DATA NOTE

The genome sequence of the Dingy Dowd, *Blastobasis adustella* (Walsingham, 1894) [version 1; peer review: awaiting peer review]

Douglas Boyes[1+], Finley Hutchinson[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]University of Exeter, Penryn, England, UK

[+] Deceased author

v1 **First published:** 19 Sep 2023, **8**:407
https://doi.org/10.12688/wellcomeopenres.19900.1
Latest published: 19 Sep 2023, **8**:407
https://doi.org/10.12688/wellcomeopenres.19900.1

Abstract

We present a genome assembly from an individual female *Blastobasis adustella* (the Dingy Dowd; Arthropoda; Insecta; Lepidoptera; Blastobasidae). The genome sequence is 557.4 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.29 kilobases in length. Gene annotation of this assembly on Ensembl identified 9,783 protein coding genes.

Keywords

Blastobasis adustella, Dingy Dowd, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

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.

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

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

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) and the Darwin Tree of Life Discretionary Award (218328).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Boyes D, Hutchinson F, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Dingy Dowd, *Blastobasis adustella* (Walsingham, 1894) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:407 https://doi.org/10.12688/wellcomeopenres.19900.1

First published: 19 Sep 2023, **8**:407 https://doi.org/10.12688/wellcomeopenres.19900.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Gelechioidea; Blastobasidae; *Blastobasis*; *Blastobasis adustella* (Walsingham, 1894) (NCBI:txid1869501).

Background

Five species of *Blastobasis* Zeller, 1855 have been recorded in the British Isles, all of which have been introduced here. The members of this genus have a characteristic resting posture, with their wings held overlapping and wrapped around the abdomen (Sterling *et al.*, 2012). *Blastobasis adustella is* thought to originate from Madeira (De Prins *et al.*, 2009), and is one of the two *Blastobasis* species that have spread the fastest here, now occurring across most of Britain and Ireland, although in Ireland it is recorded mainly in coastal regions.

It can be separated from most of the other British species by its dark ground colour and presence of an oblique white streak across each forewing at about one quarter of the wing length, forming an 'inverted V' shape. However, the rarer *B. vittata* Wollaston, 1858 is very similar, and examination of the genitalia may be required to separate some individuals, although the latter species is generally smaller, slimmer and more warmly coloured. Similarly, *B. maroccanella* Amsel, 1952 may appear very similar to both species, but is no longer considered to be a part of the UK fauna after British specimens were re-identified as *vittata* (Dickson *et al.*, 2022). Records of *B. adustella* in the UK were initially referred to as *B. lignea* Walsingham, 1894 but were misidentifications, as *B. lignea* is now understood to be a junior synonym of *B. vittata*.

B. adustella is highly polyphagous, with the larvae feeding in spinnings on a wide range of plants and plant material, but seeming to show a preference for the leaves of yew (*Taxus baccata)*. The purplish larvae occur from September to June and are very similar to those of *B. lacticolella* Rebel, 1940, but can be separated from that species by the prothoracic plate being much darker than the head (Smart, 2021). They pupate in June in a silken cocoon amongst leaf litter or on the foodplant. The single-brooded adults are readily attracted to light and are on the wing from mid-May to December, peaking in June and again in September (Dickson, 2018).

We present a chromosomally complete genome sequence for *Blastobasis adustella*, based on one female specimen from Wytham Woods, Oxfordshire, as part of the Darwin Tree of Life Project. This project is a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland, with a view to further understanding of their evolutionary relationships and biodiversity.

Genome sequence report

The genome was sequenced from one female *Blastobasis adustella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 50-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 73-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 9 missing joins or misjoins and removed 3 haplotypic duplications, reducing the scaffold number by 13.33%, and increasing the scaffold N50 by 5.97%.

