



DATA NOTE

# The genome sequence of the Oak Beauty, *Biston strataria* (Hufnagel, 1767) [version 1; peer review: 2 approved]

Douglas Boyes<sup>1+</sup>, David C. Lees<sup>id</sup><sup>2</sup>, Peter W.H. Holland<sup>3</sup>,  
University of Oxford and Wytham Woods Genome Acquisition Lab,  
Natural History Museum 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

<sup>1</sup>UK Centre for Ecology & Hydrology, Wallingford, England, UK

<sup>2</sup>Natural History Museum, London, England, UK

<sup>3</sup>University of Oxford, Oxford, England, UK

+ Deceased author

**V1** First published: 15 Feb 2024, 9:35  
<https://doi.org/10.12688/wellcomeopenres.20835.1>  
Latest published: 15 Feb 2024, 9:35  
<https://doi.org/10.12688/wellcomeopenres.20835.1>

## Abstract

We present a genome assembly from an individual male *Biston strataria* (the Oak Beauty; Arthropoda; Insecta; Lepidoptera; Geometridae). The genome sequence is 424.0 megabases in span. Most of the assembly is scaffolded into 16 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.61 kilobases in length. Gene annotation of this assembly on Ensembl identified 18,406 protein coding genes.

## Keywords

*Biston strataria*, Oak Beauty, genome sequence, chromosomal, Lepidoptera



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

## Open Peer Review

Approval Status

	1	2
<b>version 1</b> 15 Feb 2024	 <a href="#">view</a>	 <a href="#">view</a>

1. **Kumar Saurabh Singh** , Wageningen University & Research, Wageningen, The Netherlands
2. **Tyler Scott Alioto** , Fundacion Centro Nacional de Analisis Genomico, Barcelona, Spain

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](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** **Boyes D:** Investigation, Resources; **Lees DC:** Investigation, Resources; **Holland PWH:** Writing – Original Draft Preparation, Writing – Review & Editing;

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, <https://doi.org/10.35802/206194>] and the Darwin Tree of Life Discretionary Award [218328, <https://doi.org/10.35802/218328>]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 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, Lees DC, Holland PWH *et al.* **The genome sequence of the Oak Beauty, *Biston strataria* (Hufnagel, 1767) [version 1; peer review: 2 approved]** Wellcome Open Research 2024, 9:35 <https://doi.org/10.12688/wellcomeopenres.20835.1>

**First published:** 15 Feb 2024, 9:35 <https://doi.org/10.12688/wellcomeopenres.20835.1>

## 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; Ennominae; *Biston*; *Biston strataria* (Hufnagel, 1767) (NCBI:txid722658).

## Background

The Oak Beauty *Biston strataria* (synonym *stratarius*) is a large geometrid moth (wingspan 40–50 mm) with two jagged chestnut brown bands, edged in black, crossing the speckled white and black forewings. *B. strataria* is found across northern, central and eastern Europe, including southern parts of Scandinavia, central and southern counties of Britain, and scattered sites across Ireland. There are also records from Russia, Georgia, Turkmenistan and Kazakhstan (GBIF Secretariat, 2023).

*B. strataria* is a close relative of the Peppered moth *B. betularia*; the two species have similar wing size and shape, although the markings are quite different. As with *B. betularia*, melanism has been reported in *B. strataria*. At least two distinct melanic forms are reported: ab. *robinaria* in which the pale areas of the forewings are suffused with black scales and ab. *melanaria*, widely recorded in the Netherlands, which is uniformly black (West, 2005). Neither melanic variant has been common in Britain, even during industrial periods when melanic *B. betularia* increased in frequency. West (2005) suggests that ab. *robinaria* is caused by a dominant allele with viable heterozygotes and homozygotes; Ford (1967) reported that one of the melanic forms is lethal when homozygous, but it is unclear which he was referring to.

