

## The genome sequence of the Birch Marble, *Apotomis betuletana* (Haworth, 1811)

Mulley, John; Boyes, Douglas

### Wellcome Open Research

DOI:

[10.12688/wellcomeopenres.18984.1](https://doi.org/10.12688/wellcomeopenres.18984.1)

Published: 09/02/2023

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Mulley, J., & Boyes, D. (2023). The genome sequence of the Birch Marble, *Apotomis betuletana* (Haworth, 1811). *Wellcome Open Research*, 8, 66.  
<https://doi.org/10.12688/wellcomeopenres.18984.1>

#### Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# **The genome sequence of the Birch Marble, *Apotomis betuletana* (Haworth, 1811)**

## **Authors**

Douglas Boyes, 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, John F. Mulley, Darwin Tree of Life Consortium

## **Abstract**

We present a genome assembly from an individual male *Apotomis betuletana* (the Birch Marble; Arthropoda; Insecta; Lepidoptera; Tortricidae). The genome sequence is 684 megabases in span. Most of the assembly is scaffolded into 28 chromosomal pseudomolecules with the Z sex chromosome assembled. The mitochondrial genome has also been assembled and is 15.8 kilobases in length. Gene annotation of this assembly on Ensembl identified 21,717 protein coding genes.

## **Keywords**

*Apotomis betuletana*, Birch Marble, genome sequence, chromosomal, Lepidoptera

[[John F Mulley](#)

[School of Natural Sciences, Bangor University, Deiniol Road, Bangor, LL57 2UW](#)

[j.mulley@bangor.ac.uk](mailto:j.mulley@bangor.ac.uk)]

## Main body

### Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Tortricoidea; Tortricidae; Olethreutinae; Olethreutini; *Apotomis*; *Apotomis betuleтана* (Haworth, 1811) (NCBI:txid1100915).

### Background

The birch marble *Apotomis betuleтана* (Tortricidae: Olethreutinae) is a relatively large (16–20 mm wingspan) ‘micromoth’, widely reported from across the Western palearctic.

*A. betuleтана* was classified as ‘common’ in a recent assessment of British microlepidoptera (Davis, 2012). Adults fly between June and September in the UK and are readily attracted to light, and traps baited with (Z)-10-tetradecenyl acetate will attract males (Booij and Voerman, 1984). The species was named in 1811 as *Tortrix betuleтана* by Adrian Hardy Haworth in volume III of his ‘Lepidoptera Britannica’ (Haworth, 1811), with the species name recognising the association with birch (*Betula* sp.), which is the primary, and possibly only, larval foodplant. Larvae feed in rolled and folded birch leaves during early summer.

The basal two-thirds of the adult wing are dark brown to black, with the distal third white, sometimes with small yellowish or brown spots or patches. The overall appearance, as with several other genera of Tortricidae, is of a bird dropping, and the species is often referred to as a bird dropping mimic. Cott, in his 1940 work on adaptive colouration, considered this phenomenon to be a case of ‘special resemblance’ (Cott, 1940), but it is actually an example of masquerade. Masquerade is differentiated from crypsis in that the latter makes potential prey items difficult to detect and therefore confuses predators’ sensory processes, whereas masquerading individuals resemble inanimate, typically inedible, items, such as sticks, stones, or bird droppings, which confuse predators’ cognitive abilities (Skelhorn *et al.*, 2010). Masquerade is most effective when the object being copied is common in the environment (Skelhorn *et al.*, 2011; Skelhorn, 2015), raising the possibility that declines in woodland birds (Burns *et al.*, 2020) may counterintuitively lead to a decline in the abundance of this species.

The genome of *A. betuleтана* will aid research into mechanisms of masquerade in Lepidoptera. Here we present a chromosomally complete genome sequence for *A. betuleтана*, based on one male specimen from Wytham Woods, Oxfordshire.

### Genome sequence report

The genome was sequenced from one male *Apotomis betuleтана* (Figure 1) collected from Wytham Woods, UK (latitude 51.77, longitude –1.33). A total of 34-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 71-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 27 missing or mis-joins and removed seven haplotypic duplications, reducing the scaffold number by 24.05%.