The final assembly has a total length of 557.4 Mb in 39 sequence scaffolds with a scaffold N50 of 19.4 Mb (Table 1). Most (98.67%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. It is assumed that the karyotype for this sample is ZO as no evidence was found to support the existence of a W chromosome in the assembly. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). Several scaffolds consisting largely of centromeric repeats that, by Hi-C data, associate with Chromosome 1, but cannot be accurately placed. These were submitted as unlocalised. 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 58 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.0% (single = 97.1%, duplicated = 0.9%), using the lepidoptera_odb10 reference set (n = 5,286).

Figure 1. Photograph of the *Blastobasis adustella* (ilBlaAdus2) specimen used for genome sequencing.

Table 1. Genome data for *Blastobasis adustella*, ilBlaAdus2.1.

Project accession data		
Assembly identifier	ilBlaAdus2.1	
Species	*Blastobasis adustella*	
Specimen	ilBlaAdus2	
NCBI taxonomy ID	1869501	
BioProject	PRJEB44983	
BioSample ID	SAMEA7520179	
Isolate information	ilBlaAdus2, female: whole organism (DNA sequencing) ilBlaAdus1: whole organism (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	58	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.0%[S:97.1%,D:0.9%], F:0.4%,M:1.6%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	98.67%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6548409	
10X Genomics Illumina	ERR6054745, ERR6054746, ERR6054747, ERR6054748	
Hi-C Illumina	ERR6054749	
Genome assembly		
Assembly accession	GCA_907269095.1	
Accession of alternate haplotype	GCA_907269055.1	
Span (Mb)	557.4	
Number of contigs	50	
Contig N50 length (Mb)	18.0	
Number of scaffolds	39	
Scaffold N50 length (Mb)	19.4	
Longest scaffold (Mb)	33.7	
Genome annotation		
Number of protein-coding genes	9,783	
Number of non-coding genes	1,000	
Number of gene transcripts	17,171	

* 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/ilBlaAdus2.1/dataset/CAJSME01.1/busco.

