8**:480
https://doi.org/10.12688/wellcomeopenres.20151.1
Latest published: 26 Mar 2026, **8**:480
https://doi.org/10.12688/wellcomeopenres.20151.2

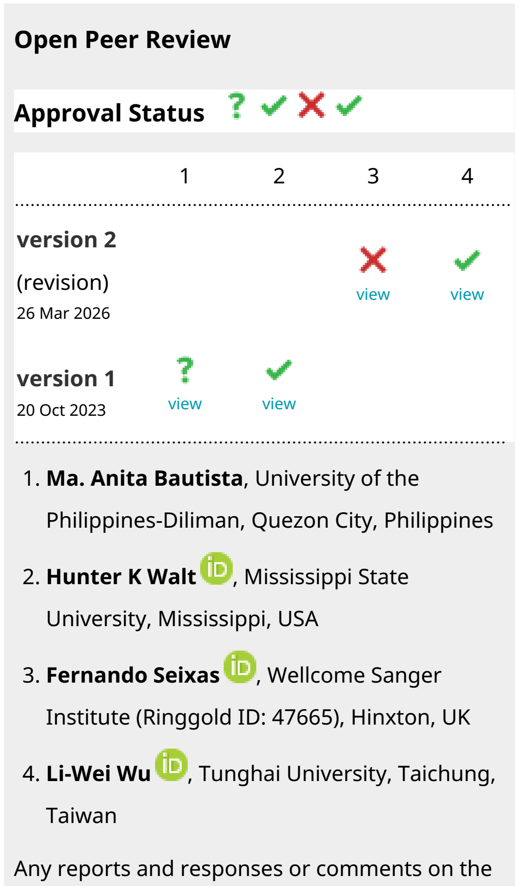
Abstract

We present a genome assembly from an individual male *Habrosyne pyritoides* (the Buff Arches; Arthropoda; Insecta; Lepidoptera; Drepanidae). The genome sequence is 400.6 megabases in span. The whole assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.59 kilobases in length. Gene annotation of this assembly on Ensembl identified 17,018 protein coding genes. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Habrosyne pyritoides, Buff Arches, 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) 26 Mar 2026			✗ view	✓ view
version 1 20 Oct 2023	? view	✓ view		

1. **Ma. Anita Bautista**, University of the Philippines-Diliman, Quezon City, Philippines
2. **Hunter K Walt**, Mississippi State University, Mississippi, USA
3. **Fernando Seixas**, Wellcome Sanger Institute (Ringgold ID: 47665), Hinxton, UK
4. **Li-Wei Wu**, Tunghai University, Taichung, Taiwan

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; **Boyes C**: 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: © 2026 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, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Buff Arches, *Habrosyne pyritoides* (Hufnagel, 1766) [version 2; peer review: 2 approved, 1 approved with reservations, 1 not approved]** Wellcome Open Research 2026, **8**:480 https://doi.org/10.12688/wellcomeopenres.20151.2

First published: 20 Oct 2023, **8**:480 https://doi.org/10.12688/wellcomeopenres.20151.1

REVISED **Amendments from Version 1**

Version 2 of this data note includes links to the wet lab protocols we used for nucleic acid extraction. Figure 5 has been replaced with a version in which the chromosomes are labelled on the axes, with a hyperlink to the interactive version. We updated the Merqury.FK results for this assembly, using a newer version of the software.

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; Drepanoidea; Drepanidae; Thyatirinae; *Habrosyne*; *Habrosyne pyritoides* (Hufnagel, 1766) (NCBI:txid721137).

Background

Habrosyne pyritoides (Buff Arches) is a macro-moth in the family Drepanidae. The species is common throughout England and Wales but is scarce in southern Scotland. The species has declined in abundance by 62% since the 1970s but has increased its range in the UK (Randle *et al.*, 2019). It is found in central Europe and there is also a cluster of records from Japan (GBIF Secretariat, 2023).

H. pyritoides is a moth of open woodland, particularly favouring coppiced areas where its foodplants are common. The larvae feed mainly on bramble, but also use dewberry, and have been found to use raspberry in captivity. The adult moth flies at dusk and in the UK is on the wing from June to August with a partial second brood in autumn in some years. As well as being attracted to light, the moth can be found at nectar or by sugaring (Waring *et al.*, 2017).

