DATA NOTE

The genome sequence of the London Dowd, *Blastobasis lacticolella* (Rebel, 1939) [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:** 17 Aug 2023, **8**:352
https://doi.org/10.12688/wellcomeopenres.19839.1
Latest published: 17 Aug 2023, **8**:352
https://doi.org/10.12688/wellcomeopenres.19839.1

Abstract

We present a genome assembly from an individual male *Blastobasis lacticolella* (the London Dowd; Arthropoda; Insecta; Lepidoptera; Blastobasidae). The genome sequence is 577.1 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 15.63 kilobases in length. Gene annotation of this assembly on Ensembl identified 10,302 protein coding genes.

Keywords

Blastobasis lacticolella, London 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 London Dowd, *Blastobasis lacticolella* (Rebel, 1939) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:352 https://doi.org/10.12688/wellcomeopenres.19839.1

First published: 17 Aug 2023, **8**:352 https://doi.org/10.12688/wellcomeopenres.19839.1

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 lacticolella* (Rebel, 1939) (NCBI:txid2561016).

Background

Thirty-three species of the genus *Blastobasis* are known in Europe, and most of which are endemic to Madeira (De Prins *et al.*, 2009). Five species of *Blastobasis* Zeller, 1855 have been recorded in Britain and Ireland, all of which have been introduced here. All species in the genus rest with their wings held overlapping and wrapped around the abdomen, giving them a distinctive resting posture (Sterling *et al.*, 2012). *Blastobasis lacticolella* was introduced to the west of mainland Europe accidentally, and there are now records from the Netherlands, Britain and Ireland. Since its introduction to the UK, it has spread rapidly and it is now found across most of Britain, but there are few records from much of Ireland, where it occurs mainly in coastal areas (GBIF Secretariat, 2022).

Blastobasis lacticolella is the largest of the *Blastobasis* species recorded in the UK, reaching 11 mm. The species is straw coloured, with varying degrees of brownish shading; lighter individuals should be instantly recognisable, while at the darker end of the spectrum they could be confused with light examples of *B. vittata* Wollaston, 1858, although the larger size and broader wings of *lacticolella* should help to distinguish it. *B. lacticolella* is further characterised by a dark streak or spot on the dorsum at 1/3, and two more dark spots alongside each other at around 2/3 (Sterling *et al.*, 2012).

Like its relative *B. adustella* Walsingham, 1894, *B. lacticolella* larvae are highly polyphagous, feeding in silk tubes or spinnings on a variety of seeds and fruits as well as moss, dried leaves, oak galls and dead insects (Smart, 2021). They can be distinguished from the former species by their prothoracic plate, which is orange or brown as opposed to black. Larvae are active from June to August and again from September to May, with pupation occurring in June and August and the adults flying from mid-May through to December (Dickson, 2018).

We present a chromosomally complete genome sequence for *Blastobasis lacticolella*, based on one male 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.

Genome sequence report

The genome was sequenced from one male *Blastobasis lacticolella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.33). A total of 32-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 63-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 87 missing joins or mis-joins and removed 32 haplotypic duplications, reducing the assembly length by 3.72% and the scaffold number by 50.56%, and increasing the scaffold N50 by 10.59%.

Figure 1. Photograph of the *Blastobasis lacticolella* (ilBlaLact1) specimen used for genome sequencing.

The final assembly has a total length of 577.1 Mb in 44 sequence scaffolds with a scaffold N50 of 20.5 Mb (Table 1). Most (99.91%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 2–Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 58 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 97.7% (single = 96.7%, duplicated = 0.9%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at https://links.tol.sanger.ac.uk/species/2561016.

Table 1. Genome data for *Blastobasis lacticolella*, ilBlaLact1.1.

Project accession data		
Assembly identifier	ilBlaLact1.1	
Species	*Blastobasis lacticolella*	
Specimen	ilBlaLact1	
NCBI taxonomy ID	2561016	
BioProject	PRJEB42117	
BioSample ID	SAMEA7519826	
Isolate information	ilBlaLact1, male: whole organism (DNA sequencing and Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	58	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:97.7%[S:96.7%,D:0.9%], F:0.6%,M:1.7%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.91%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6548403	
10X Genomics Illumina	ERR6002598, ERR6002601, ERR6002593, ERR6002595	
Hi-C Illumina	ERR6002596, ERR6002597, ERR6003038	
Genome assembly		
Assembly accession	GCA_905147135.1	
Accession of alternate haplotype	GCA_905147205.1	
Span (Mb)	577.1	
Number of contigs	115	
Contig N50 length (Mb)	11.6	
Number of scaffolds	44	
Scaffold N50 length (Mb)	20.5	
Longest scaffold (Mb)	35.4	
Genome annotation		
Number of protein-coding genes	10,302	
Number of non-coding genes	934	
Number of gene transcripts	18,275	

* 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/ilBlaLact1.1/dataset/CAJHUS01.1/busco.