The larvae of *B. strataria* feed on the leaves of many deciduous trees and, despite the common name, are not oak specialists. Indeed, the German common name, Pappel-Dickleibspanner, refers to living on poplar trees (*Populus* spp.). At rest, the mature larvae grip twigs of the food plant with posterior claspers and prolegs, and extend their thin, lumpy body at a sharp angle. This ‘twig-like’ posture is thought to be an example of masquerade, where the larva is clearly visible to predators but misidentified; this contrasts to crypsis in which an individual is not detected (Skelhorn *et al.*, 2009). Attempts to test the masquerade hypothesis have given supportive, but not conclusive, support (Skelhorn *et al.*, 2009)

The genome sequence of *Biston strataria* was determined as part of the Darwin Tree of Life project. The complete genome sequence will aid research into the molecular basis of wing colour polymorphism and into adaptations to polyphagy, and will contribute to the growing set of resources for studying molecular evolution in the Lepidoptera.

## Genome sequence report

The genome was sequenced from one male *Biston strataria* (Figure 1) collected from Wytham Woods, Oxfordshire, UK



**Figure 1.** Photograph of the *Biston strataria* (ilBisStrt2) specimen used for genome sequencing.

(51.77, −1.34). A total of 48-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 one missing join and removed one haplotypic duplication, reducing the scaffold number by 4.35%.

The final assembly has a total length of 424.0 Mb in 21 sequence scaffolds with a scaffold N50 of 31.2 Mb (Table 1). The snailplot 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 (99.94%) of the assembly sequence was assigned to 16 chromosomal-level scaffolds, representing 15 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 by synteny to *Biston betularia* (GCA\_905404145.2) (Boyes *et al.*, 2022). 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 68.5 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.6% (single = 98.3%, duplicated = 0.3%), 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/722658>.

## Genome annotation report

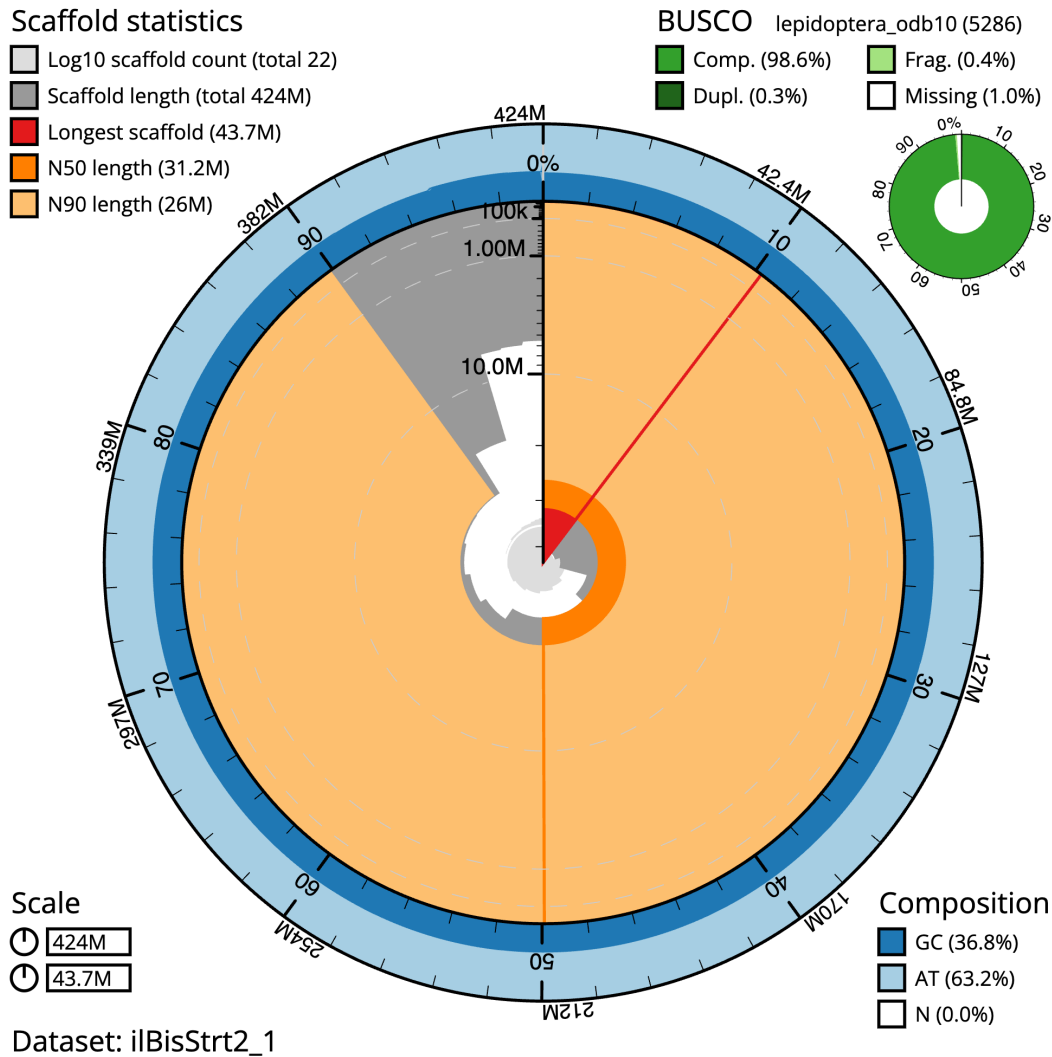
The *Biston strataria* genome assembly (GCA\_950106695.1) was annotated using the Ensembl rapid annotation pipeline