The moth shows very little variation, and its appearance is unmistakable. It has as a forewing size of 17–20mm. The wings are slaty grey, with a delicate pattern of orange, brown and white lines. The moth's specific name of *pyritoides* means 'like pyrites' (fool's gold), referring to the distinctive orange markings on the wings (Marren, 2019).

A genome sequence from *H. pyritoides* will be useful for comparative studies across the Lepidoptera. The genome of *H. pyritoides* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *H. pyritoides* based on a male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Habrosyne pyritoides* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.33). A total of 52-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 101-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 3 missing joins or mis-joins and removed 4 haplotypic duplications, reducing the scaffold number by 18.42%.

The final assembly has a total length of 400.6 Mb in 31 sequence scaffolds with a scaffold N50 of 14.1 Mb (Table 1). 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. The whole assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 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). 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 combined primary and alternate assemblies achieve an estimated QV of 60.1. The *k*-mer completeness is 98.30% for the primary assembly, 96.31% for the alternate haplotype, and 99.57% for the combined assemblies. The primary assembly has a BUSCO v5.3.2 completeness of 98.9% (single = 98.7%, duplicated = 0.3%), using the lepidoptera_odb10 reference set (n = 5,286).

Figure 1. Photograph of the *Habrosyne pyritoides* (ilHabPyri1) specimen used for genome sequencing.

Genome annotation report

The *Habrosyne pyritoides* genome assembly (GCA_907165245.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 17 239 transcribed mRNAs from 17 018 protein-coding genes. The average transcript length is 6 644.44 bp, with an average of 5.80 exons per transcript. Annotation files may be downloaded from the Ensembl annotation page.

Methods

Sample acquisition and nucleic acid extraction

A male *Habrosyne pyritoides* (specimen ID Ox000531, ToLID ilHabPyri1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.77, longitude –1.33) on 2020-06-25 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

Table 1. Genome data for *Habrosyne pyritoides*, ilHabPyri1.1.

Project accession data	
Assembly identifier	ilHabPyri1.1
Assembly release date	2021-05-17
Species	*Habrosyne pyritoides*
Specimen	ilHabPyri1
NCBI taxonomy ID	721137
BioProject	PRJEB44836
BioSample ID	SAMEA7701298
Isolate information	ilHabPyri1, male: abdomen (DNA sequencing), head and thorax (Hi-C scaffolding and RNA sequencing)

Table 1. *Continued*

Assembly metrics*		***Benchmark***
Consensus quality (QV)	Primary: 61.1; alternate: 59.7; combined: 60.1	*≥ 40*
k-mer completeness	Primary: 98.30%; alternate: 96.31%; combined: 99.57%	*≥ 95%*
BUSCO**	C:98.9%[S:98.7%,D:0.3%], F:0.3%,M:0.8%, n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	100%	*≥ 90%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6454728	
10X Genomics Illumina	ERR6054722, ERR6054723, ERR6054725, ERR6054724	
Hi-C Illumina	ERR6054726	
PolyA RNA-Seq Illumina	ERR9434975	
Genome assembly		
Assembly accession	GCA_907165245.1	
Accession of alternate haplotype	GCA_907165225.1	
Span (Mb)	400.6	
Number of contigs	38	
Contig N50 length (Mb)	14.1	
Number of scaffolds	31	
Scaffold N50 length (Mb)	14.1	
Longest scaffold (Mb)	23.7	
Genome annotation		
Number of protein-coding genes	17,018	
Number of gene transcripts	17,239	

*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/Habrosyne%20pyritoides/dataset/ilHabPyri1_1.1/busco.

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The ilHabPyri1 sample was weighed and triaged to determine the appropriate extraction protocol.

Abdomen tissue was cryogenic disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit using the Manual MagAttract protocol. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using the 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30, following the protocol. 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, following the manual SPRI protocol. 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.

