Check for updates

DATA NOTE

The genome sequence of the Buff Arches, *Habrosyne pyritoides* (Hufnagel, 1766) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Clare Boyes [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

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]Independent researcher, Welshpool, Wales, UK

[+] Deceased author

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

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.

Keywords
Habrosyne pyritoides, Buff Arches, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

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

Author roles: Boyes D: Investigation, Resources; **Boyes C**: Writing – Original Draft Preparation;

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 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:480 https://doi.org/10.12688/wellcomeopenres.20151.1

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

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; 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

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

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 estimated Quality Value (QV) of the final assembly is 61.1 with *k*-mer completeness of 100%, and the 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).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/721137.

Genome annotation report

The *Habrosyne pyritoides* genome assembly (GCA_907165245.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Habrosyne_pyritoides_GCA_907165245.1/Info/Index). The resulting annotation includes 17,239 transcribed mRNAs from 17,018 protein-coding non-coding genes.

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.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilHabPyri1 sample was weighed and

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)

Assembly metrics*		Benchmark
Consensus quality (QV)	61.1	≥ 50
k-mer completeness	100%	≥ 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%	≥ 95%
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.
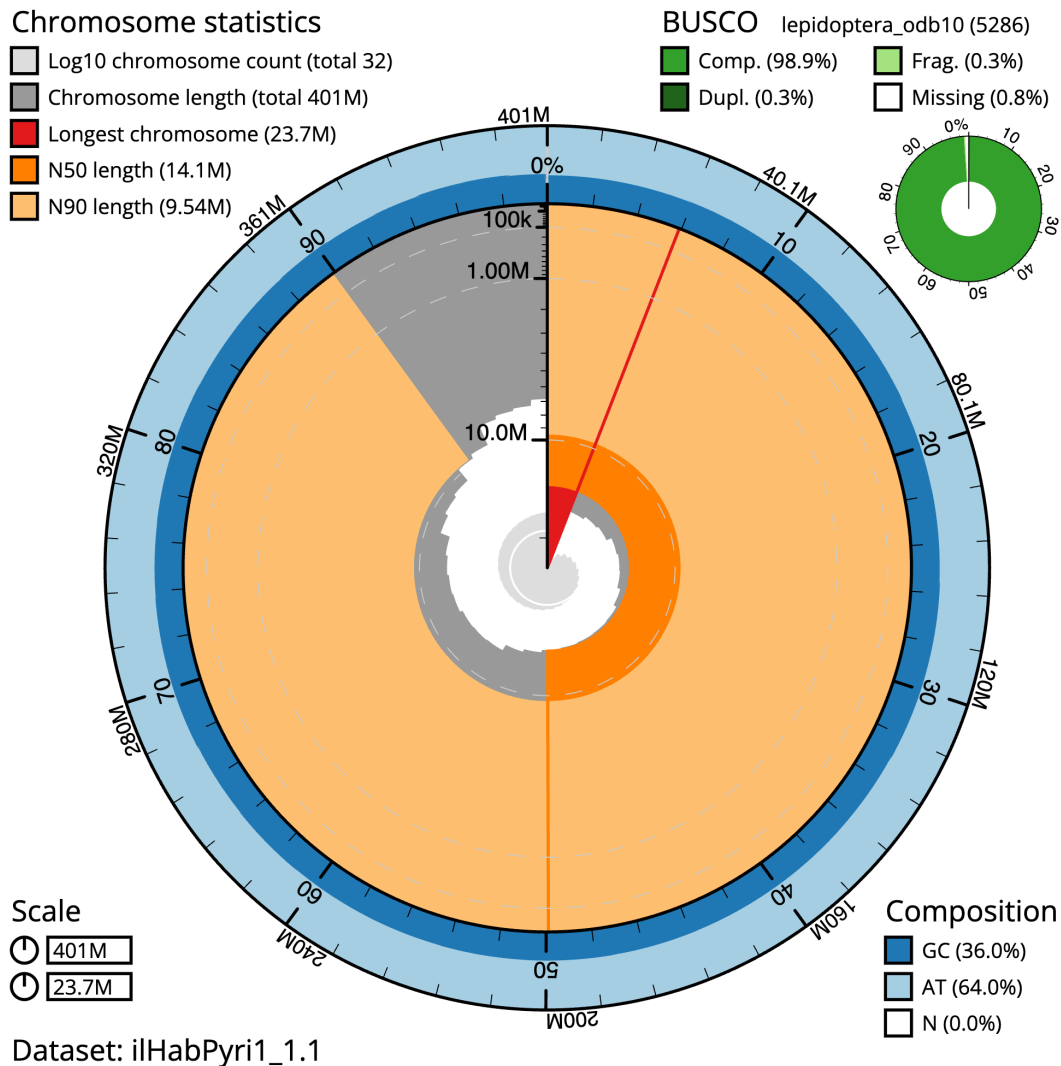
Figure 2. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 400,568,986 bp assembly. The distribution 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.