The final assembly has a total length of 684.2 Mb in 60 sequence scaffolds with a scaffold N50 of 24.9 Mb (Table 1). Most (99.63%) of the assembly sequence was assigned to 28 chromosomal-level scaffolds, representing 27 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figures 2–5; Table 2). The assembly has a BUSCO v5.3.2 (Manni *et al.*, 2021) completeness of 98.3% (single 97.5%, duplicated 0.7%) using the *lepidoptera\_odb10* reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

## Genome annotation report

The *A. betuletana* genome assembly GCA\_932273695.1 was annotated using the Ensembl rapid annotation pipeline (Table 1; [https://rapid.ensembl.org/Apotomis\\_betuletana\\_GCA\\_932273695.1/](https://rapid.ensembl.org/Apotomis_betuletana_GCA_932273695.1/)). The resulting annotation includes 21,968 transcribed mRNAs from 21,717 protein-coding genes.

## Methods

### Sample acquisition and nucleic acid extraction

Two *A. betuletana* specimens (ilApoBetu1 and ilApoBetu2) were collected in Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.77, longitude –1.33) on 31 August 2020 by netting. The specimens were collected and identified by Douglas Boyes (University of Oxford) and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilApoBetu1 sample was weighed and dissected on dry ice. Whole organism tissue 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 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.

### 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 Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from

whole organism tissue of ilApoBetu2 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

## Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge\_dups (Guan *et al.*, 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with freebayes (Garrison and Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou, McCarthy and Durbin, 2022). The assembly was checked for contamination 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.*, 2022), which performed annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores were generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

## Genome annotation

The BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *A. betuletana* assembly (GCA\_932273695.1) in Ensembl Rapid Release.

## Ethics and compliance issues

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](#). 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. All efforts are undertaken to minimise the suffering of animals. 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: *Apotomis betuletana* (birch marble). Accession number [PRJEB49039](#); <https://identifiers.org/ena.embl/PRJEB49039>. (Wellcome Sanger Institute, 2021)

The genome sequence is released openly for reuse. The *Apotomis betuletana* 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.

## 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).

## 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

- Allio, R. *et al.* (2020) 'MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics', *Molecular Ecology Resources*, 20(4), pp. 892–905. Available at: <https://doi.org/10.1111/1755-0998.13160>.
- Booij, C.J.H. and Voerman, S. (1984) 'New sex attractants for 35 tortricid and 4 other lepidopterous species, found by systematic field screening in the Netherlands', *Journal of Chemical Ecology*, 10(1), pp. 135–144. Available at: <https://doi.org/10.1007/BF00987650>.
- Brůna, T. *et al.* (2021) 'BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database', *NAR Genomics and Bioinformatics*, 3(1), pp. 1–11. Available at: <https://doi.org/10.1093/nargab/lqaa108>.
- Burns, F. *et al.* (2020) *The state of the UK's birds 2020*. Sandy, Bedfordshire: The RSPB, BTO, WWT, DAERA, JNCC, NatureScot, NE and NRW. Available at: <https://www.bto.org/our-science/publications/state-uks-birds/state-uks-birds-2020> (Accessed: 25 January 2023).
- Challis, R. *et al.* (2020) 'BlobToolKit - interactive quality assessment of genome assemblies', *G3: Genes, Genomes, Genetics*, 10(4), pp. 1361–1374. Available at: <https://doi.org/10.1534/g3.119.400908>.
- Cheng, H. *et al.* (2021) 'Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm', *Nature Methods*, 18(2), pp. 170–175. Available at: <https://doi.org/10.1038/s41592-020-01056-5>.
- Cott, H.B. (1940) *Adaptive coloration in animals*. London: Methuen & Co. Ltd.

