Check for updates

DATA NOTE

# The genome sequence of the Broad-barred Knot-horn,

# *Acrobasis consociella* (Hübner, 1813)

[version 1; peer review: 2 approved]

Douglas Boyes[1+], Liam C. Crowley [2], Finley Hutchinson[3], Denise C. Wawman [2],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]Department of Biology, University of Oxford, Oxford, England, UK
[3]University of Exeter, Penryn, England, UK

+ Deceased author

## Abstract

We present a genome assembly from one female *Acrobasis consociella* (the Broad-barred Knot-horn; Arthropoda; Insecta; Lepidoptera; Pyralidae). The genome sequence is 598.4 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z and W sex chromosomes. The mitochondrial genome has also been assembled and is 15.22 kilobases in length.

## Keywords

Acrobasis consociella, Broad-barred Knot-horn, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

## Open Peer Review

### Approval Status ✓ ✓

|  | 1 | 2 |
|---|---|---|
| **version 1**<br>07 Aug 2024 | ✓<br>view | ✓<br>view |

1. **Frederico Hickmann** [iD], The Ohio State University College of Food Agricultural and Environmental Sciences (Ringgold ID: 155699), Wooster, USA

2. **Jerome H L Hui** [iD], The Chinese University of Hong Kong, Hong Kong, Hong Kong

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; **Crowley LC**: Investigation, Resources; **Hutchinson F**: Investigation, Resources; **Wawman DC**: Writing – Original Draft Preparation, Writing – Review & Editing;

**How to cite this article:** Boyes D, Crowley LC, Hutchinson F *et al.* **The genome sequence of the Broad-barred Knot-horn,** *Acrobasis consociella* **(Hübner, 1813) [version 1; peer review: 2 approved]** Wellcome Open Research 2024, **9**:429 https://doi.org/10.12688/wellcomeopenres.21553.1

**First published:** 07 Aug 2024, **9**:429 https://doi.org/10.12688/wellcomeopenres.21553.1

## 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; Pyraloidea; Pyralidae; Phycitinae; *Acrobasis*; *Acrobasis consociella* (Hübner, 1813) (NCBI:txid1100900).

## Background

*Acrobasis consociella*, the Broad-barred Knot-horn or Grey Oak Knot-horn, is a micromoth in the family Pyralidae. Its forewings are 9–11 mm long with an oblique first crossline, about one third of the way along the wing, that is white and edged with black on its distal side. Proximal to this crossline the wing is a light greyish colour and beyond it the wing has more coppery or mauve shades (Sterling *et al.*, 2023).

In Great Britain *Acrobasis consociella* is absent from Scotland and Northern England, but common further south. In Ireland records are more sparse and limited to coastal regions. The main flight season is from late May to September, although it has been recorded into November, when it is found in woodland, parks, gardens and along hedgerows (Sterling *et al.*, 2023). The larvae feed from September to June, gregariously in masses of leaves held together by silk, on oaks *Quercus* spp. (Sterling *et al.*, 2023; Visakorpi *et al.*, 2019)

It has been predicted that caterpillars eating oak leaves early in the season would have a negative effect on both tree growth and acorn production, and might lead to changes in concentrations of compounds such as polyphenols in the leaves. However, Visakorpi *et al.* (2019) found no evidence of this when *Acrobasis consociella* fed on pedunculate oak *Quercus robur* at different levels of abundance. Instead, the variation seen could be attributed to plant chemotype and varying environmental conditions (Visakorpi *et al.*, 2019).

We present a chromosomal-level whole genome sequence for a female *Acrobasis consociella*, based on one specimen collected in Wytham Woods, Oxfordshire as part of the Darwin Tree of Life Project.

## Genome sequence report

The genome was sequenced from a female *Acrobasis consociella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, -1.34). A total of 45-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 8 missing joins or mis-joins and removed one haplotypic duplication, reducing the scaffold number by 4.08%.