dissected on dry ice with tissue set aside for Hi-C sequencing. Abdomen tissue was cryogenically 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. 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. 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 head and thorax tissue of ilHabPyri1 in the Tree of Life Laboratory at the WSI using
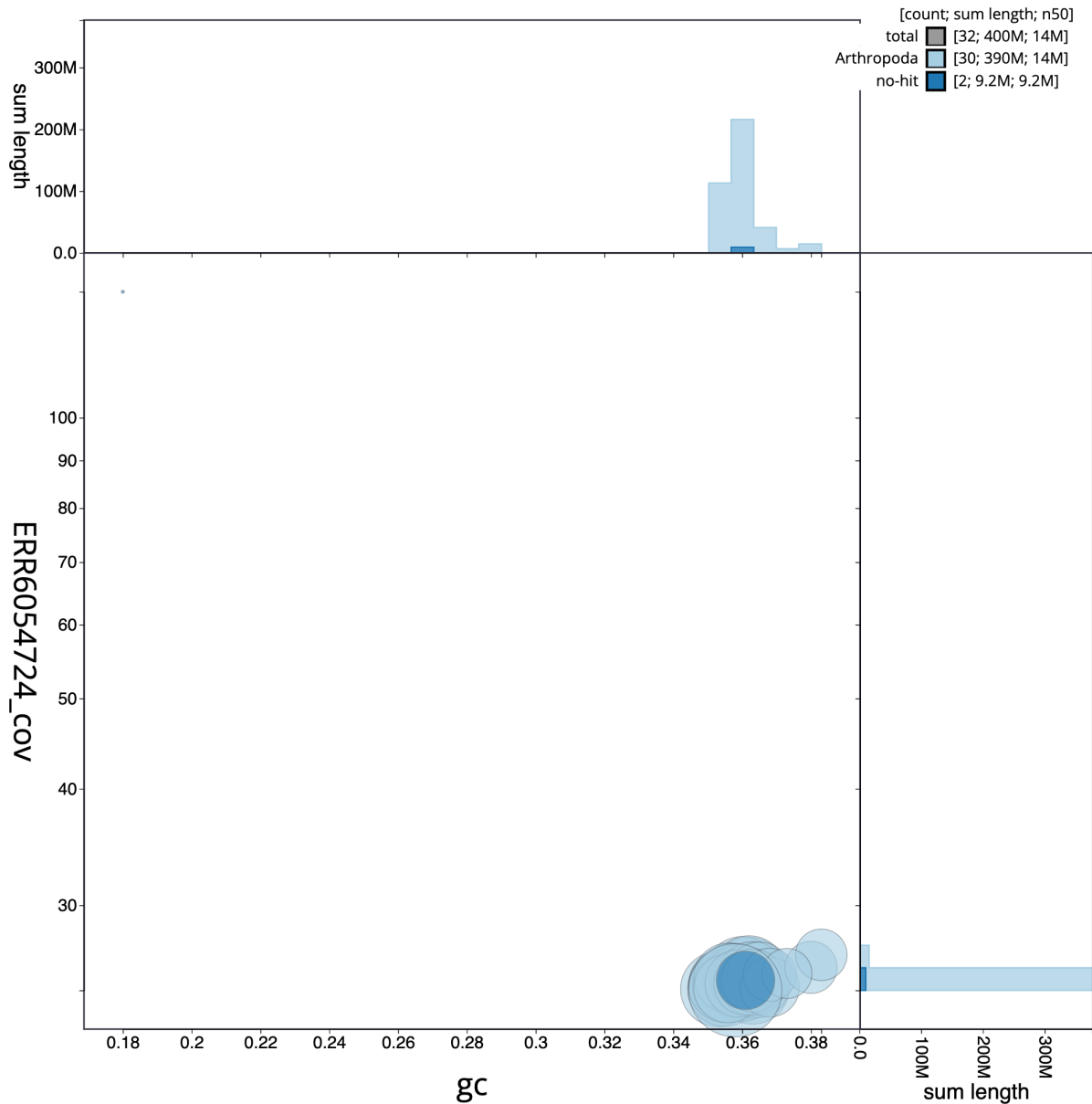
Figure 3. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Habrosyne%20pyritoides/dataset/ilHabPyri1_1.1/blob.

TRIzol, according to the manufacturer's instructions. 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.