- Davis, T. (2012) 'A review of the status of Microlepidoptera in Britain', *Butterfly Conservation Report* [Preprint]. Available at: [chrome-extension://http://www.gloucestershire-butterflies.org.uk/download/The\\_Status\\_of\\_Micro\\_Lepidoptera\\_in\\_Britain.pdf](http://www.gloucestershire-butterflies.org.uk/download/The_Status_of_Micro_Lepidoptera_in_Britain.pdf) (Accessed: 6 December 2022).
- Garrison, E. and Marth, G. (2012) 'Haplotype-based variant detection from short-read sequencing'. Available at: <http://arxiv.org/abs/1207.3907>.
- Guan, D. *et al.* (2020) 'Identifying and removing haplotypic duplication in primary genome assemblies', *Bioinformatics*, 36(9), pp. 2896–2898. Available at: <https://doi.org/10.1093/bioinformatics/btaa025>.
- Harry, E. (2022) 'PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.' Available at: <https://github.com/wtsi-hpag/PretextView> (Accessed: 19 October 2022).
- Haworth, A.W. (1811) *Lepidoptera Britannica; sistens digestionem novam insectorum lepidopterorum quæ in Magnâ Britannîâ reperiuntur, larvarum pabulo, temporeque pascendi; expansione alarum; mensibusque volandi; synonymis atque locis observationibusque variis. Vol III*. London: Richard Taylor & Co.
- Howe, K. *et al.* (2021) 'Significantly improving the quality of genome assemblies through curation', *GigaScience*. Oxford University Press. Available at: <https://doi.org/10.1093/gigascience/giaa153>.
- Kerpedjiev, P. *et al.* (2018) 'HiGlass: Web-based visual exploration and analysis of genome interaction maps', *Genome Biology*, 19(1). Available at: <https://doi.org/10.1186/s13059-018-1486-1>.
- Manni, M. *et al.* (2021) 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes', *Molecular Biology and Evolution*, 38(10), pp. 4647–4654. Available at: <https://doi.org/10.1093/molbev/msab199>.
- Rao, S.S.P. *et al.* (2014) 'A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping', *Cell*, 159(7), pp. 1665–1680. Available at: <https://doi.org/10.1016/j.cell.2014.11.021>.
- Rhie, A. *et al.* (2021) 'Towards complete and error-free genome assemblies of all vertebrate species', *Nature*, 592(7856), pp. 737–746. Available at: <https://doi.org/10.1038/s41586-021-03451-0>.
- Skelhorn, J. *et al.* (2010) 'Masquerade: Camouflage Without Crypsis', *Science*, 327(5961), pp. 51–51. Available at: <https://doi.org/10.1126/science.1181931>.
- Skelhorn, J. *et al.* (2011) 'Density-dependent predation influences the evolution and behavior of masquerading prey', *Proceedings of the National Academy of Sciences*, 108(16), pp. 6532–6536. Available at: <https://doi.org/10.1073/pnas.1014629108>.
- Skelhorn, J. (2015) 'Masquerade', *Current Biology*, 25(15), pp. R643–R644. Available at: <https://doi.org/10.1016/j.cub.2015.02.069>.
- Uliano-Silva, M. *et al.* (2022) 'MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads', *bioRxiv* [Preprint]. Available at: <https://doi.org/10.1101/2022.12.23.521667>.
- Zhou, C., McCarthy, S.A. and Durbin, R. (2022) 'YaHS: yet another Hi-C scaffolding tool', *bioRxiv* [Preprint]. Available at: <https://doi.org/10.1101/2022.06.09.495093>.

[\[add dataset citation\]](#)