The final assembly has a total length of 598.4 Mb in 46 sequence scaffolds with a scaffold N50 of 20.4 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets



**Figure 1. Photograph of the *Acrobasis consociella* (ilAcrCons1) specimen used for genome sequencing.**

of scaffolds assigned to different phyla. Most (99.83%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 29 autosomes and the Z and W sex chromosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). The Z and W chromosomes were assigned based on read coverage statistics. 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 66.6 with $k$-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.8% (single = 98.4%, duplicated = 0.4%), 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/1100900.

## Methods

### Sample acquisition and nucleic acid extraction

A *Acrobasis consociella* (specimen ID Ox000597, ToLID ilAcrCons1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2020-07-05 using a light trap. The specimen was collected by Douglas Boyes (University Of Oxford) and identified by Douglas Boyes (University Of Oxford) and preserved on dry ice.

The specimen used for Hi-C sequencing (specimen ID Ox003032, ToLID ilAcrCons2) was collected in a light trap at the same location on 2022-07-22. The specimen was collected by Liam Crowley (University of Oxford) and Finley Hutchinson (University of Exeter), identified by Finley Hutchinson, and then preserved on dry ice.

**Table 1. Genome data for *Acrobasis consociella*, ilAcrCons1.1.**

| Project accession data | |
|---|---|
| Assembly identifier | ilAcrCons1.1 |
| Species | *Acrobasis consociella* |
| Specimen | ilAcrCons1 |
| NCBI taxonomy ID | 1100900 |
| BioProject | PRJEB65197 |
| BioSample ID | SAMEA7701461 |
| Isolate information | ilAcrCons1<br>ilAcrCons2 |

| Assembly metrics* | | Benchmark |
|---|---|---|
| Consensus quality (QV) | 66.6 | ≥50 |
| *k*-mer completeness | 100.0% | ≥95% |
| BUSCO** | C:98.8%[S:98.4%,D:0.4%], F:0.3%,M:0.9%,n:5,286 | C ≥95% |
| Percentage of assembly mapped to chromosomes | 99.83% | ≥95% |
| Sex chromosomes | WZ | *localised homologous pairs* |
| Organelles | Mitochondrial genome: 15.22 kb | *complete single alleles* |

| Raw data accessions | |
|---|---|
| PacificBiosciences Sequel IIe | ERR11867199, ERR11867198 |
| Hi-C Illumina | ERR11872557 |

| Genome assembly | |
|---|---|
| Assembly accession | GCA_963555685.1 |
| *Accession of alternate haplotype* | GCA_963555645.1 |
| Span (Mb) | 598.4 |
| Number of contigs | 57 |
| Contig N50 length (Mb) | 19.8 |
| Number of scaffolds | 46 |
| Scaffold N50 length (Mb) | 20.4 |
| Longest scaffold (Mb) | 40.98 |

* 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/CAUVWO01/dataset/CAUVWO01/busco.