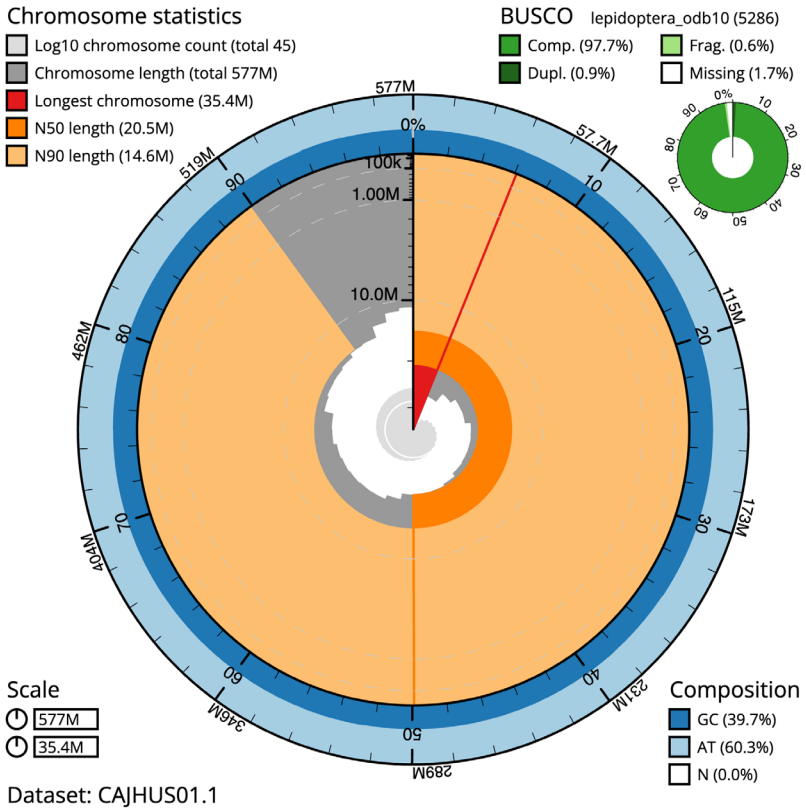

Figure 2. Genome assembly of *Blastobasis lacticolella*, ilBlaLact1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 577,083,784 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (35,431,177 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (20,548,006 and 14,576,097 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/ilBlaLact1.1/dataset/CAJHUS01.1/snail.

Genome annotation report

The *Blastobasis lacticolella* genome assembly (GCA_905147135.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Blastobasis_lacticolella_GCA_905147135.1/Info/Index). The resulting annotation includes 18,275 transcribed mRNAs from 10,302 protein-coding and 934 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Blastobasis lacticolella* (specimen ID Ox000033, individual ilBlaLact1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.33) on 2019-06-29 using a light trap. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilBlaLact1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. 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 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.