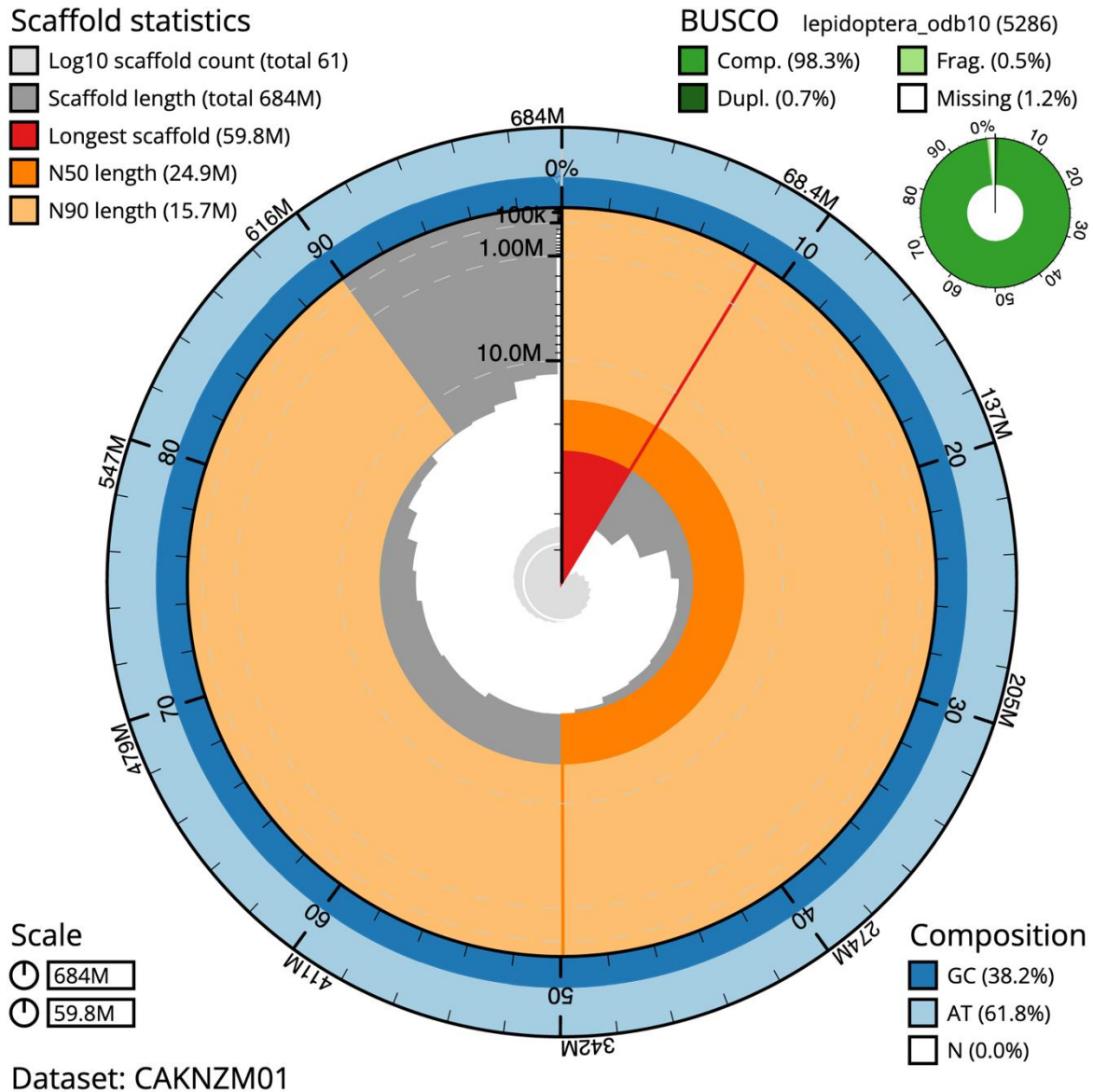
Wellcome Sanger Institute (2021) The genome sequence of the Birch Marble, *Apotomis betulana* (Haworth, 1811), European Nucleotide Archive, [dataset], accession number PRJEB49039