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 ilAcrCons1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was sheared into an average fragment size
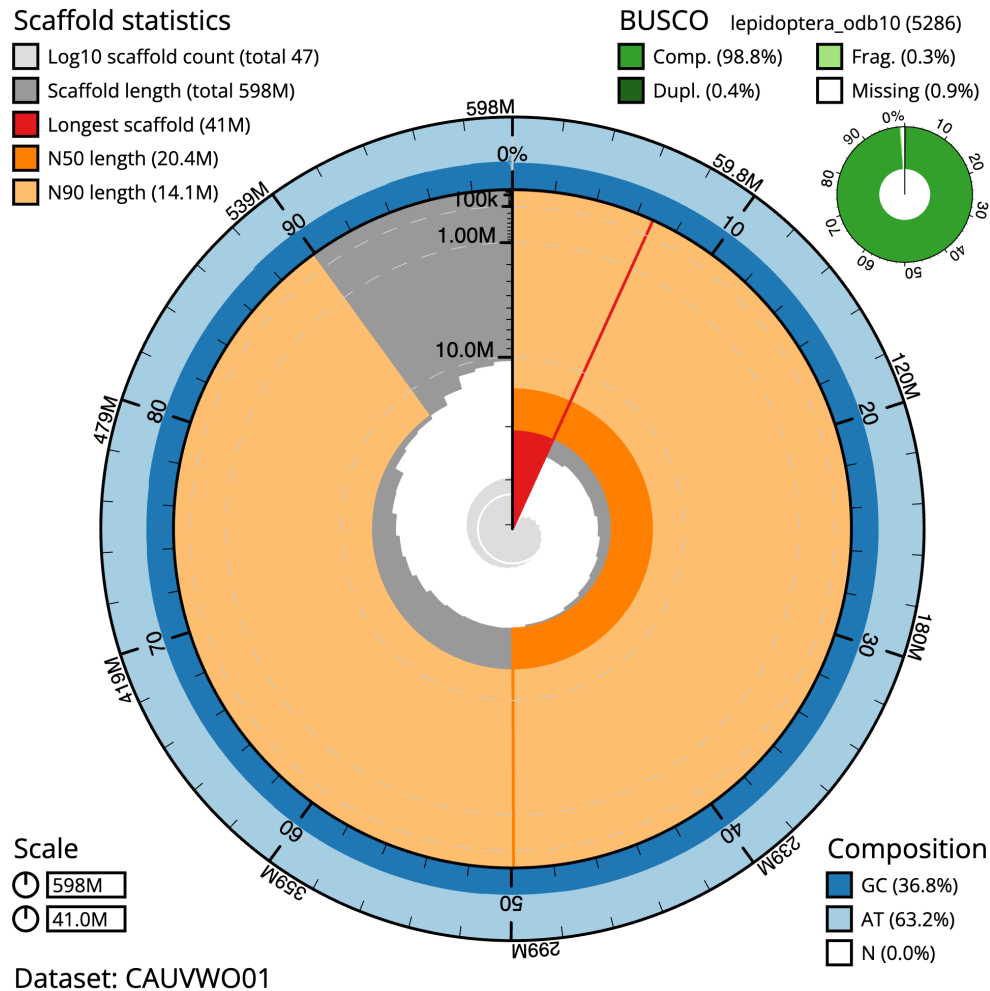
**Figure 2. Genome assembly of *Acrobasis consociella*, ilAcrCons1.1: metrics.** The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 598,408,520 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (40,980,569 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (20,377,658 and 14,055,802 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/CAUVWO01/dataset/CAUVWO01/snail.