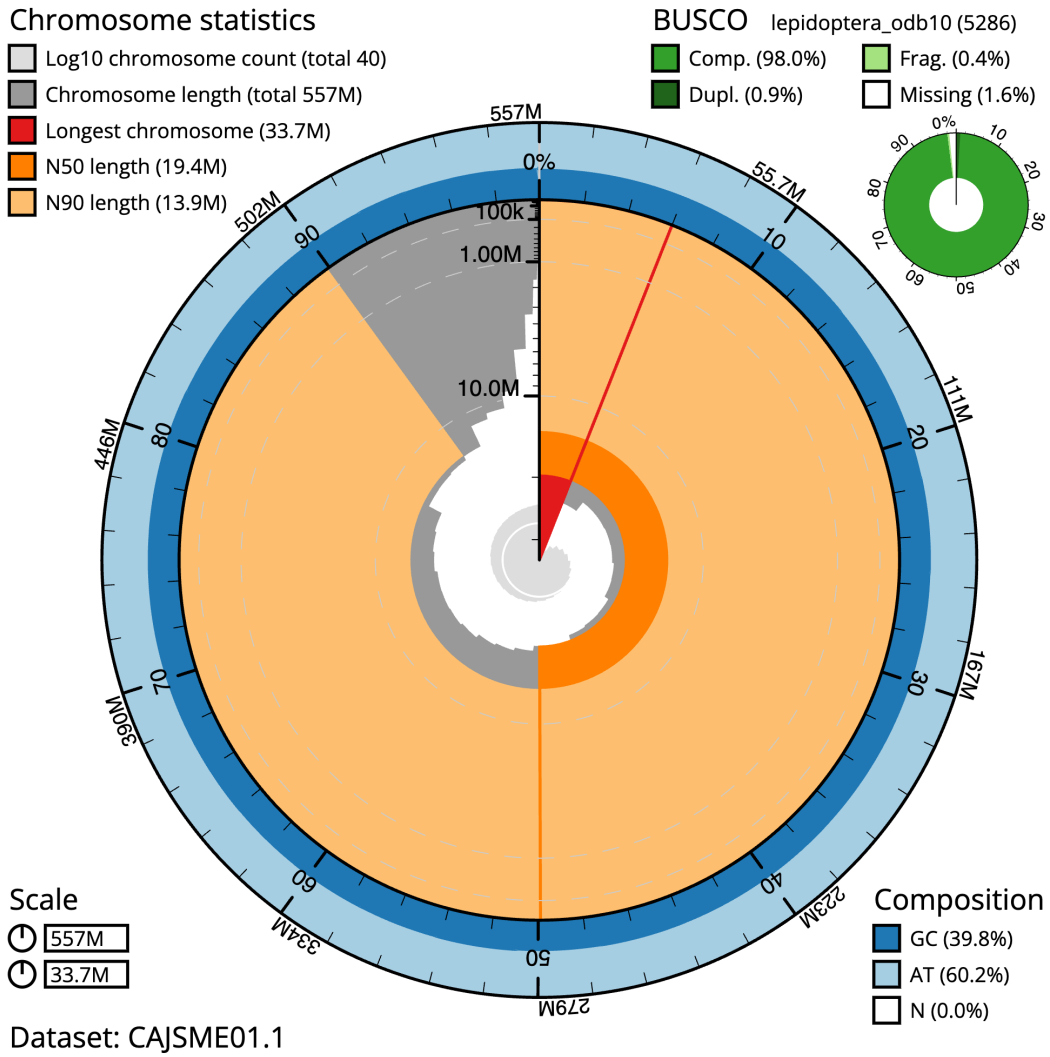

Figure 2. Genome assembly of *Blastobasis adustella*, ilBlaAdus2.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 557,440,016 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (33,699,264 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (19,379,192 and 13,890,508 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/ilBlaAdus2.1/dataset/CAJSME01.1/snail.

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/1869501.

Genome annotation report

The *Blastobasis adustella* genome assembly (GCA_907269095.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Blastobasis_adustella_GCA_907269095.1/Info/Index). The resulting annotation includes 17,171 transcribed mRNAs from 9,783 protein-coding and 1,000 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

The *Blastobasis adustella* specimen used for DNA sequencing was specimen ID Ox000209, ToLID ilBlaAdus2, while the specimen used for Hi-C data was specimen ID Ox000210, ToLID ilBlaAdus1. Both specimens were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2019-08-24 using a light trap. The specimens were collected and identified by Douglas Boyes (University of Oxford). They were preserved on dry ice.