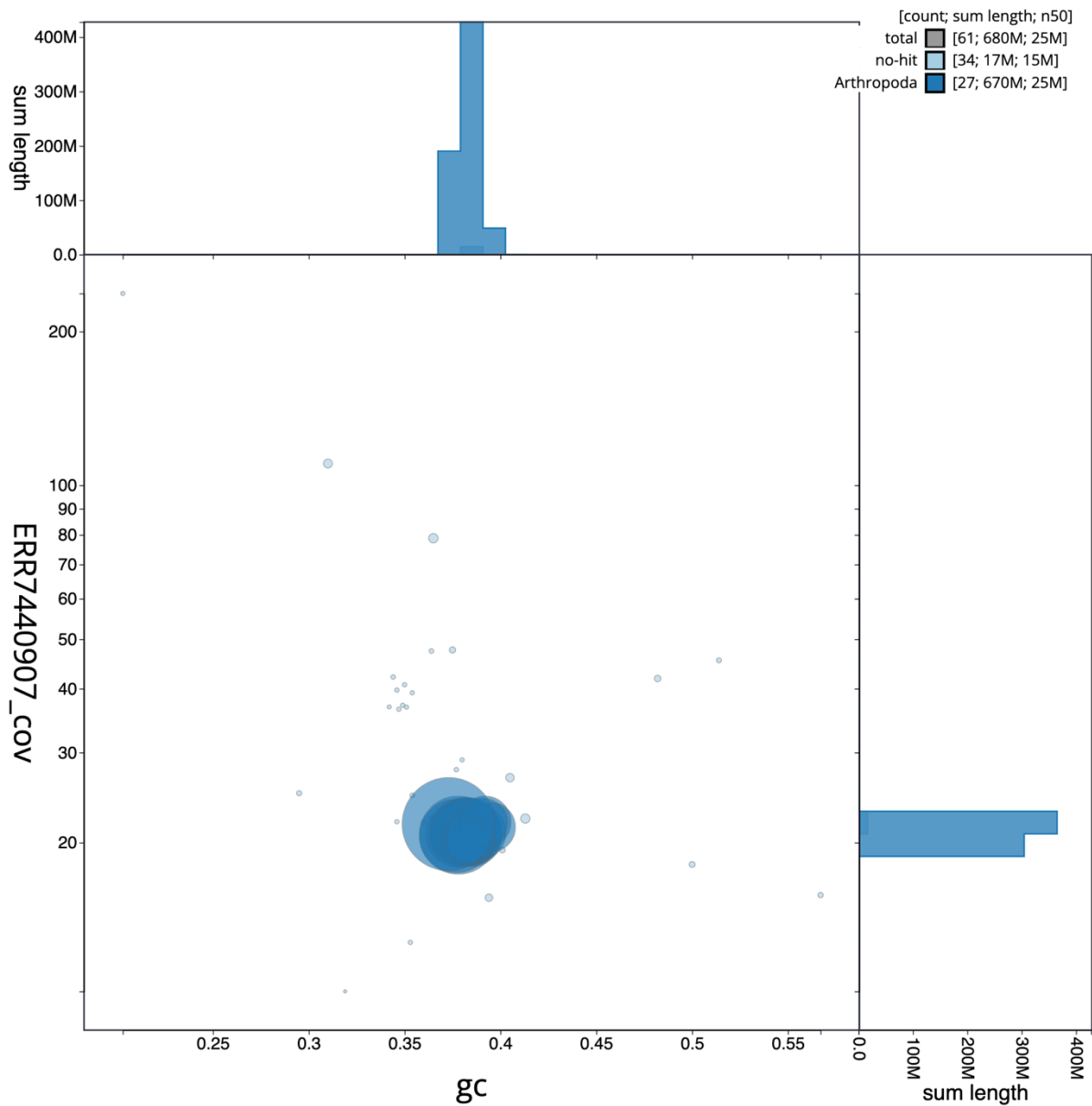
## Figures



Figure 1. Image of the *Apotomis betuletana* (ilApoBetu1) specimen used for genome sequencing.