of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

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

### Sequencing
Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences Sequel IIe instrument. Hi-C data were also generated from tissue of ilAcrCons2 using the Arima2 kit and sequenced on an 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
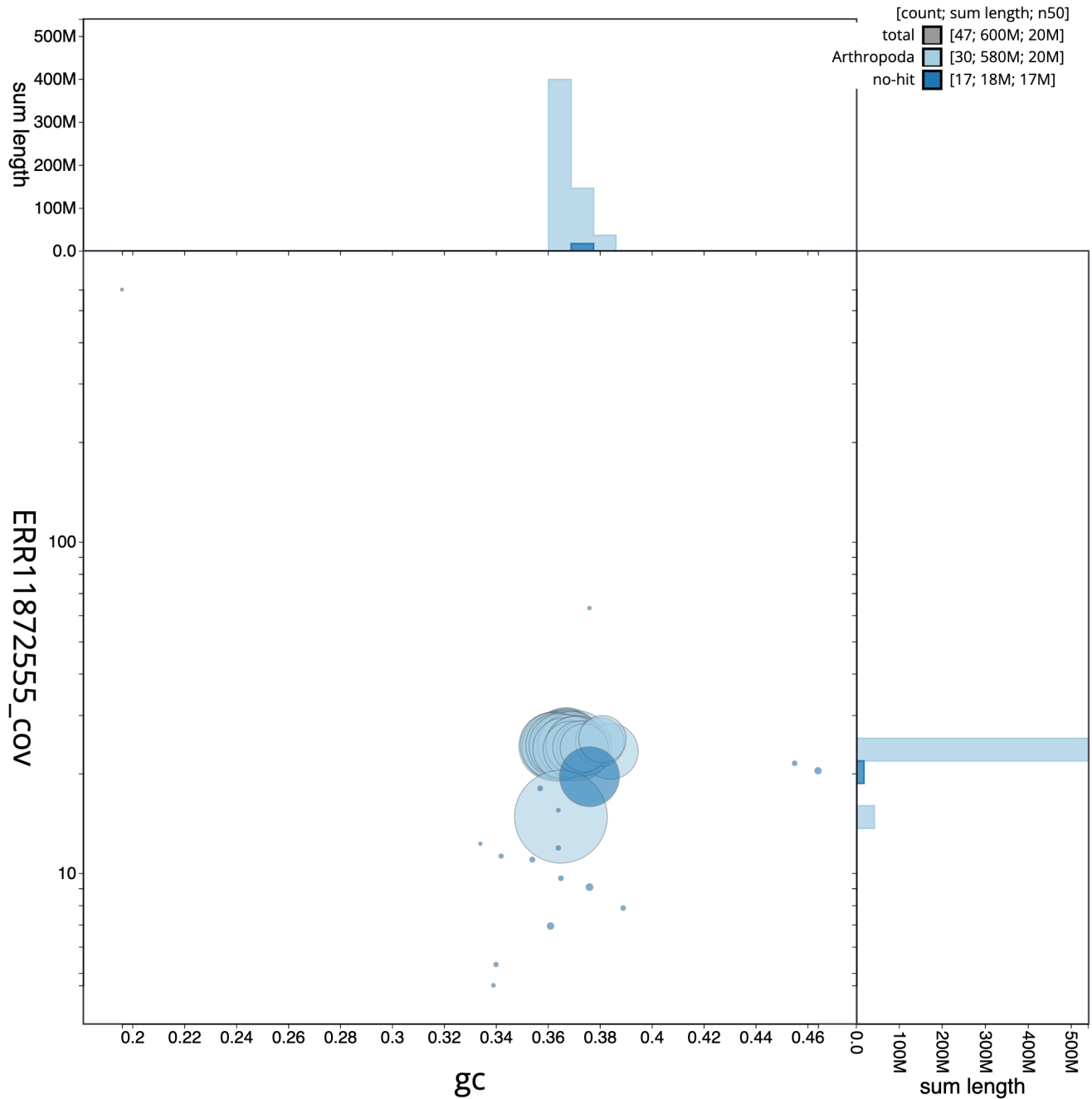
**Figure 3. Genome assembly of *Acrobasis consociella*, ilAcrCons1.1: BlobToolKit GC-coverage plot.** Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAUVWO01/dataset/CAUVWO01/blob.

scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) 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.
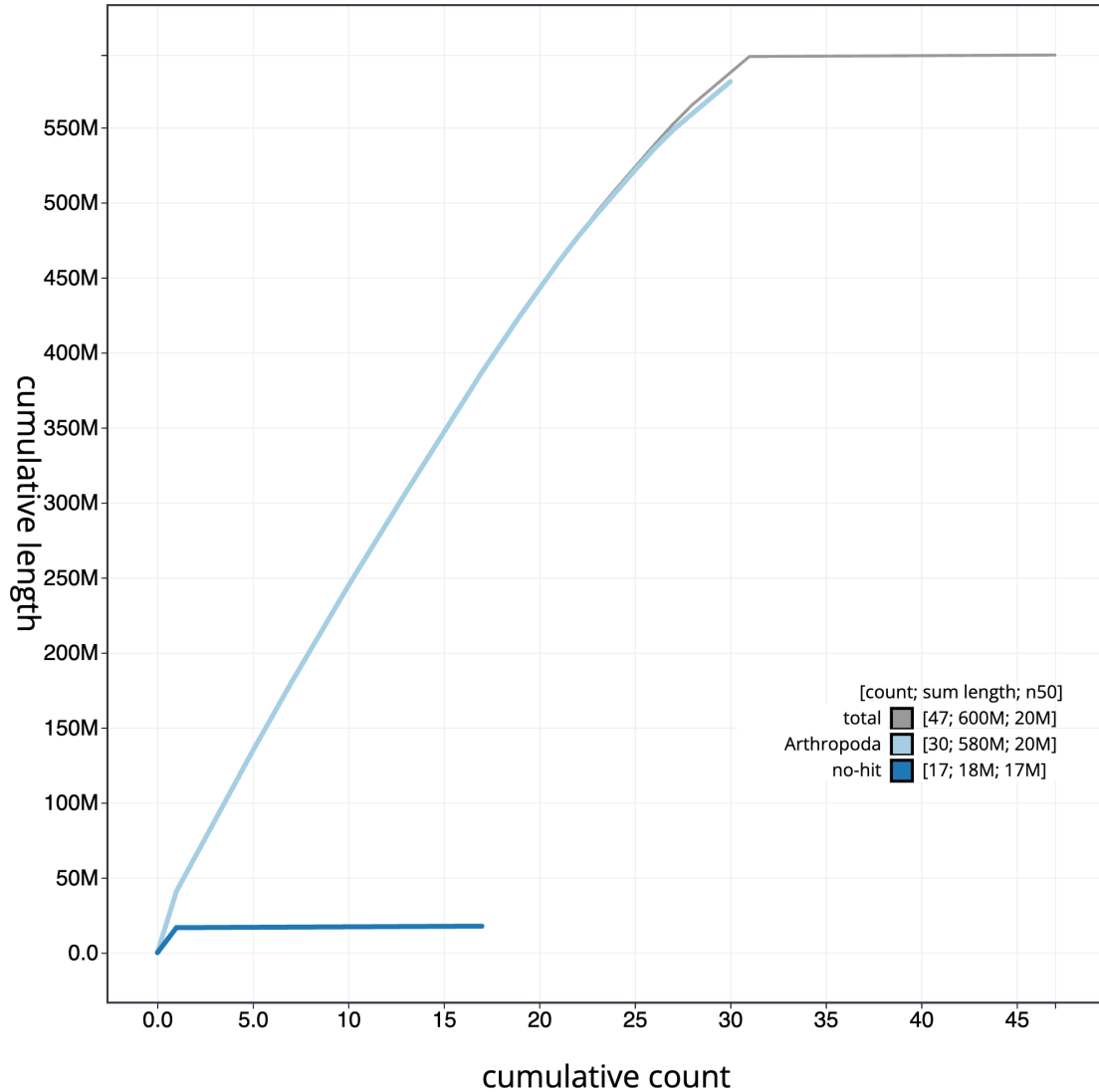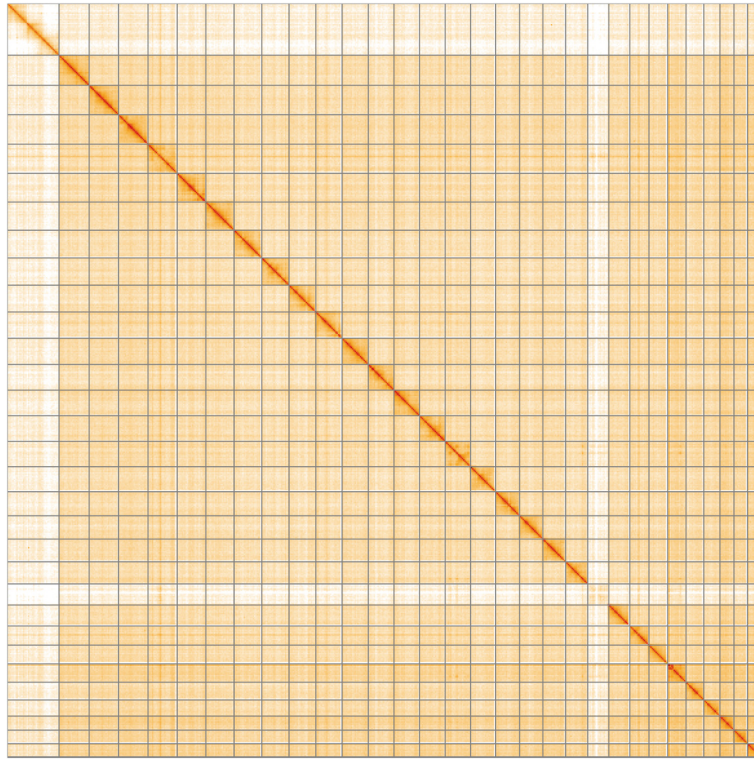
**Figure 4. Genome assembly of *Acrobasis consociella*, ilAcrCons1.1: BlobToolKit cumulative sequence plot.** The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/CAUVWO01/dataset/CAUVWO01/cumulative.

