DATA NOTE

The genome sequence of the Dark Spectacle, *Abrostola triplasia* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes[1+], Owen T. Lewis[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 Oxford, Oxford, England, UK

[+] Deceased author

v1 **First published:** 27 Jun 2023, **8**:278
https://doi.org/10.12688/wellcomeopenres.19624.1
Latest published: 27 Jun 2023, **8**:278
https://doi.org/10.12688/wellcomeopenres.19624.1

Abstract

We present a genome assembly from an individual male *Abrostola triplasia* (the Dark Spectacle; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 362.7 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.34 kilobases in length. Gene annotation of this assembly on Ensembl identified 11,532 protein coding genes.

Keywords

Abrostola triplasia, Dark Spectacle, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 27 Jun 2023	✓ view	✓ view

1. **Jesper Bechsgaard**, Aarhus University, Aarhus, Denmark
Jeppe Bayer Pedersen, Aarhus University, Aarhus, Denmark

2. **William Hemstrom**, University of California, California, USA

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; **Lewis OT**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194) 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, Lewis OT, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Dark Spectacle, *Abrostola triplasia* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2023, **8**:278 https://doi.org/10.12688/wellcomeopenres.19624.1

First published: 27 Jun 2023, **8**:278 https://doi.org/10.12688/wellcomeopenres.19624.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Noctuidae; Plusiinae; *Abrostola*; *Abrostola triplasia* (Linnaeus, 1758) (NCBI:txid254365).

Background

The Dark Spectacle (*Abrostola triplasia*) is a widespread and common moth species in the UK and Ireland, where it is most frequent and abundant in the west (Randle *et al.*, 2019; Waring *et al.*, 2017). It especially favours areas with acidic soils, but unlike the closely-related Spectacle, *Abrostola tripartita* (Hufnagel, 1766) there are relatively few records from Scotland, most of them from southern Scotland. Distribution trends for this species in the UK and Ireland show a long term decline of 38%, in contrast to *A. tripartita* which has increased its distribution over the same period (Randle *et al.*, 2019). *Abrostola triplasia* has been recorded across much of temperate Europe and Asia to China and Japan (GBIF Secretariat, 2023).

The main larval foodplants are nettle (*Urtica dioica*) and hop (*Humulus lupulus*). The distinctive larva with its two yellow "eye spots" feeds mostly at night and it has been suggested that it is a snake mimic (Henwood *et al.*, 2020).

Here we present a chromosomally complete genome sequence for *Abrostola triplasia* based on one male specimen from Wytham Woods, Oxfordshire, UK. A genome sequence for *Abrostola triplasia* will contribute to a growing data set of resources for understanding lepidopteran biology. The genome of *Abrostola triplasia* was sequenced as part of the Darwin Tree of Life Project, 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 *Abrostola triplasia* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.34). A total of 39-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected one mis-join.

The final assembly has a total length of 362.7 Mb in 35 sequence scaffolds with a scaffold N50 of 12.8 Mb (Table 1). Most (99.96%) of the assembly sequence was assigned to 31 chromosomal-level scaffolds, representing 30 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.

Figure 1. Photograph of the *Abrostola triplasia* (ilAbrTril1) specimen used for genome sequencing.

The estimated Quality Value (QV) of the final assembly is 68.4 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.9% (single = 98.7%, duplicated = 0.2%), 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/254365.

Genome annotation report

The *Abrostola triplasia* genome assembly (GCA_946251915.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Abrostola_triplasia_GCA_946251915.1/Info/Index). The resulting annotation includes 20,311 transcribed mRNAs from 11,532 protein-coding and 1,479 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

A male *Abrostola triplasia* (specimen ID Ox001901, individual ilAbrTril1) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.34) on 2021-06-16 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 ilAbrTril1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Head and thorax tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system

Table 1. Genome data for *Abrostola triplasia*, ilAbrTril1.1.