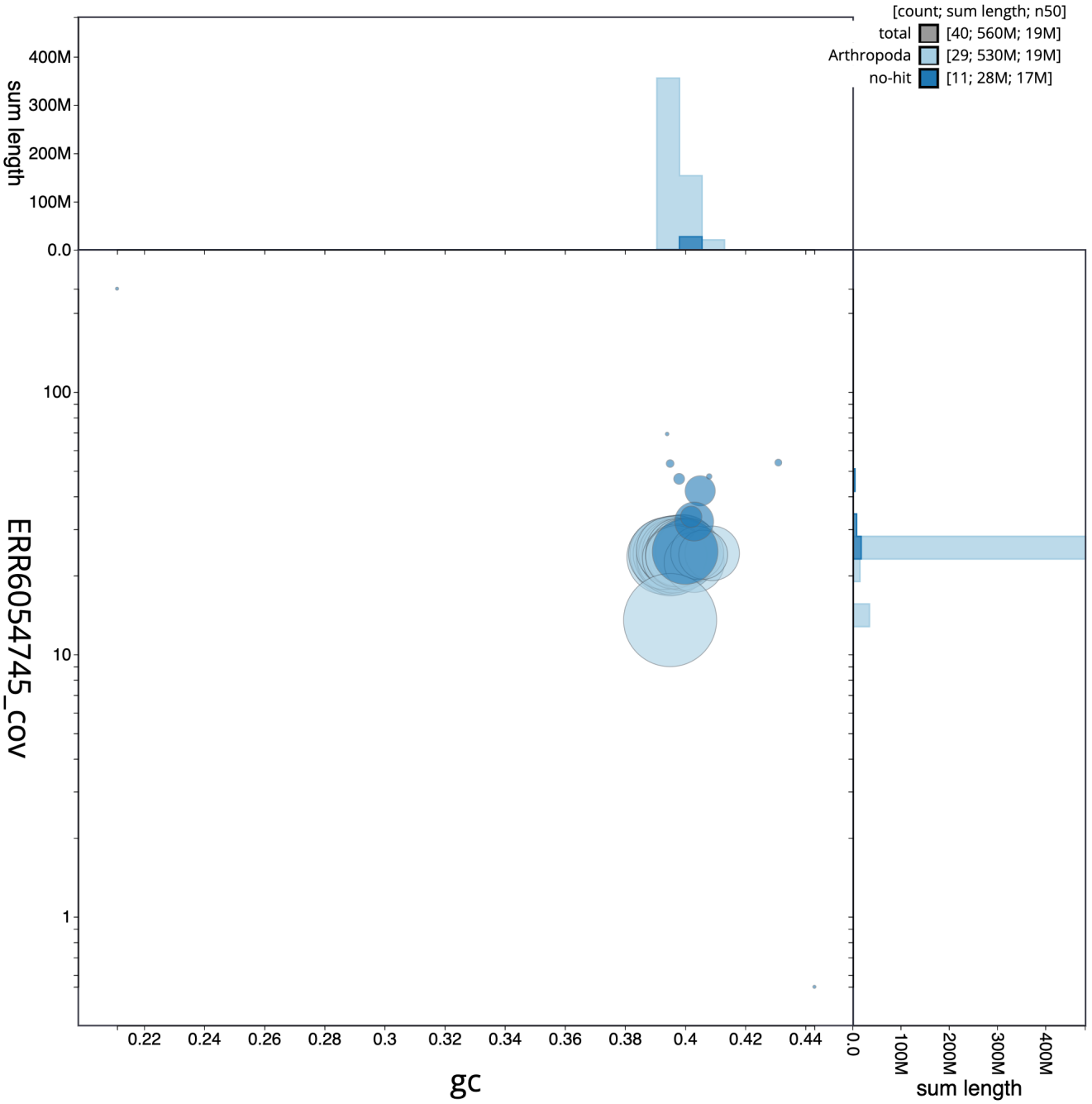

Figure 3. Genome assembly of *Blastobasis adustella*, ilBlaAdus2.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilBlaAdus2.1/dataset/CAJSME01.1/blob.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilBlaAdus2 sample was weighed and dissected on dry ice. Tissue from the whole organism was disrupted using a Nippi Powermasher fitted with a BioMasher pestle. 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