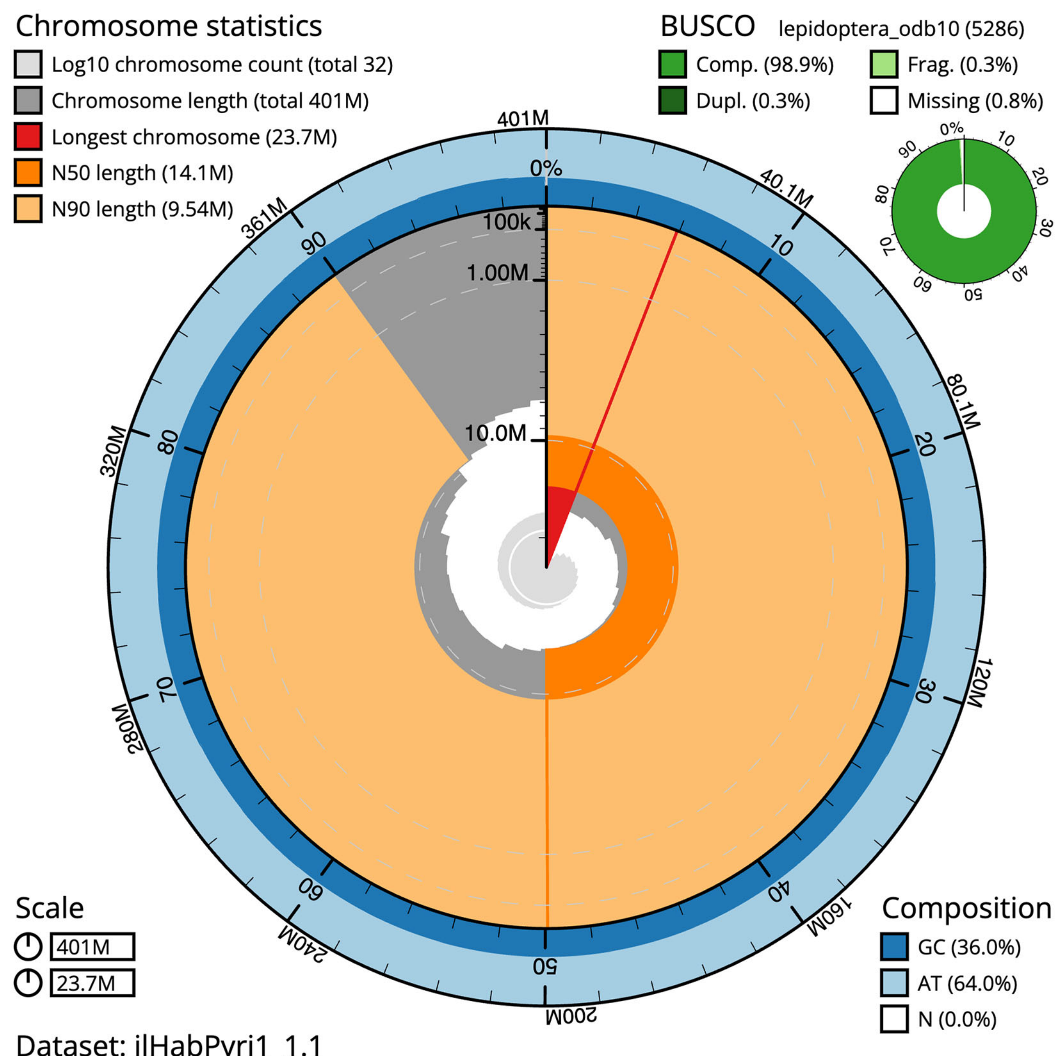


Figure 2. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 400,568,986 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (23,742,291 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (14,121,114 and 9,540,854 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/Habrosyne%20pyritoides/dataset/ilHabPyri1_1.1/snail.

RNA was extracted from head and thorax tissue of ilHabPyri1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the protocol published on protocols.io. RNA was then eluted in 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 and 10X Genomics read cloud 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), Illumina HiSeq 4000 (RNA-Seq) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from head and thorax tissue of ilHabPyri1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