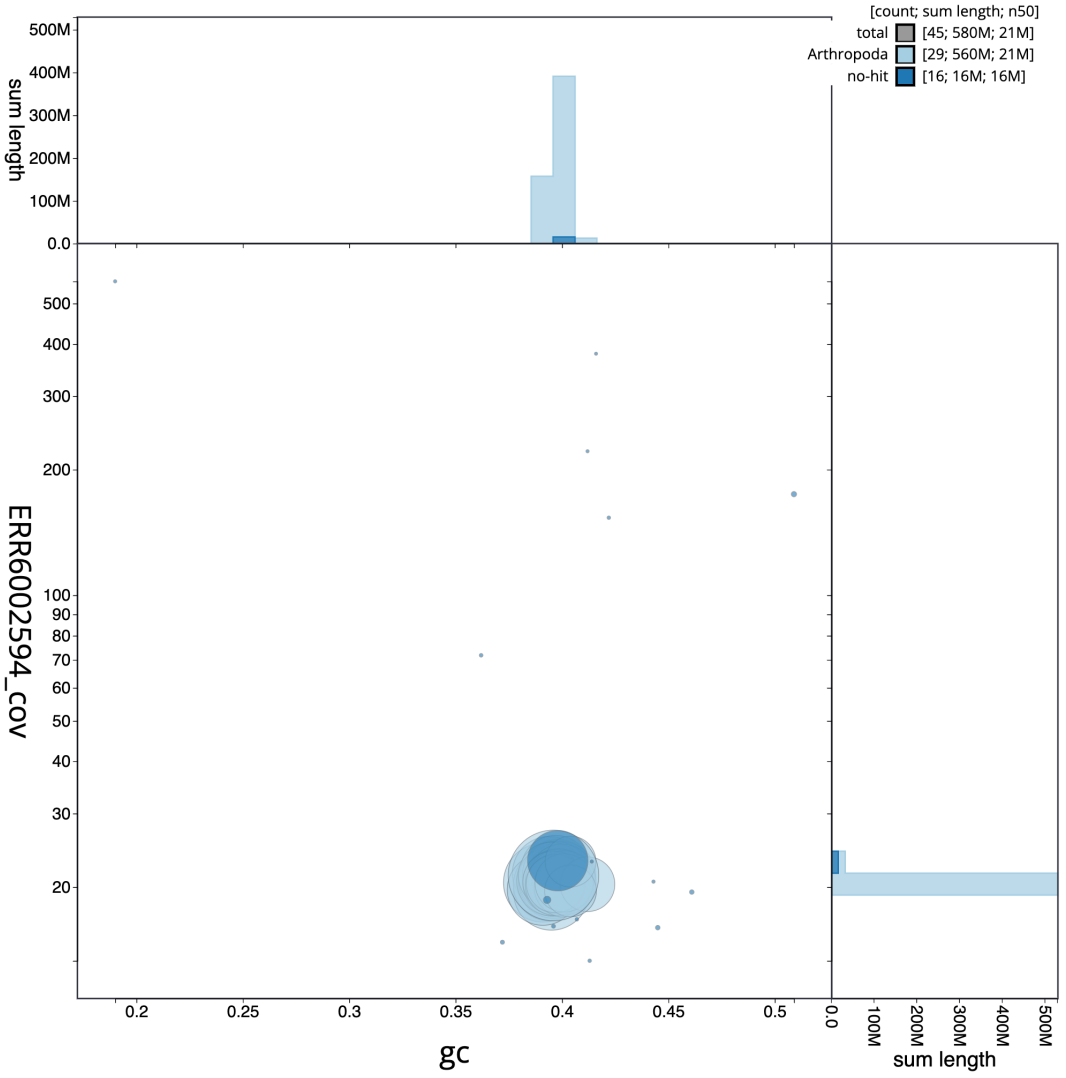

Figure 3. Genome assembly of *Blastobasis lacticolella*, ilBlaLact1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilBlaLact1.1/dataset/CAJHUS01.1/blob.

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 remaining tissue of ilBlaLact1 using the Arima2 kit and sequenced on the HiSeq X Ten instrument.

Genome assembly, curation and evaluation

Assembly was carried out with HiCanu (Nurk *et al.*, 2020) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) 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.

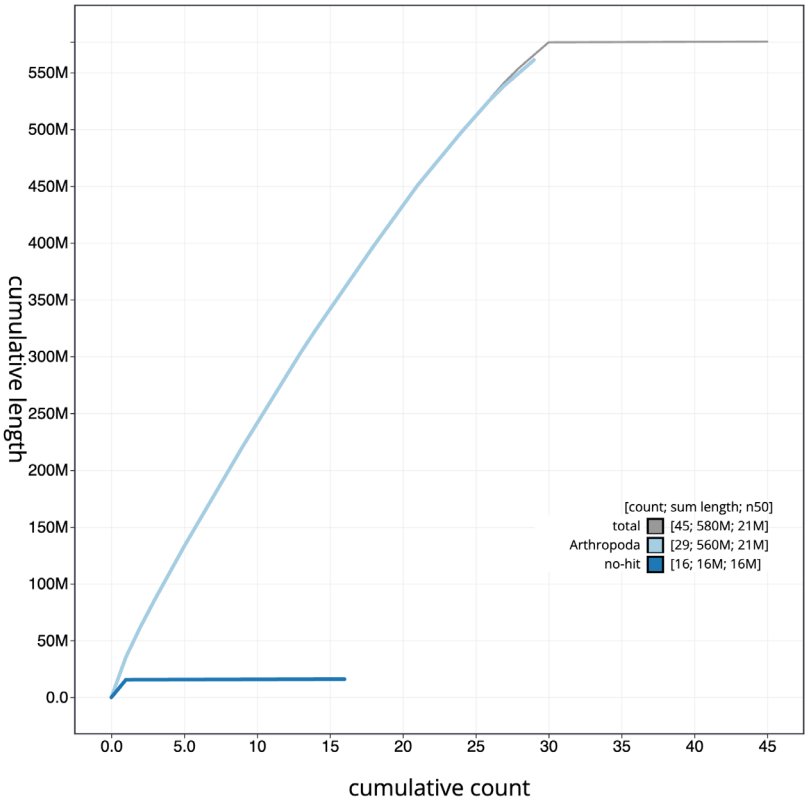

Figure 4. Genome assembly of *Blastobasis lacticolella*, ilBlaLact1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilBlaLact1.1/dataset/CAJHUS01.1/cumulative.