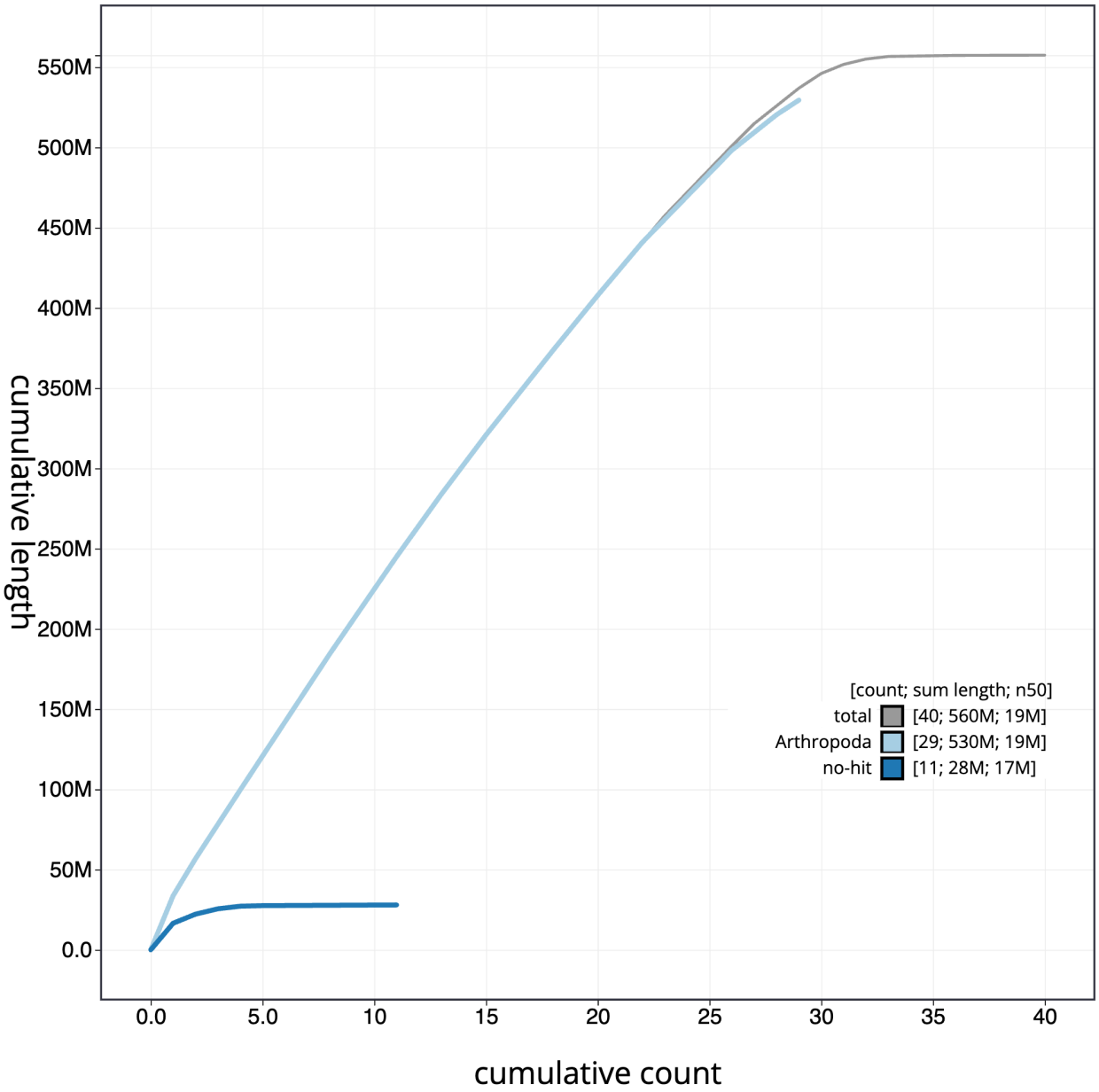

Figure 4. Genome assembly of *Blastobasis adustella*, ilBlaAdus2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilBlaAdus2.1/dataset/CAJSME01.1/cumulative.

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.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and HiSeq X Ten (10X) instruments. Hi-C data were also generated from whole organism tissue of ilBlaAdus2 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