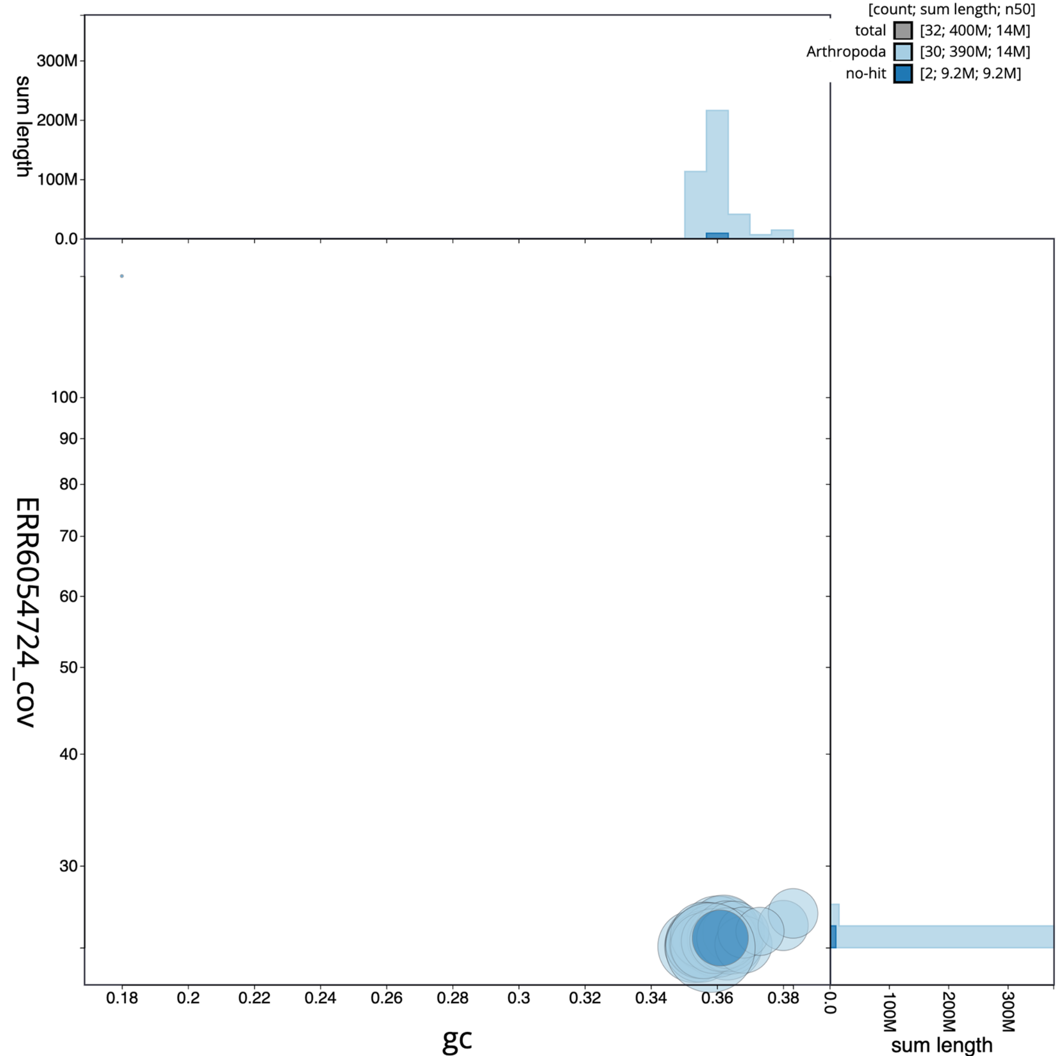


Figure 3. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Habrosyne%20pyritoides/dataset/ilHabPyri1_1.1/blob.

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). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (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 analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