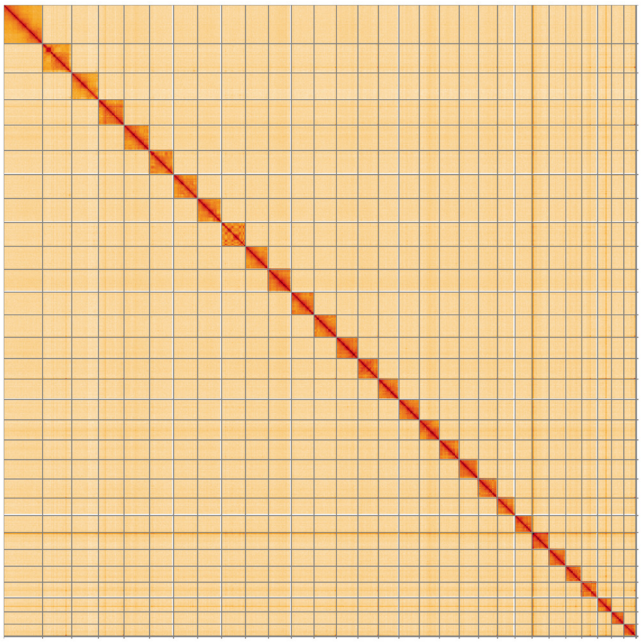

Figure 5. Genome assembly of *Blastobasis lacticolella*, ilBlaLact1.1: Hi-C contact map of the ilBlaLact1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=J3xCJ2DESky4XaVsB50DEw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Blastobasis lacticolella*, ilBlaLact1.

INSDC accession	Chromosome	Length (Mb)	GC%
LR990040.1	1	26.65	39.0
LR990041.1	2	24.43	39.5
LR990042.1	3	23.27	39.5
LR990043.1	4	23.19	40.0
LR990044.1	5	22.02	39.5
LR990045.1	6	21.92	39.5
LR990046.1	7	21.88	40.0
LR990047.1	8	21.75	39.0
LR990048.1	9	21.09	39.5
LR990049.1	10	20.74	39.5
LR990050.1	11	20.67	39.5
LR990051.1	12	20.55	39.5
LR990052.1	13	19.48	40.0
LR990053.1	14	18.7	39.5
LR990054.1	15	18.67	39.5
LR990055.1	16	18.52	39.5
LR990056.1	17	18.36	40.0
LR990057.1	18	18.14	40.0
LR990058.1	19	17.57	40.0
LR990059.1	20	17.49	40.0
LR990060.1	21	15.85	39.5
LR990061.1	22	15.58	39.5
LR990062.1	23	15.5	40.0
LR990063.1	24	15.23	40.0
LR990064.1	25	14.58	39.5
LR990065.1	26	14.43	40.0
LR990066.1	27	12.63	41.0
LR990067.1	28	11.31	40.5
LR990068.1	29	10.9	40.5
LR990039.1	Z	35.43	39.5
LR990069.1	MT	0.02	19.0

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 lacticolella* assembly (GCA_905147135.1). 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 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	4.0.7	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/
HiCanu	1	https://github.com/marbl/canu
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	2	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

Data availability

European Nucleotide Archive: *Blastobasis lacticolella* (London dowd). Accession number PRJEB42117; https://identifiers.org/ena.embl/PRJEB42117. (Wellcome Sanger Institute, 2020)

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

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: **Blastobasidae.** In: 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.

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

GBIF Secretariat: ***Blastobasis lacticolella* Rebel, 1939.** *GBIF Backbone Taxonomy. Checklist dataset.* 2022; [Accessed 26 July 2023].
Reference Source

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

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

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

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

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

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

Nurk S, Walenz BP, Rhie A, *et al.*: **HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads.** *Genome Res.* 2020; **30**(9): 1291–1305.
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 (Volume 2).** Lancashire & Cheshire Fauna Society, Rishton. 2021.
Reference Source

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

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

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**. 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 London Dowd, *Blastobasis lacticolella* (Rebel, 1939).** European Nucleotide Archive. [dataset], accession number PRJEB42117, 2020.