**Table 1. Genome data for *Biston strataria*, ilBisStrt2.1.**

Project accession data		
Assembly identifier	ilBisStrt2.1	
Species	<i>Biston strataria</i>	
Specimen	ilBisStrt2	
NCBI taxonomy ID	722658	
BioProject	PRJEB61133	
BioSample ID	SAMEA10107025	
Isolate information	ilBisStrt2, male: thorax (DNA and RNA sequencing) ilBisStrt1: head and thorax (Hi-C sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	68.5	≥ 50
k-mer completeness	100.0%	≥ 95%
BUSCO**	C:98.6%[S:98.3%,D:0.3%], F:0.4%,M:1.0%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.94%	≥ 95%
Sex chromosomes	Z	localised homologous pairs
Organelles	Mitochondrial genome: 15.61 kb	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11242112	
Hi-C Illumina	ERR11217105, ERR11217106	
PolyA RNA-Seq Illumina	ERR12245551	
Genome assembly		
Assembly accession	GCA_950106695.1	
Accession of alternate haplotype	GCA_950106675.1	
Span (Mb)	424.0	
Number of contigs	70	
Contig N50 length (Mb)	9.3	
Number of scaffolds	21	
Scaffold N50 length (Mb)	31.2	
Longest scaffold (Mb)	43.73	
Genome annotation		
Number of protein-coding genes	18,406	
Number of gene transcripts	18,552	

\* 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/ilBisStrt2\\_1/dataset/ilBisStrt2\\_1/busco](https://blobtoolkit.genomehubs.org/view/ilBisStrt2_1/dataset/ilBisStrt2_1/busco).



**Figure 2. Genome assembly of *Biston strataria*, ilBisStrt2.1: 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 424,056,544 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 (43,732,088 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (31,208,321 and 25,983,362 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/ilBisStrt2\\_1/dataset/ilBisStrt2\\_1/snail](https://blobtoolkit.genomehubs.org/view/ilBisStrt2_1/dataset/ilBisStrt2_1/snail).

(Table 1; [https://rapid.ensembl.org/Biston\\_stratarius\\_GCA\\_950106695.1/Info/Index](https://rapid.ensembl.org/Biston_stratarius_GCA_950106695.1/Info/Index)). The resulting annotation includes 18,552 transcribed mRNAs from 18,406 protein-coding genes.