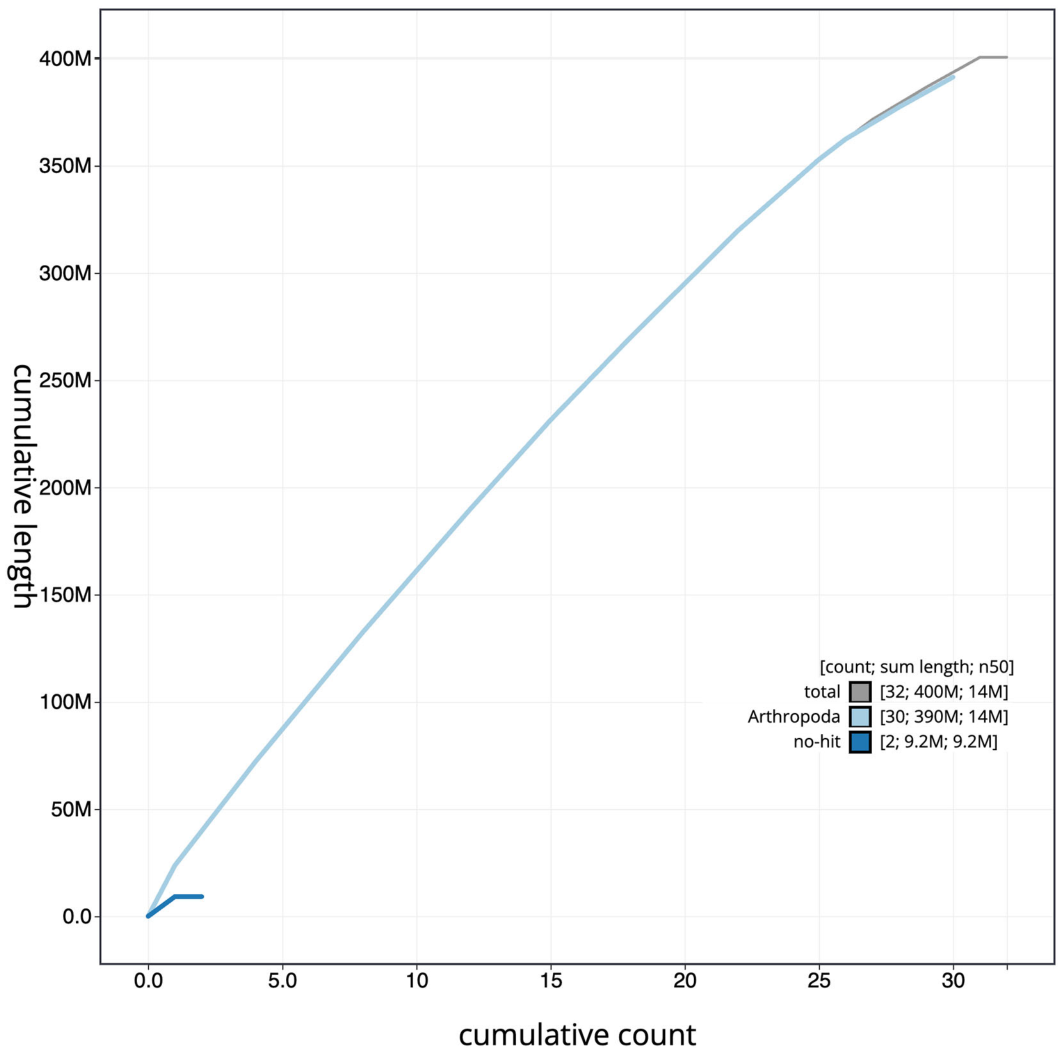


Figure 4. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Habrosyne%20pyritoides/dataset/ilHabPyri1_1.1/cumulative.

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 *Habrosyne pyritoides* assembly (GCA_907165245.1) 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.