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.

### 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

**Figure 5. Genome assembly of *Acrobasis consociella*, ilAcrCons1.1: Hi-C contact map of the ilAcrCons1.1 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=DNmAKRq8QySpznTvuK-JGQ.

**Table 2. Chromosomal pseudomolecules in the genome assembly of *Acrobasis consociella*, ilAcrCons1.**

| INSDC accession | Chromosome | Length (Mb) | GC% |
|---|---|---|---|
| OY743105.1 | 1 | 23.87 | 36.5 |
| OY743106.1 | 2 | 23.35 | 36.5 |
| OY743107.1 | 3 | 23.33 | 36.5 |
| OY743108.1 | 4 | 23.18 | 37.0 |
| OY743109.1 | 5 | 22.75 | 36.5 |
| OY743110.1 | 6 | 22.34 | 36.5 |
| OY743111.1 | 7 | 21.96 | 36.0 |
| OY743112.1 | 8 | 21.64 | 36.5 |
| OY743113.1 | 9 | 21.24 | 36.5 |
| OY743114.1 | 10 | 20.82 | 36.5 |
| OY743115.1 | 11 | 20.79 | 36.5 |
| OY743116.1 | 12 | 20.38 | 36.0 |
| OY743117.1 | 13 | 20.32 | 36.5 |
| OY743118.1 | 14 | 20.28 | 36.5 |
| OY743119.1 | 15 | 20.12 | 37.0 |

| INSDC accession | Chromosome | Length (Mb) | GC% |
|---|---|---|---|
| OY743120.1 | 16 | 19.83 | 36.5 |
| OY743121.1 | 17 | 19.1 | 36.5 |
| OY743122.1 | 18 | 18.46 | 37.0 |
| OY743123.1 | 19 | 18.07 | 37.0 |
| OY743124.1 | 20 | 17.51 | 36.5 |
| OY743126.1 | 21 | 16.61 | 37.0 |
| OY743127.1 | 22 | 15.2 | 37.0 |
| OY743128.1 | 23 | 14.82 | 37.0 |
| OY743129.1 | 24 | 14.58 | 38.5 |
| OY743130.1 | 25 | 14.06 | 37.5 |
| OY743131.1 | 26 | 12.8 | 37.0 |
| OY743132.1 | 27 | 11.14 | 38.0 |
| OY743133.1 | 28 | 10.79 | 37.5 |
| OY743134.1 | 29 | 10.35 | 38.0 |
| OY743125.1 | W | 16.74 | 37.5 |
| OY743104.1 | Z | 40.98 | 36.5 |
| OY743135.1 | MT | 0.02 | 20.0 |