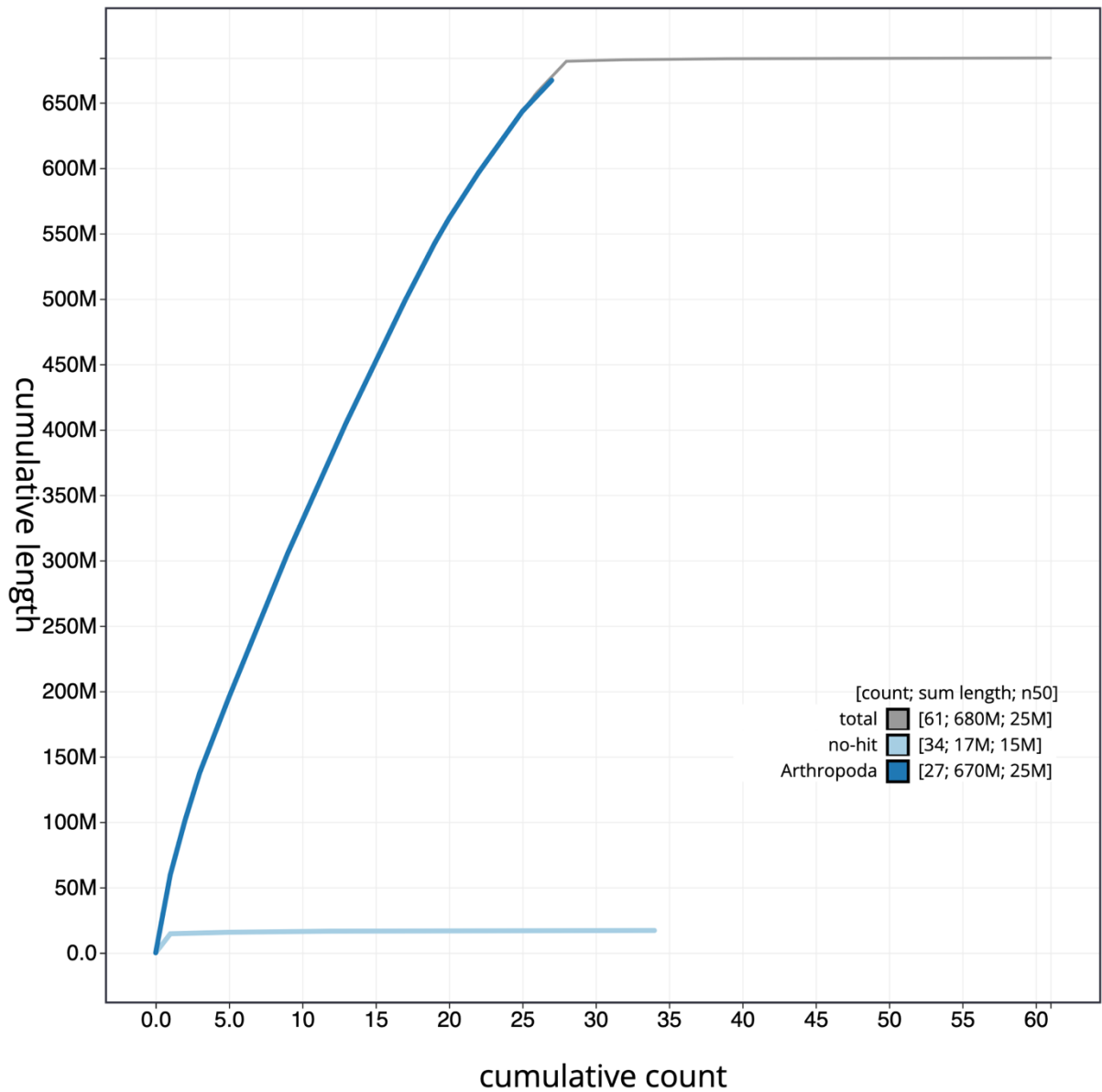
**Figure 2: Genome assembly of *Apotomis betuletana*, ilApoBetu1.1: metrics.** The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 684,240,599 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (59,814,915 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (24,873,308 and 15,741,175 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/ilApoBetu1.1/dataset/CAKNZM01/snail>.