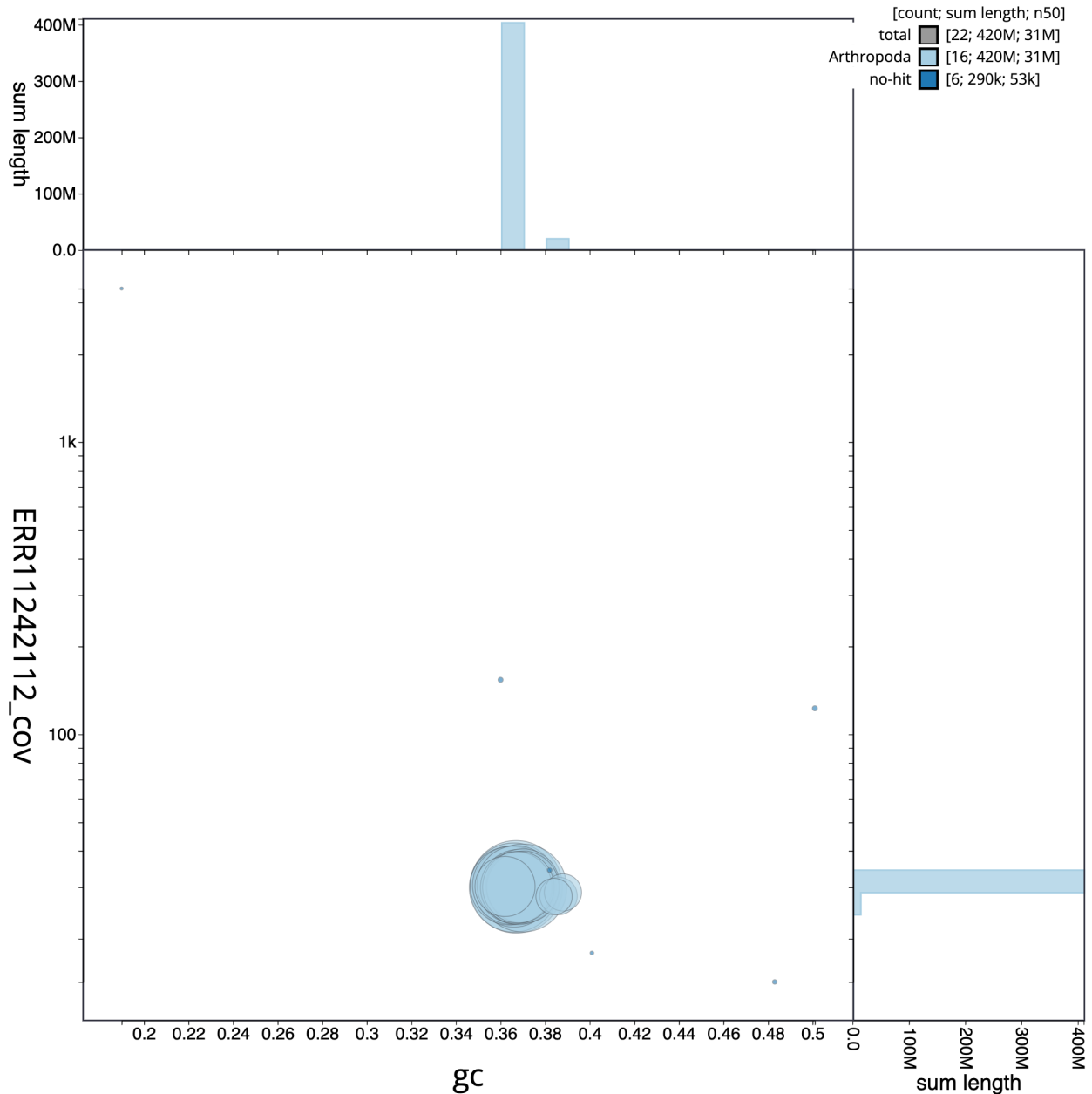
## Methods

### Sample acquisition and nucleic acid extraction

A male *Biston strataria* (specimen ID Ox001101, ToLID ilBisStrt2) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77,

longitude  $-1.34$ ) on 2021-03-31 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice. This specimen was used for DNA and RNA sequencing.