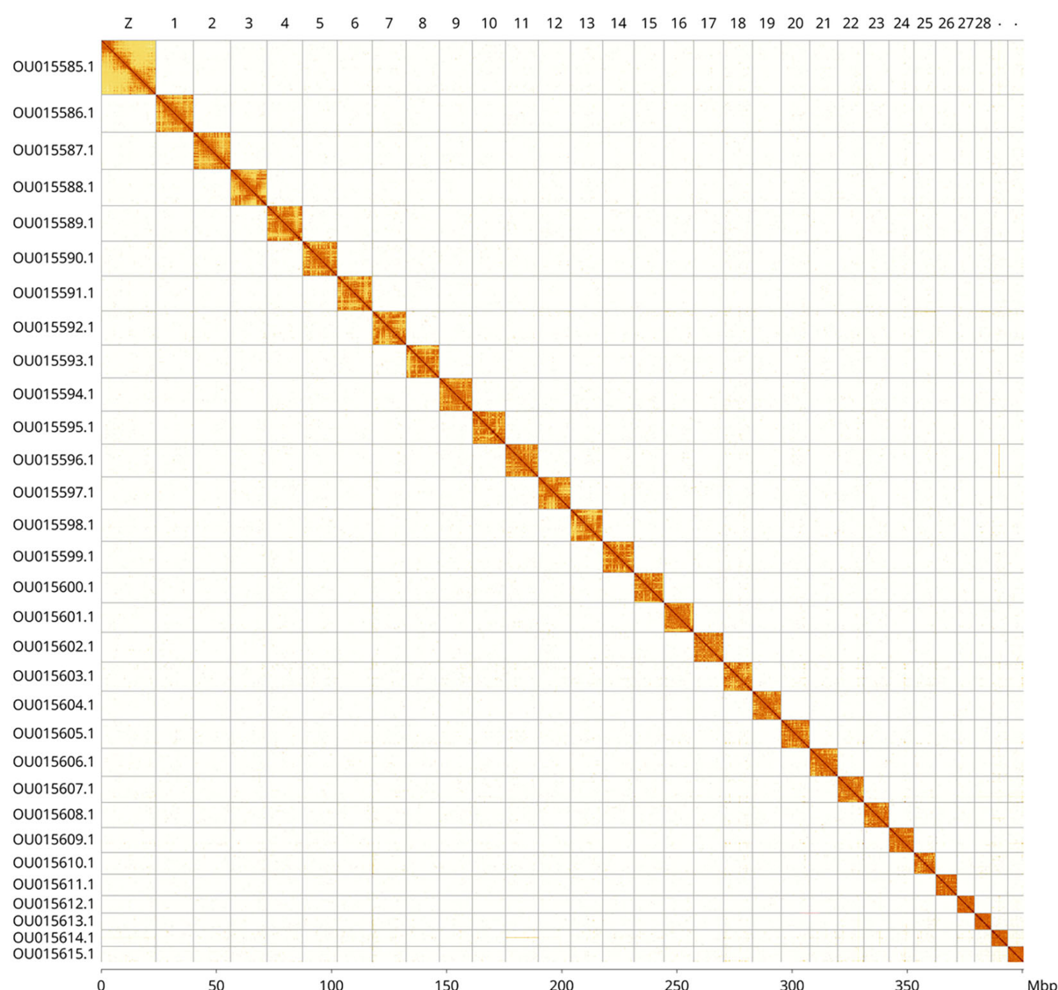


Figure 5. Hi-C contact map of the *Habrosyne pyritoides* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot. Chromosomes are shown in order of size from left to right and top to bottom. An interactive HiGlass version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=B3zYGzGyRUqj-nevo640eHg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Habrosyne pyritoides*, ilHabPyri1.

INSDC accession	Chromosome	Length (Mb)	GC%
OU015586.1	1	16.4	36.2
OU015587.1	2	16.07	35.9
OU015588.1	3	15.88	36
OU015589.1	4	15.32	35.3
OU015590.1	5	15.17	35.5
OU015591.1	6	15.11	36.2
OU015592.1	7	14.76	36.3
OU015593.1	8	14.43	35.5
OU015594.1	9	14.36	35.6
OU015595.1	10	14.31	35.5

Table 2. *Continued*

INSDC accession	Chromosome	Length (Mb)	GC%
OU015596.1	11	14.24	35.7
OU015597.1	12	14.12	36
OU015598.1	13	13.89	35.9
OU015599.1	14	13.62	35.6
OU015600.1	15	12.99	35.6
OU015601.1	16	12.93	36.1
OU015602.1	17	12.88	35.6
OU015603.1	18	12.62	36.4
OU015604.1	19	12.53	36.3
OU015605.1	20	12.37	35.9
OU015606.1	21	12.18	36.3
OU015607.1	22	11.29	36.5
OU015608.1	23	10.98	36.1
OU015609.1	24	10.74	36.3
OU015610.1	25	9.54	36.8
OU015611.1	26	9.2	36.1
OU015612.1	27	7.6	36.8
OU015613.1	28	7.35	38
OU015614.1	29	7.19	38.3
OU015615.1	30	6.77	37.3
OU015585.1	Z	23.74	35.8
OU015616.1	MT	0.02	18.2

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
FreeBayes	1.3.1-17-gaa2ace8	https://github.com/freebayes/freebayes
gEVAL	N/A	https://geval.org.uk/
Hifiasm	0.14-r312	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
Merqury.FK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	v2.11.3	https://github.com/marcelauliano/MitoHiFi
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
SALSA	2.2	https://github.com/salsa-rs/salsa

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.

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.