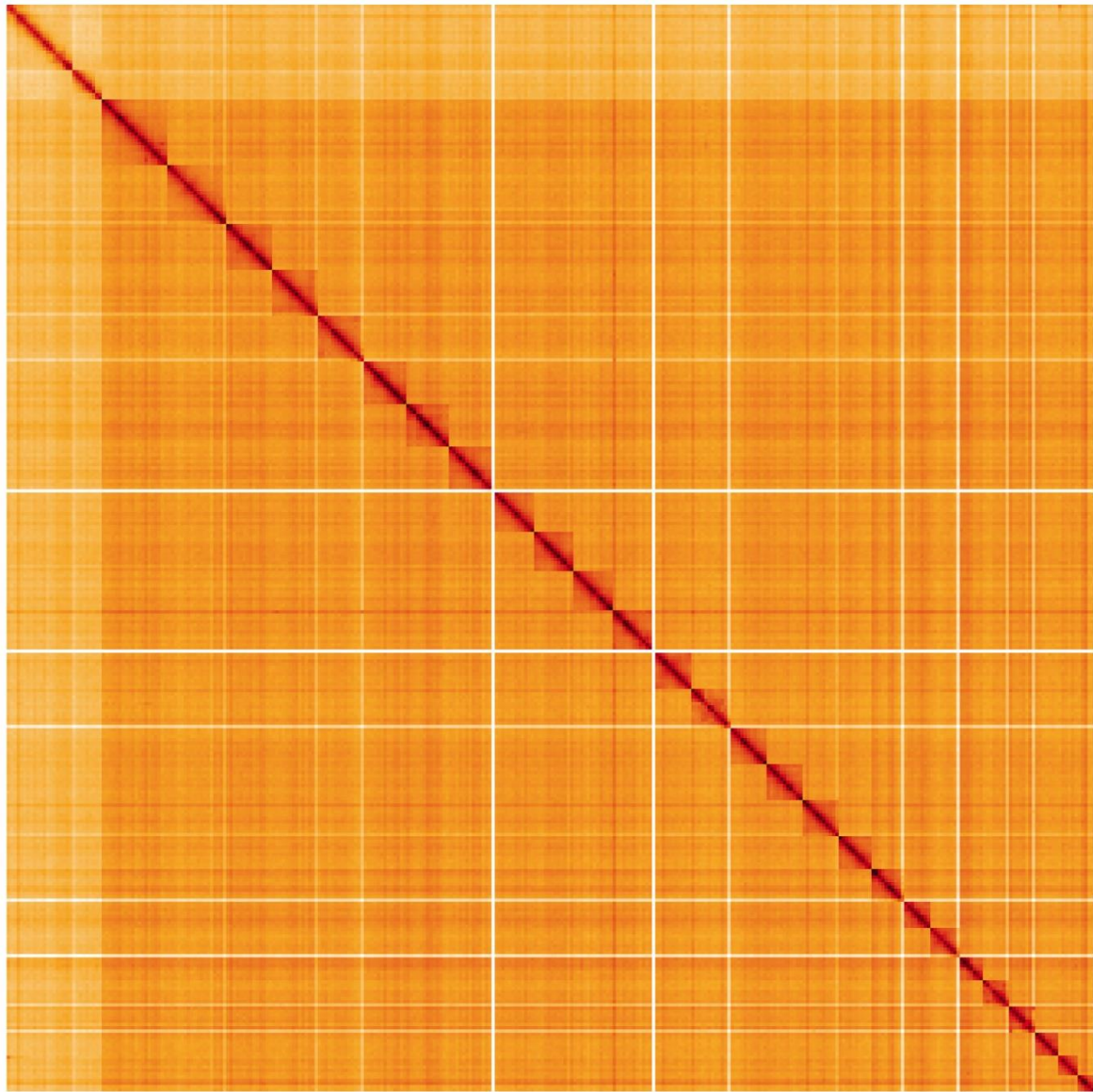
**Figure 3: Genome assembly of *Apotomis betuletana*, ilApoBetu1.1: GC coverage.**

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/ilApoBetu1.1/dataset/CAKNZM01/blob>.



**Figure 4: Genome assembly of *Apotomis betuletana*, ilApoBetu1.1: cumulative sequence.** 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/ilApoBetu1.1/dataset/CAKNZM01/cumulative>.