The specimen used for Hi-C sequencing (specimen ID NHMUK014043020, ToLID ilBisStrt1) was collected in a light trap from High Wycombe, Buckinghamshire, UK (latitude 51.63, longitude  $-0.74$ ) on 2021-03-04. The specimen

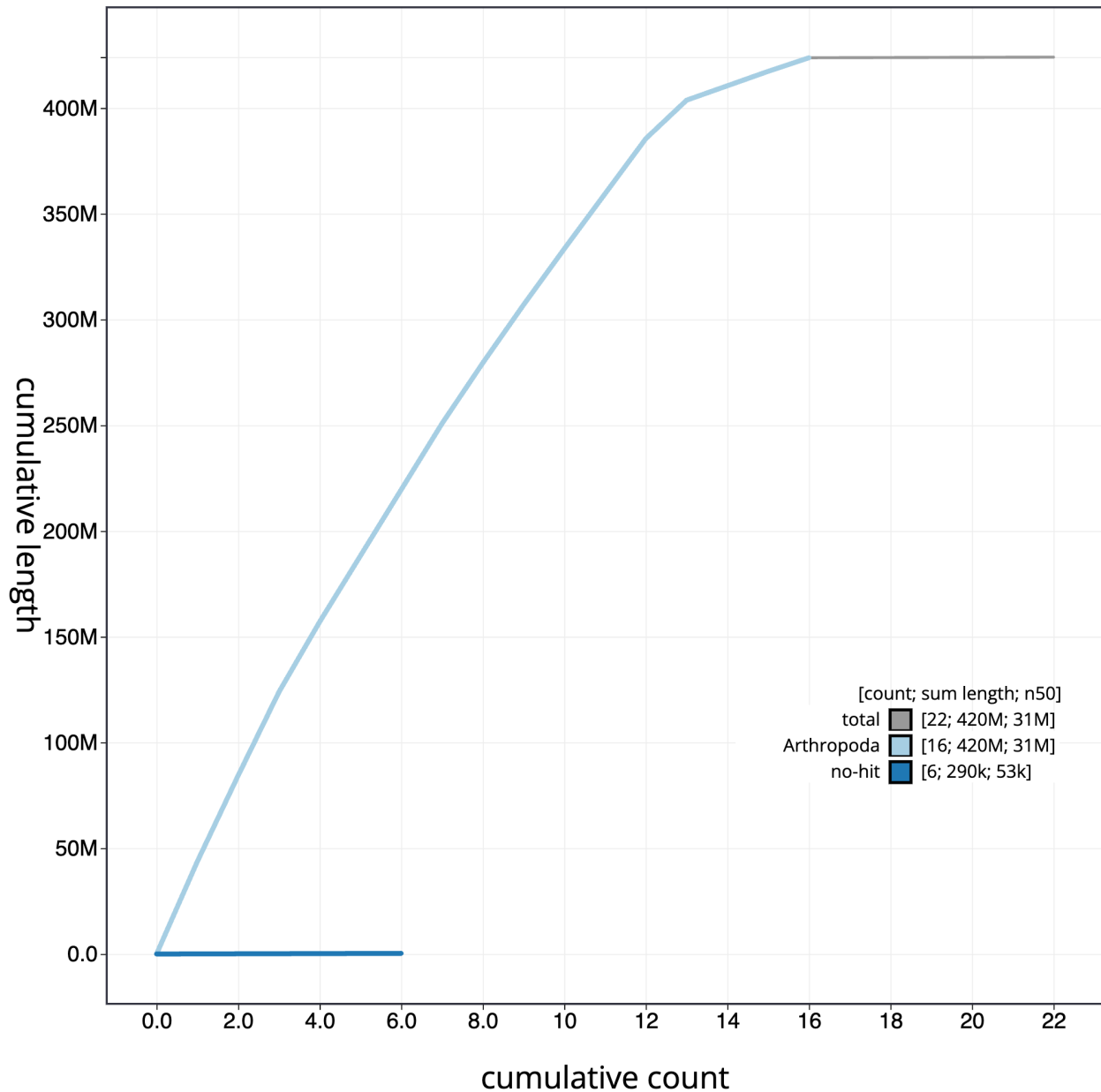


**Figure 3. Genome assembly of *Biston strataria*, ilBisStrt2.1: BlobToolKit GC-coverage plot.** Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/ilBisStrt2\\_1/dataset/ilBisStrt2\\_1/blob](https://blobtoolkit.genomehubs.org/view/ilBisStrt2_1/dataset/ilBisStrt2_1/blob).

was collected and identified by David Lees (Natural History Museum) and preserved by dry freezing at  $-80^{\circ}\text{C}$ .

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.

In sample preparation, the ilBisStrt2 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the thorax 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



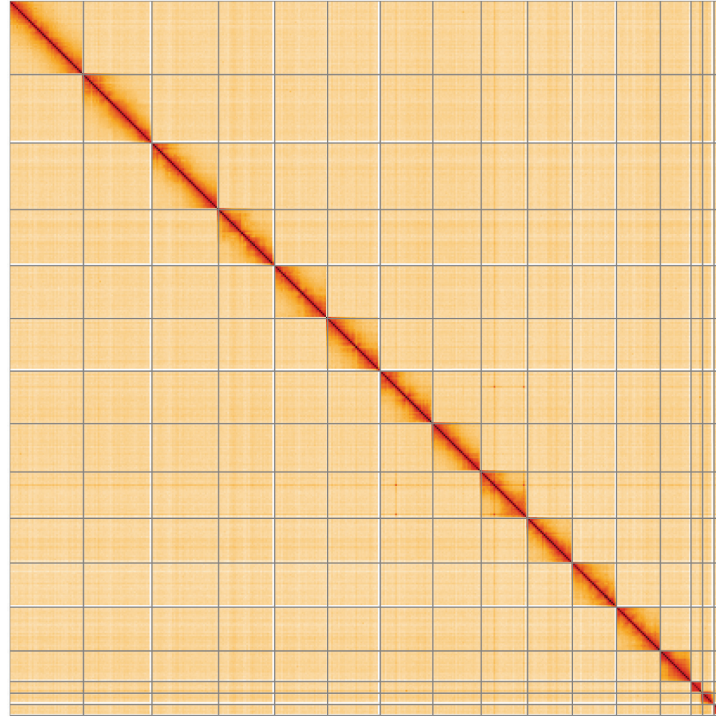
**Figure 4. Genome assembly of *Biston strataria*, ilBisStrt2.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscodegenes taxrule. An interactive version of this figure is available at [https://blobtoolkit.genomehubs.org/view/ilBisStrt2\\_1/dataset/ilBisStrt2\\_1/cumulative](https://blobtoolkit.genomehubs.org/view/ilBisStrt2_1/dataset/ilBisStrt2_1/cumulative).

with speed setting 31 (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.

RNA was extracted from remaining thorax tissue of ilBisStrt2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using





**Figure 5. Genome assembly of *Biston strataria*, ilBisStrt2.1: Hi-C contact map of the ilBisStrt2.1 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=H2uSqyMdS6ST2KSsBHf3Cg>.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Biston strataria*, ilBisStrt2.**

INSDC accession	Chromosome	Length (Mb)	GC%
OX467106.1	1	43.73	36.5
OX467107.1	2	40.6	36.5
OX467108.1	3	39.47	37.0
OX467109.1	4	33.24	36.5
OX467110.1	5	31.34	36.5
OX467111.1	6	31.21	36.5
OX467112.1	7	31.19	36.5
OX467113.1	8	28.64	37.0
OX467114.1	9	27.5	37.0
OX467115.1	10	26.51	37.0
OX467116.1	11	26.11	36.5
OX467117.1	12	25.98	37.0
OX467119.1	13	6.88	38.5
OX467120.1	14	6.75	39.0
OX467121.1	15	6.43	38.5
OX467118.1	Z	18.18	36.0
OX467122.1	MT	0.02	19.0

a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

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

### Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from head and thorax tissue of ilBisStrt1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000, 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 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.

#### Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Biston strataria* assembly (GCA\_950106695.1) in Ensembl Rapid Release.

#### Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission

of materials by a Darwin Tree of Life Partner is subject to the ‘Darwin Tree of Life Project Sampling Code of Practice’, which can be found in full on the Darwin Tree of Life website [here](#). By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

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

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

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

**Table 3. Software tools: versions and sources.**

Software tool	Version	Source
BlobToolKit	4.2.1	<a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>
BUSCO	5.3.2	<a href="https://gitlab.com/ezlab/busco">https://gitlab.com/ezlab/busco</a>
Hifiasm	0.16.1-r375	<a href="https://github.com/chhy123/hifiasm">https://github.com/chhy123/hifiasm</a>
HiGlass	1.11.6	<a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>
Merqury	MerquryFK	<a href="https://github.com/thegenemyers/MERQURY.FK">https://github.com/thegenemyers/MERQURY.FK</a>
MitoHiFi	3	<a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>
PretextView	0.2	<a href="https://github.com/wtsi-hpag/PretextView">https://github.com/wtsi-hpag/PretextView</a>
purge_dups	1.2.5	<a href="https://github.com/dfguan/purge_dups">https://github.com/dfguan/purge_dups</a>
sanger-tol/genomenote	v1.0	<a href="https://github.com/sanger-tol/genomenote">https://github.com/sanger-tol/genomenote</a>
sanger-tol/readmapping	1.1.0	<a href="https://github.com/sanger-tol/readmapping/tree/1.1.0">https://github.com/sanger-tol/readmapping/tree/1.1.0</a>
YaHS	1.2a.2	<a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>

## Data availability

European Nucleotide Archive: *Biston stratarius*. Accession number PRJEB61133; <https://identifiers.org/ena.embl/PRJEB61133> (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Biston strataria* 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 Natural History Museum Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.7139035>.

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 mitochondrial 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<sup>®</sup>3 for LI PacBio**. *protocols.io*. 2023. [Publisher Full Text](#)
- Bernt M, Donath A, Jühling F, et al.: **MITOS: Improved de novo metazoan mitochondrial genome annotation**. *Mol Phylogenet Evol*. 2013; **69**(2): 313–319. [PubMed Abstract](#) | [Publisher Full Text](#)
- Boyes D, Wright C; University of Oxford and Wytham Woods Genome Acquisition Lab, et al.: **The genome sequence of the peppered moth, *Biston betularia* Linnaeus, 1758 [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2022; **7**: 97. [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](#)
- 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**. *protocols.io*. 2023b. [Publisher Full Text](#)
- Di Tommaso P, Chatzou M, Floden EW, et al.: **Nextflow enables reproducible computational workflows**. *Nat Biotechnol*. 2017; **35**(4): 316–319. [PubMed Abstract](#) | [Publisher Full Text](#)
- do Amaral RJV, Bates A, Denton A, et al.: **Sanger Tree of Life RNA Extraction: Automated MagMax<sup>™</sup> mirVana**. *protocols.io*. 2023. [Publisher Full Text](#)
- Ford EB: *Moths*. London: New Naturalist, Collins. 1967.
- GBIF Secretariat: ***Biston strataria* (Hufnagel, 1767)**. *GBIF Backbone Taxonomy*. 2023; [Accessed 30 December 2023]. [Reference Source](#)
- Guan D, McCarthy SA, Wood J, et al.: **Identifying and removing haplotypic duplication in primary genome assemblies**. *Bioinformatics*. 2020; **36**(9): 2896–2898. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps**. 2022; [Accessed 19 October 2022]. [Reference Source](#)
- Howe K, Chow W, Collins J, et al.: **Significantly improving the quality of genome assemblies through curation**. *GigaScience*. Oxford University Press, 2021; **10**(1): gjaa153. [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](#)
- Manni M, Berkeley MR, Seppely M, et al.: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes**. *Mol Biol Evol*. 2021; **38**(10): 4647–4654. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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.: **Merquy: Reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biol*. 2020; **21**(1): 245. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Simão FA, Waterhouse RM, Ioannidis P, et al.: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs**. *Bioinformatics*. 2015; **31**(19): 3210–3212. [PubMed Abstract](#) | [Publisher Full Text](#)
- Skellhorn J, Rowland HM, Ruxton GD: **The evolution and ecology of masquerade**. *Biol J Linn Soc Lond*. 2009; **99**(1): 1–8. [Publisher Full Text](#)
- Strickland M, Cornwell C, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Manual SPRI**. *protocols.io*. 2023. [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].