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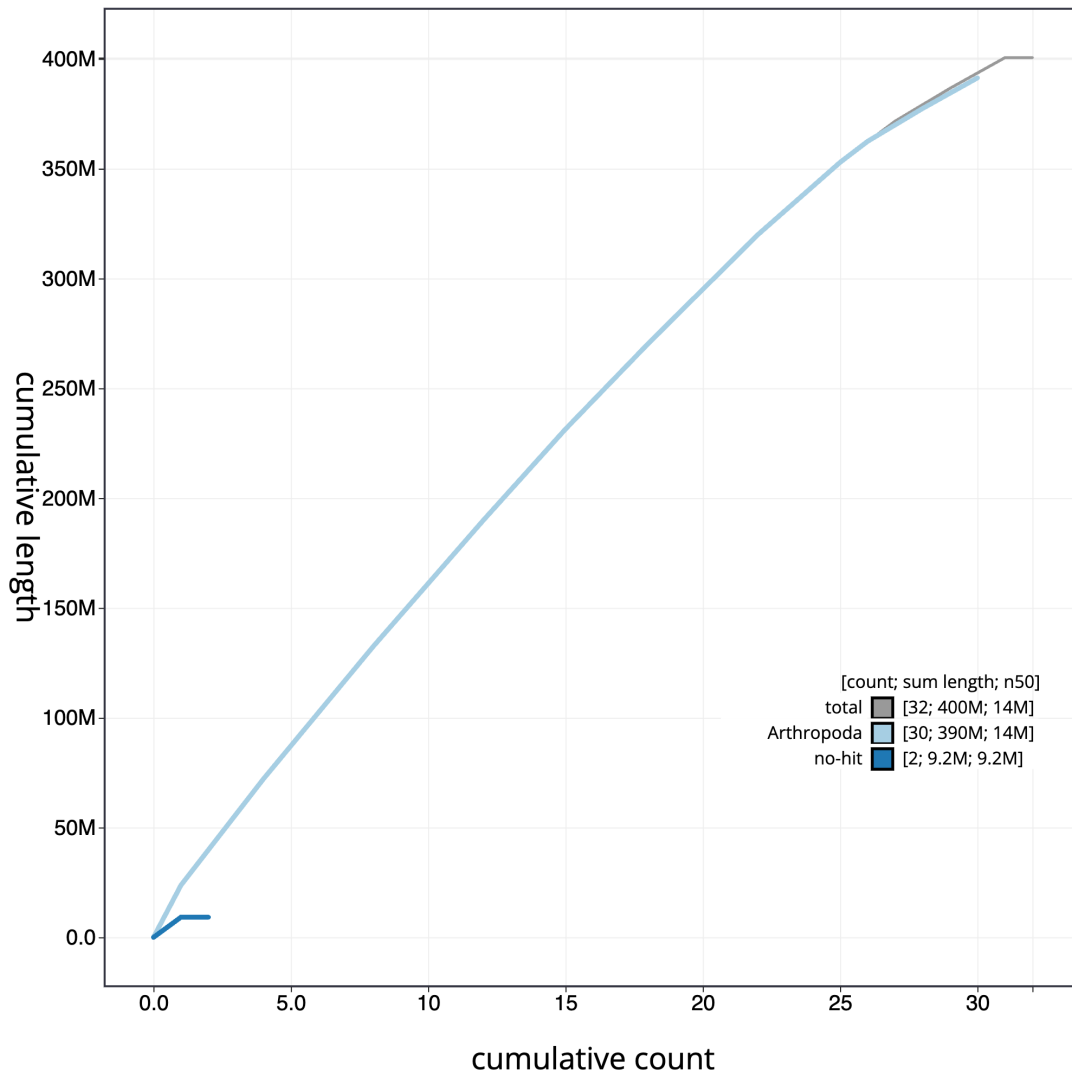
Figure 4. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure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

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) 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.

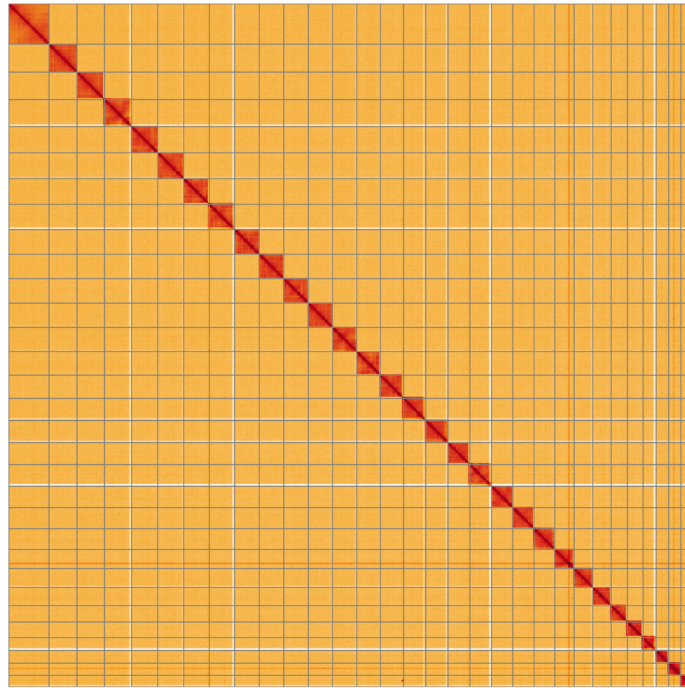Table 3 contains a list of relevant software tool versions and sources.

Figure 5. Genome assembly of *Habrosyne pyritoides*, ilHabPyri1.1: Hi-C contact map of the ilHabPyri1.1 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=B3zYGzGyRUqjnevo640eHg.

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
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

INSDC accession	Chromosome	Length (Mb)	GC%
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	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	v2.11.3	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
SALSA	2.2	https://github.com/salsa-rs/salsa
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

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. 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: *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.

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**

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**

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**

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].
Reference Source

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

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

Harry E: **PretextView (Paired REad TEXTURE Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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.
Reference Source

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

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

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

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