**Figure 5: Genome assembly of *Apotomis betuletana*, ilApoBetu1.1: Hi-C contact map.**

Hi-C contact map of the ilApoBetu1.1 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/?d=DsBR-8qXSpyKsith2rNGKw>.

## Tables

**Table 1: Genome data for *Apotomis betuletana*, ilApoBetu1.1.**

Project accession data		
Assembly identifier		ilApoBetu1.1
Species		<i>Apotomis betuletana</i>
Specimen		ilApoBetu1
NCBI taxonomy ID		1100915
BioProject		PRJEB49039
BioSample ID		SAMEA7701588
Isolate information		ilApoBetu1 (PacBio and 10X) ilApoBetu2 (Hi-C)
Assembly metrics*		Benchmark
Consensus quality (QV)	57.6	$\geq 50$
k-mer completeness	99.99%	$\geq 95\%$
BUSCO**	C:98.3%[S:97.5%,D:0.7%], F:0.5%,M:1.2%,n:5,286	$C \geq 95\%$
Percentage of assembly mapped to chromosomes	99.63%	$\geq 95\%$
Sex chromosomes	Z chromosome	localised homologous pairs
Organelles	Mitochondrial genome assembled	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II		ERR7815860
10X Genomics Illumina		ERR7440907–ERR7440910
Hi-C Illumina		ERR7569936
Genome assembly		
Assembly accession		GCA_932273695.1
Accession of alternate haplotype		GCA_932273875.1
Span (Mb)		684.2
Number of contigs		134
Contig N50 length (Mb)		12.4
Number of scaffolds		60
Scaffold N50 length (Mb)		24.9



Longest scaffold (Mb)	59.8
<b>Genome annotation</b>	
Number of protein-coding genes	21,717
Number of gene transcripts	21,968

\* 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/ilApoBetu1.1/dataset/CAKNZM01/busco>.

**Table 2: Chromosomal pseudomolecules in the genome assembly of *Apotomis betuletana*, ilApoBetu1.**

INSDC accession	Chromosome	Size (Mb)	GC%
OW026299.1	1	41.42	37.8
OW026300.1	2	36.39	37.7
OW026301.1	3	29.81	38.1
OW026302.1	4	28.07	37.8
OW026303.1	5	27.84	38.1
OW026304.1	6	27.76	38.4
OW026305.1	7	27.36	38.3
OW026306.1	8	26.96	38
OW026307.1	9	25.71	38.2
OW026308.1	10	24.87	38.4
OW026309.1	11	24.76	38.2
OW026310.1	12	24.8	37.9
OW026311.1	13	23.63	38
OW026312.1	14	23.53	38.5
OW026313.1	15	23.02	38.3
OW026314.1	16	22.98	38.4
OW026315.1	17	22.07	38.5

OW026316.1	18	21.35	38.8
OW026317.1	19	19.4	38.5
OW026318.1	20	17.54	39.2
OW026319.1	21	17.27	38.4
OW026320.1	22	15.82	39.5
OW026321.1	23	15.74	38.4
OW026322.1	24	15.48	39.1
OW026323.1	25	14.55	38.5
OW026324.1	26	12.06	38.6
OW026325.1	27	11.69	38.3
OW026298.1	Z	59.81	37.3
OW026326.1	MT	0.02	20.3
-	unplaced	2.52	38.6

**Table 3. Software tools and versions used.**

Software tool	Version	Source
BlobToolKit	3.5.0	Challis <i>et al.</i> , 2020
freebayes	1.3.1-17-gaa2ace8	Garrison and Marth, 2012
Hifiasm	0.16.1	Cheng <i>et al.</i> , 2021
HiGlass	1.11.6	Kerpedjiev <i>et al.</i> , 2018
Long Ranger ALIGN	2.2.2	<a href="https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines">https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines</a>
MitoHiFi	2	Uliano-Silva <i>et al.</i> , 2022
PretextView	0.2	Harry, 2022
purge_dups	1.2.3	Guan <i>et al.</i> , 2020
YaHS	1.0	Zhou <i>et al.</i> 2022