[Reference Source](#)

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, Krashennikova 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 Md, 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](#)

Wellcome Sanger Institute: **The genome sequence of the Oak Beauty, *Biston strataria* (Hufnagel, 1767)**. European Nucleotide Archive, [dataset], accession number PRJEB61133. 2023.

West BK: ***Biston strataria* Hufn. (Lep.: Geometridae): Melanic forms in north-west Kent**. *Entomologists' Record*. 2005; **117**: 158.

[Reference Source](#)

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

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

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 01 August 2024

<https://doi.org/10.21956/wellcomeopenres.23054.r78834>

© 2024 Alioto T. 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.



**Tyler Scott Alioto** 

Fundacion Centro Nacional de Analisis Genomico, Barcelona, Catalonia, Spain

The genome note submitted by Boyes *et al.* reports the genome sequence of the Oak Beauty, *Biston strataria*. The assembly is chromosome-scale and of high quality, meeting the minimum standards recommended by the Earth Biogenome Project. All protocols are appropriate and well-documented. All data conforms with FAIR principles, with read data and assemblies being available in the ENA. Blobtoolkit figures are interactive. Additional QC data are provided at <https://tolqc.cog.sanger.ac.uk/>, supplementing the figures published in the data note.

**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:** genomics, genome assembly, gene prediction, annotation

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

Reviewer Report 21 May 2024

<https://doi.org/10.21956/wellcomeopenres.23054.r81629>

© 2024 Singh K. 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.



**Kumar Saurabh Singh** 

Wageningen University & Research, Wageningen, Gelderland, The Netherlands

The authors have assembled the genome of the male Oak Beauty moth, *Biston strataria* (Hufnagel, 1767), utilizing advanced genome sequencing technologies such as Pacific Biosciences single-molecule HiFi sequencing and chromosome conformation Hi-C data. They have deposited the raw datasets in the NCBI and included the accessions for accessing this data. While I found the manuscript comprehensive, I have a few comments on areas that were unclear to me or suggestions for further enhancing the clarity and quality of the manuscript.

1. The introduction discusses other moth species closely related to *B. strataria*. I believe including a phylogenetic tree illustrating the position of *B. strataria* would greatly enhance the manuscript, effectively conveying the rationale for sequencing this moth species.

2. I couldn't find details regarding the coverage and quantity of sequencing data generated for the Hi-C and RNAseq libraries. Additionally, annotating the Hi-C contact map with chromosome numbers would be helpful for readers.

3. Regarding the manual assembly correction, if the Hi-C map was utilized, could you provide a graphical representation indicating where corrections were applied within the contact map? Additionally, it was unclear how correcting a single missing join resulted in a significant reduction of 4.35% in scaffold count. I recommend providing more detailed information on this aspect to help readers understand the connection within the correction process.

3. In the annotation section, there seems to be no explanation regarding whether the RNAseq data was utilized as external evidence for predicting gene models. I recommend adding those details in that section.

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

Yes

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

Yes

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

Yes

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

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Bioinformatics, genomics, transcriptomics, metabolomics, data-integration

**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.**

---