Project accession data		
Assembly identifier	ilAbrTril1.1	
Species	*Abrostola triplasia*	
Specimen	ilAbrTril1	
NCBI taxonomy ID	254365	
BioProject	PRJEB55021	
BioSample ID	SAMEA10979163	
Isolate information	ilAbrTril1, male: head and thorax (DNA sequencing and Hi-C scaffolding), abdomen (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	68.4	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.9%[S:98.7%,D:0.2%], F:0.2%,M:0.9%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.96%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10008903	
Hi-C Illumina	ERR10015060	
PolyA RNA-Seq Illumina	ERR10890699	
Genome assembly		
Assembly accession	GCA_946251915.1	
Accession of alternate haplotype	GCA_946251925.1	
Span (Mb)	362.7	
Number of contigs	39	
Contig N50 length (Mb)	12.7	
Number of scaffolds	35	
Scaffold N50 length (Mb)	12.8	
Longest scaffold (Mb)	21.2	
Genome annotation		
Number of protein-coding genes	11,532	
Number of non protein-coding genes	1,479	
Number of gene transcripts	20,311	

* 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/ilAbrTril1.1/dataset/CAMIUI01/busco.

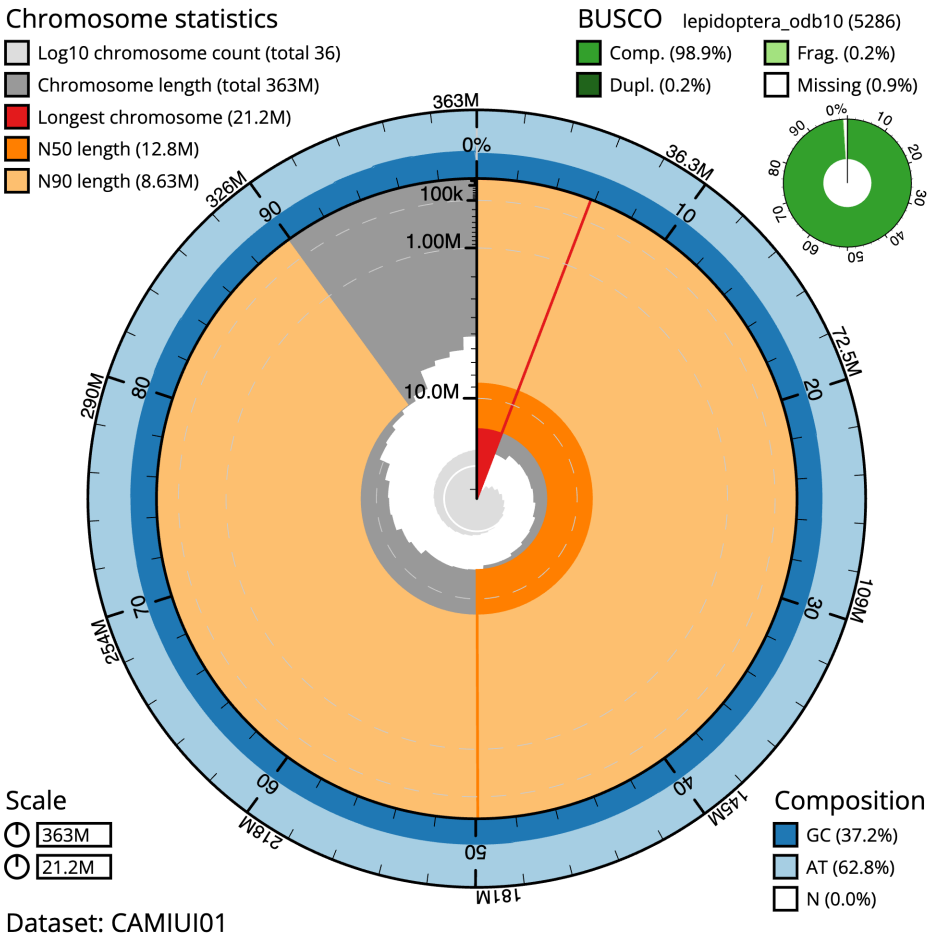

Figure 2. Genome assembly of *Abrostola triplasia*, ilAbrTril1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 362,722,748 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (21,208,463 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,758,438 and 8,625,622 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/ilAbrTril1.1/dataset/CAMIUI01/snail.

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.

RNA was extracted from abdomen tissue of ilAbrTril1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments.