Data availability

European Nucleotide Archive: *Habrosyne pyritoides* (buff arches). Accession number PRJEB44836; https://identifiers.org/ena.embl/PRJEB44836 (Wellcome Sanger Institute, 2021). The genome sequence is released openly for reuse. The *Habrosyne pyritoides* 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.

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

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL — a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510.
PubMed Abstract | Publisher Full Text | Free Full Text

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** *arXiv:1207.3907v2.* 2012. [Accessed 26 July 2023].
Publisher Full Text

GBIF Secretariat: ***Habrosyne pyritoides* Hufnagel, 1767.** GBIF Backbone Taxonomy. 2023. [Accessed 15 September 2023].

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput. Biol.* 2019; **15**(8): e1007273.
PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
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].

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.

Oxford University Press.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol. Biol. Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Marren P: *Emperors, Admirals, and Chimney Sweepers.* Toller Fratrum, Dorset: Little Toller; 2019.

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: *Atlas of Britain & Ireland's Larger Moths.* Newbury: NatureBureau; 2019.

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

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

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.** *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 ed. Bloomsbury Wildlife Guides; 2017.

Wellcome Sanger Institute: The genome sequence of the Buff Arches, Habrosyne pyritoides (Hufnagel, 1766). [dataset]. *European Nucleotide Archive.* 2021. accession number PRJEB44836.

Open Peer Review

Current Peer Review Status:

Version 2

Reviewer Report 07 May 2026

https://doi.org/10.21956/wellcomeopenres.28652.r152909

© **2026 Wu L.** 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.

Li-Wei Wu
Tunghai University, Taichung, Taichung City, Taiwan

The article presents a high-quality genome assembly for *Habrosyne pyritoides*, generated using a combination of long-read Pacific Biosciences HiFi sequencing, Hi-C data, and Illumina short reads. The results assembly comprises 30 autosomes and one Z chromosome, consistent with the chromosome number reported for many lepidopteran insects. The result also shows high completeness, with a BUSCO score of 98.9%. Overall, the authors provide a valuable and high-quality genomic resource that will be useful for future comparative genomic, evolutionary, and functional studies.

One point that may merit clarification concerns the DNA source. Because abdominal tissue was used and the specimens were collected from the field, potential contamination from gut-associated microorganisms could be a concern. It would be helpful if the authors could clarify whether microbial contamination was considered during sample preparation or whether any specific bioinformatic filtering steps were applied to detect and remove possible contaminants during genome assembly and annotation. Overall, this appears to be a well-executed work with a sound methodological workflow. The results are reasonable and provide a useful genomic resource for future research. I look forward to seeing further studies that build upon this genome assembly.

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: conservation genetics, phylogenetics

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 06 May 2026

https://doi.org/10.21956/wellcomeopenres.28652.r152094

© 2026 Seixas F. 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.

Fernando Seixas
Wellcome Sanger Institute (Ringgold ID: 47665), Hinxton, England, UK

In the Genome Note, the authors present a chromosome-level genome assembly and annotation for *Habrosyne pyritoides.*

As a Genome Note, I would expect that it should provide a complete and clear explanation of the molecular and analytical methods used to generate the data, the genome assembly and its annotation. However, I find the manuscript in its current state lacking in this respect. In particular the methodology section lacks sufficient detail; key aspects such as protocols, library preparation, and which technologies or platforms were used are often not adequately described and often different technologies are mentioned in the same paragraph. The manuscript also does not seem to follow the typical structure of other Genome notes, which I found a bit confusing and made it difficult to read. Please find more detailed comments below:

Genome Sequence Report
1. The raw data generated for PacBio and Chromium 10X should be reported – e.g. X Gb (gigabases) from Y million reads. The coverage then depends on the assembly size or estimated haploid genome size (which are only generated downstream). And could you please also specify the amount of HiC data generated?
2. “The whole assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome.”. How was the Z chromosome identified?
3. “The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.”. What is the size and accession number?

Methods
1. “Protocols for high molecular weight (HMW) DNA extraction […]”. I believe this sentence should be preceded by a subheader ‘Nucleic acid extraction’. Also, in this section the authors mention that HMW DNA extracted for the abdomen was used for 10X Chromium sequencing and RNA extracted from the head and thorax used for RNA sequencing. However there is no mention of the DNA

extractions used for PacBio and Hi-C sequencing.