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

| Software tool | Version | Source |
|---|---|---|
| BlobToolKit | 4.2.1 | https://github.com/blobtoolkit/blobtoolkit |
| BUSCO | 5.3.2 | https://gitlab.com/ezlab/busco |
| Hifiasm | 0.19.5-r587) | https://github.com/chhylp123/hifiasm |
| HiGlass | 1.11.6 | https://github.com/higlass/higlass |
| Merqury | MerquryFK | https://github.com/thegenemyers/MERQURY.FK |
| MitoHiFi | 3 | https://github.com/marcelauliano/MitoHiFi |
| PretextView | 0.2 | https://github.com/wtsi-hpag/PretextView |
| purge_dups | 1.2.5 | https://github.com/dfguan/purge_dups |
| sanger-tol/genomenote | v1.0 | https://github.com/sanger-tol/genomenote |
| sanger-tol/readmapping | 1.1.0 | https://github.com/sanger-tol/readmapping/tree/1.1.0 |
| YaHS | 1.2a.2 | https://github.com/c-zhou/yahs |

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: *Acrobasis consociella* (broad-barred knot-horn). Accession number PRJEB65197; https://identifiers.org/ena.embl/PRJEB65197 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Acrobasis consociella* 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. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

## Author information
Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.7125292.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

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

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

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

## References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

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**

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

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

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

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**

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

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

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

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

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

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**

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**

Sheerin E, Sampaio F, Oatley G, *et al.*: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.1.** *protocols.io.* 2023.
**Publisher 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**

Sterling P, Parsons M, Lewington R: **Field guide to the micro-moths of Great Britain and Ireland, Second Edition.** London: Bloomsbury Publishing, 2023.
**Reference Source**

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.
**Publisher Full Text**

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
**Publisher Full Text**

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
**Publisher Full Text**

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

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
**Publisher Full Text**

Visakorpi K, Riutta T, Martínez-Bauer AE, *et al.*: **Insect community structure covaries with host plant chemistry but is not affected by prior herbivory.** *Ecology.* 2019; **100**(8): e02739.
**PubMed Abstract** | **Publisher Full Text**

Wellcome Sanger Institute: **The genome sequence of the Broad-barred Knot-horn, *Acrobasis consociella* (Hübner, 1813).** European Nucleotide Archive, [dataset], accession number PRJEB65197, 2023.

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 06 December 2024

https://doi.org/10.21956/wellcomeopenres.23824.r112675

✓    **Jerome H L Hui** (iD)

The Chinese University of Hong Kong, Hong Kong, Hong Kong

In this data note, Boyes and colleagues reported the genome sequence of a female moth *Acrobasis consociella* (Hübner, 1813). According to the UKMoth, *A. consociella* is a "reasonably common species distributed from southern England northwards into southern Scotland and west into Ireland". Prior to this report, molecular data available for this species are rather limited (which are mainly mitochondrial cytochrome oxidase subunit 1 (COI) gene sequences deposited to the NCBI database). The newly generated genomic resource of this species is important and useful for further studies, including but not limited to understanding the genomics, ecology, impact of climate change, insect-plant interaction, identification of cryptic species, reveal of population structure, to understanding their evolution with other insects.

This genome resource is excellent from the summary statistics, with high BUSCO number scores, high sequence continuity (scaffold N50), and majority of sequences contained on the 29 pseudochromosomes (plus sex chromosomes and mitochondrion). All in all, this is a valuable contribution.

**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, evolution, invertebrates

**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 29 October 2024

https://doi.org/10.21956/wellcomeopenres.23824.r104475

✔️ **Frederico Hickmann** 🆔
   The Ohio State University College of Food Agricultural and Environmental Sciences (Ringgold ID: 155699), Wooster, Ohio, USA

This manuscript provides a comprehensive account of the complete genome sequencing, assembly, and annotation of *Acrobasis consociella*, a species of the Pyralidae family indigenous to Great Britain. The work presented is meticulous and well-structured, employing appropriate genome assembly tools relevant to current and forthcoming research endeavors. The assembly corresponds with the standard chromosome count typically observed in Pyralidae moths. Moreover, the initiative to disseminate this data openly is of considerable importance to the scientific community.

**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:* My area is agricultural entomology, focusing on molecular ecology, genomics, and pests' off-season persistence.

**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.**