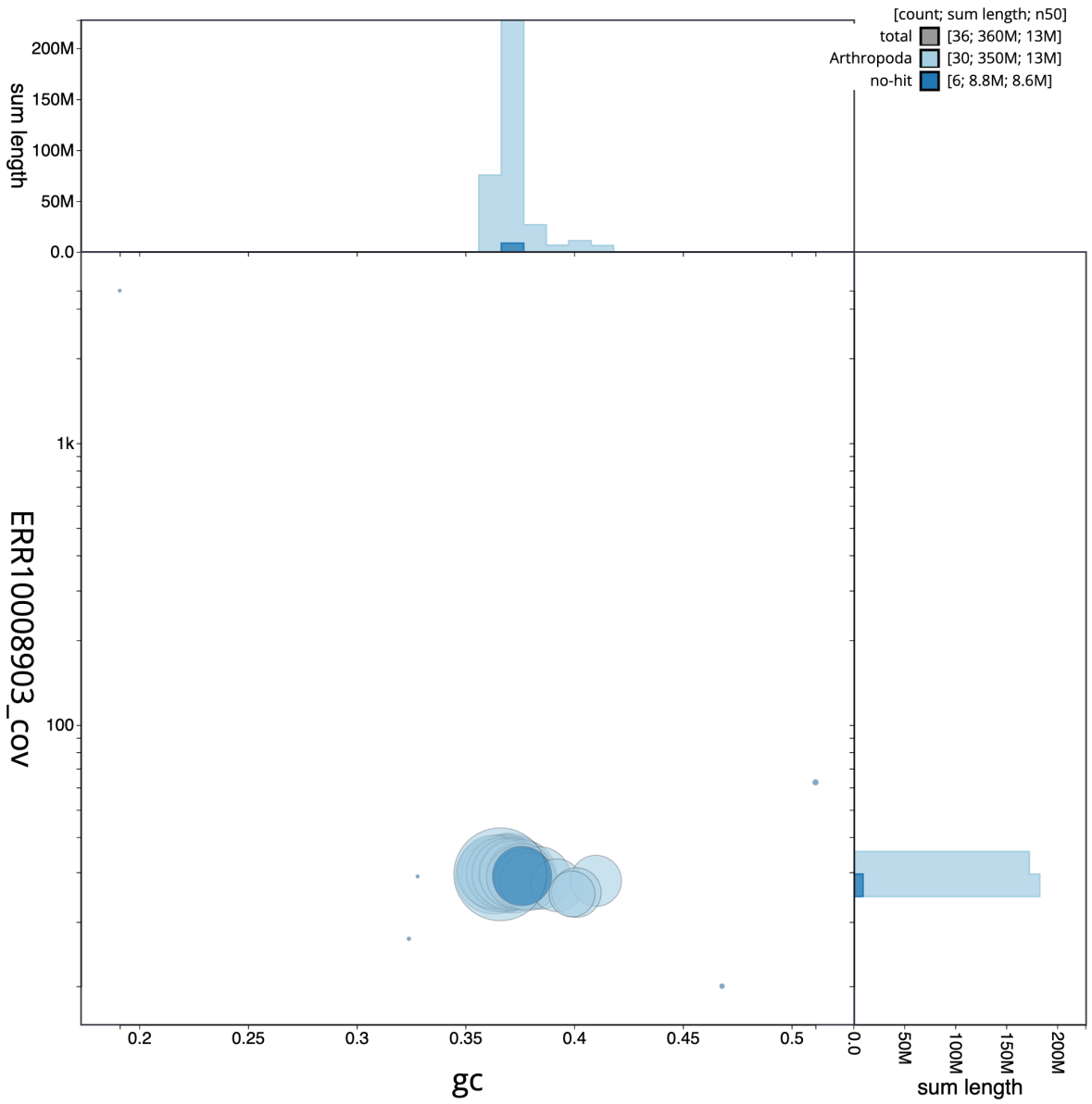

Figure 3. Genome assembly of *Abrostola triplasia*, ilAbrTril1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilAbrTril1.1/dataset/CAMIUI01/blob.