2. I believe the section currently labelled as 'Sequencing' should be divided into different subsections since different sequencing technologies (Pacbio, Chromium 10X, Hi-C and RNA) were used and each of these required different methodologies for the preparation of the libraries and different sequencing platforms. For instance, in other Genome Notes divide this is found under different sections named'PacBio HiFi library preparation and sequencing', 'Hi-C', and 'RNA library preparation and sequencing'. I believe that the authors should do the same and describe more thoroughly how the different libraries were prepared, which kits were used and protocols were used, and which platforms and e.g. types of cells were used for sequencing, etc.

3. 'Genome annotation'. Again I find this section lacking information regarding how the gene annotation was performed. For instance, what protein evidence was used for the annotation? And since RNA data was generated for this individual, was it used to help in gene annotation? If not, what was it used for?

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?

No

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

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Evolutionary Biology; Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Version 1

Reviewer Report 06 November 2024

https://doi.org/10.21956/wellcomeopenres.22313.r107849

© 2024 Walt H. 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.

Hunter K Walt
Mississippi State University, Mississippi, Mississippi, USA

In this data note, the authors provide a chromosome-level genome assembly for the Buff arches, a moth found across Europe that has also been reported in Japan. The genome is highly contiguous and has a high completeness score according to appropriate metrics. This high-quality genome is valuable as it can be used for comparative genomic studies in the Lepidoptera and also as a reference for other genomic/transcriptomic analyses in this species.

Overall, this data report is well done and the genome is similar to other genomes sequenced within this family. The report is succinctly written in the standard Darwin Tree of Life format.

I only have a couple of comments to make.

1.) MitoHiFi uses a mitogenome from a closely related organism to assemble novel mitochondrial genomes. I understand that it can automatically infer this, however, the species that was used for this assembly should be reported to ensure reproducibility.

2.) Generally, genomes should be repeat-masked before annotation as it increases accuracy. Was the genome repeat-masked before using the BRAKER2 pipeline? This should be reported to ensure the accuracy and reproducibility of the annotation. Additionally, I suggest reporting some information about the repetitive landscape such as the percent of the genome that was masked.

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: Insect genomics and evolution, transcriptomics, metagenomics

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 10 June 2024

https://doi.org/10.21956/wellcomeopenres.22313.r79526

© 2024 Bautista M. 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.

?

Ma. Anita Bautista
University of the Philippines-Diliman, Quezon City, Philippines

The paper focuses on generating an assembled genome for *Habrosyne pyritoides* (the Buff Arches, Arthropoda, Insecta, Lepidoptera, Drepanidae). The availability of this genome from another species under Drepanidae will enable comparative studies across the Lepidoptera, especially those present in the Atlantic Archipelago of Britain and Ireland.

While the paper published useful genome information, I found that some details are lacking in the methodology or a few information need to be clarified. For example, RNA was extracted and sequenced for annotation purposes, but can the authors mention if the RNAs were DNAse treated? Also, were there replications needed for RNA Seq? What platforms were used to generate the transcriptome sequences?

Finally, I found that the writing style used and even how the paper was organized was highly similar to another published paper on another Drepanidae. See **Crowley and Philips, 2023** https://wellcomeopenresearch.org/articles/8-458.

The paper must be re-written in another style.

References

1. Crowley L, Phillips D, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al.: The genome sequence of the Yellow Horned, Achlya flavicornis (Linnaeus, 1758). *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?

No

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Insect genomics, transcriptomics, metagenomics

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 04 Feb 2026

Tree of Life Team Sanger

We thank the reviewer for their comments on this data note. The similarity in structure and writing style to other published data notes, including Crowley and Phillips (2023), is intentional and reflects the standardised format used for genome notes generated as part of the Darwin Tree of Life project. These articles follow a common structure to ensure consistency and comparability across a large number of species, and to allow readers to locate key information efficiently. This approach is aligned with the scope and aims of the project, which is producing genome notes for many thousands of taxa using a consistent analytical and reporting framework. We have included links to the appropriate wet lab protocols used in the Methods section, and to the github page for production code in the Data availability section. Tree of Life team, Wellcome Sanger Institute

Competing Interests: No competing interests were disclosed.