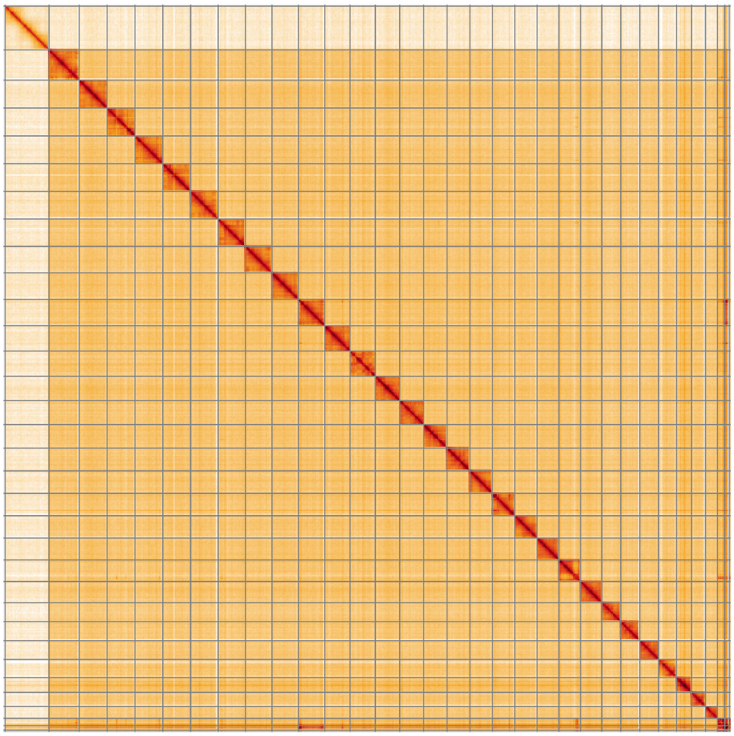

Figure 5. Genome assembly of *Blastobasis adustella*, ilBlaAdus2.1: Hi-C contact map of the ilBlaAdus2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=cg4RwO9BSmyF6YJgJPikLg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Blastobasis adustella*, ilBlaAdus2.

INSDC accession	Chromosome	Size (Mb)	GC%
OU026124.1	1	20.13	39.6
OU026115.1	2	23.04	39.5
OU026116.1	3	21.41	39.9
OU026117.1	4	21.4	39.3
OU026118.1	5	21.34	39.4
OU026119.1	6	21.3	39.9
OU026120.1	7	21.04	39.5
OU026121.1	8	21	39.6
OU026122.1	9	20.35	39.6
OU026123.1	10	20.3	39.7
OU026125.1	11	19.46	39.7
OU026126.1	12	19.38	39.3
OU026127.1	13	18.69	40
OU026128.1	14	18.29	39.6
OU026129.1	15	18.02	39.7
OU026130.1	16	17.49	39.5
OU026131.1	17	17.24	39.8
OU026132.1	18	17.14	39.9
OU026133.1	19	17.11	40
OU026134.1	20	16.75	39.9
OU026135.1	21	16.57	40
OU026136.1	22	16.34	40.1
OU026137.1	23	14.53	39.6
OU026138.1	24	14.5	39.7
OU026139.1	25	14.38	39.7
OU026140.1	26	13.89	40.3
OU026141.1	27	11.37	40.9
OU026142.1	28	10.75	40.4
OU026143.1	29	9.13	40.6
OU026114.1	Z	33.7	39.5
OU026144.1	MT	0.02	20.8
-	-	11.39	40.3

purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) 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 Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Blastobasis adustella* assembly (GCA_907269095.1). Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

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

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	0.12	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	1	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

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: *Blastobasis adustella* (dingy dowd). Accession number PRJEB44983; https://identifiers.org/ena.embl/PRJEB44983. (Wellcome Sanger Institute, 2021)

The genome sequence is released openly for reuse. The *Blastobasis adustella* 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

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

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

De Prins W, De Prins G, Larsen K: ***Blastobasis adustella* (Lepidoptera: Coleophoridae, Blastobasinae), new to the Belgian list.** *Phegea.* 2009; **37**(3): 111–118.
Reference Source

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

Dickson RJ: In: **Blastobasidae** Langmaid, J.R, Palmer, S.M and Young, M. R.(eds.) *A Field Guide to the Smaller Moths of Great Britain and Ireland.* Reading: The British Entomological and Natural History Society. 2018; 194–195.

Dickson R, Langmaid JR, Thirlwell I: **British *Blastobasis maroccanella* Amsel, 1952 are *B. vittata* (Wollaston, 1858).** *The Entomologist's Record and Journal of Variation.* 2022; **134**(2): 116–120.

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

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

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**

Smart B: **Micro-moth Field Tips: A Guide to Finding the Early Stages in Lancashire and Cheshire (2nd Edition).** Rishton: Lancashire & Cheshire Fauna Society. 2021.

Sterling P, Parsons M, Lewington R: **Field Guide to the Micro Moths of Great Britain and Ireland.** Gillingham, Dorset: British Wildlife Publishing, 2012.

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.
Reference Source

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**

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the Dingy Dowd, *Blastobasis adustella* (Walsingham, 1894).** European Nucleotide Archive. [dataset], accession number PRJEB44983, 2021.