Hi-C data were also generated from ilAbrTril1 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 purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2022), 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,

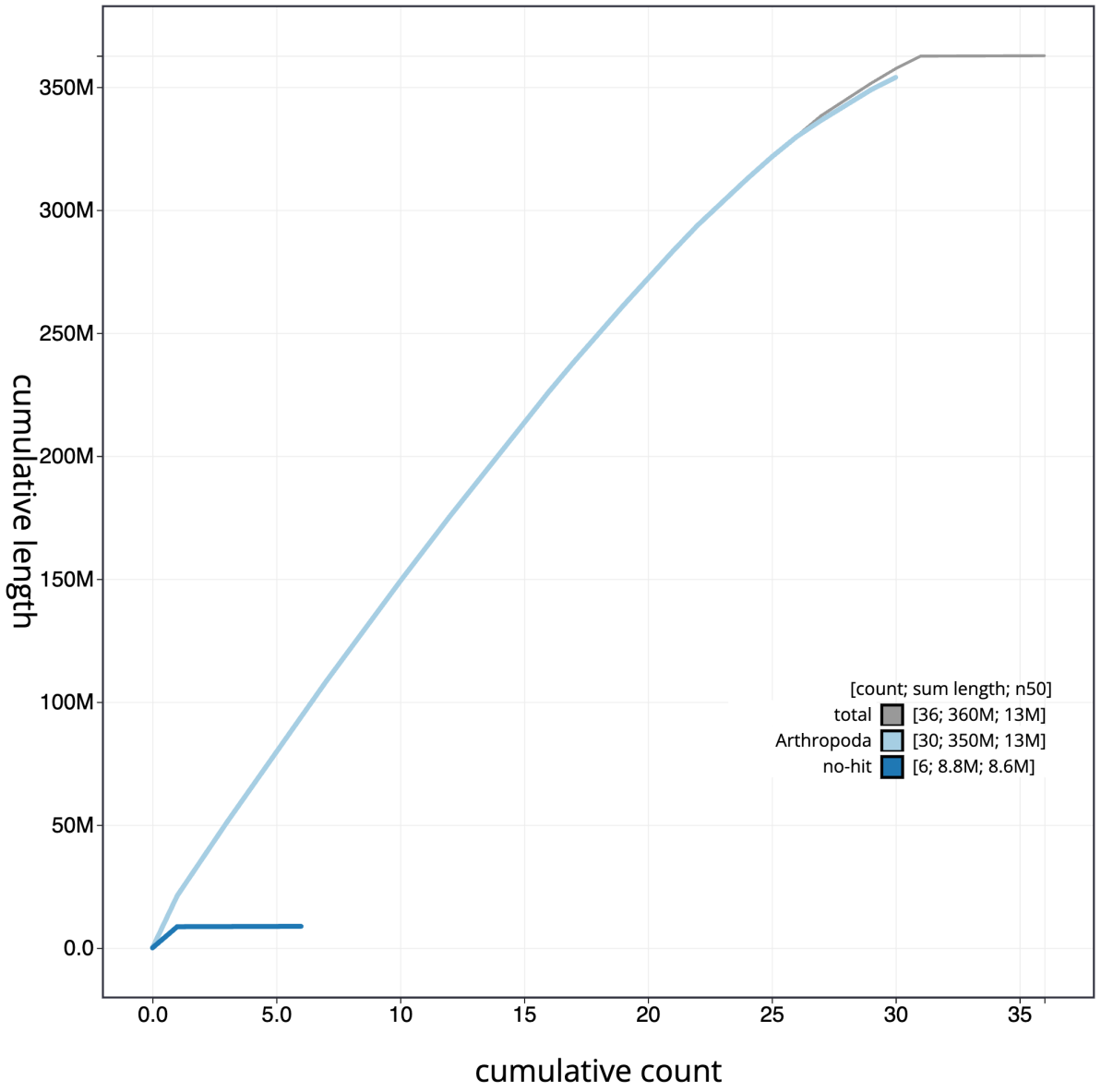

Figure 4. Genome assembly of *Abrostola triplasia*, ilAbrTril1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilAbrTril1.1/dataset/CAMIUI01/cumulative.

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 gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Abrostola triplasia* assembly (GCA_946251915.1). 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

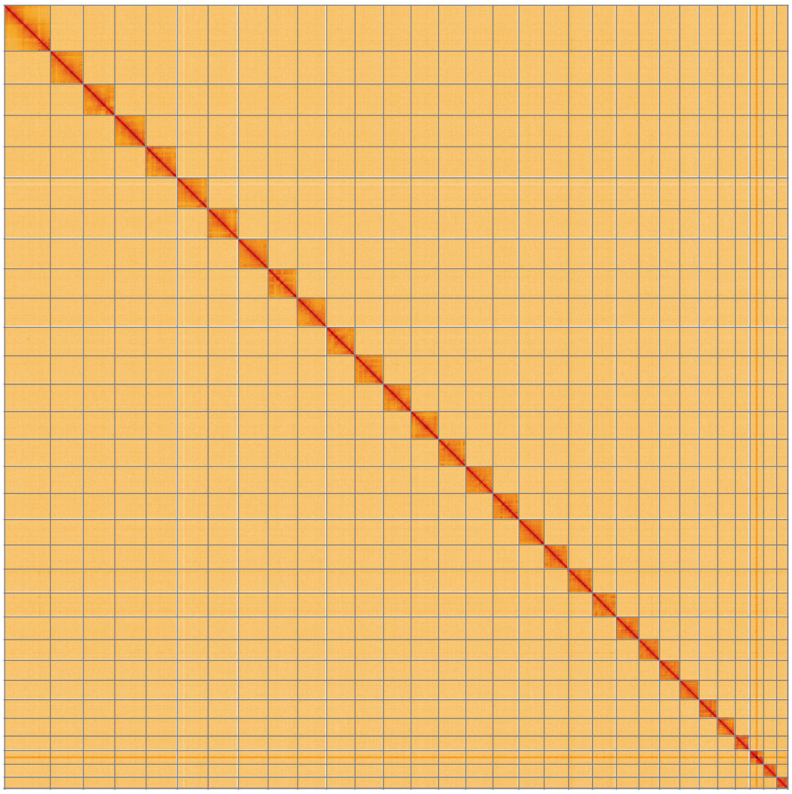

Figure 5. Genome assembly of *Abrostola triplasia*, ilAbrTril1.1: Hi-C contact map of the ilAbrTril1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=XXUZC-xDR-uMmH0_Av5lzg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Abrostola triplasia*, ilAbrTril1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX276421.1	1	15.22	37.0
OX276422.1	2	14.58	37.0
OX276423.1	3	14.48	37.0
OX276424.1	4	14.42	37.0
OX276425.1	5	14.27	36.5
OX276426.1	6	14.04	37.0
OX276427.1	7	13.76	37.0
OX276428.1	8	13.67	36.5
OX276429.1	9	13.45	37.0
OX276430.1	10	13.18	36.5
OX276431.1	11	13.17	36.5
OX276432.1	12	12.76	37.0
OX276433.1	13	12.7	37.0
OX276434.1	14	12.62	36.5
OX276435.1	15	12.6	37.0
OX276436.1	16	12.0	37.0
OX276437.1	17	11.76	37.0
OX276438.1	18	11.24	37.5
OX276439.1	19	11.1	37.0
OX276440.1	20	10.99	37.5
OX276441.1	21	10.54	37.5
OX276442.1	22	9.54	38.5
OX276443.1	23	9.36	37.5
OX276444.1	24	8.95	37.5
OX276445.1	25	8.63	37.5
OX276446.1	26	8.19	38.0
OX276447.1	27	6.71	39.0
OX276448.1	28	6.36	41.0
OX276449.1	29	5.98	40.0
OX276450.1	30	5.1	40.0
OX276420.1	Z	21.21	36.5
OX276451.1	MT	0.02	19.5

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
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
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
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	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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.

Data availability

European Nucleotide Archive: *Abrostola triplasia* (dark spectacle). Accession number PRJEB55021; https://identifiers.org/ena.embl/PRJEB55021. (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *Abrostola triplasia* 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): 313–9.
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

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

GBIF Secretariat: ***Abrostola triplasia* (Linnaeus, 1758).** *GBIF Backbone Taxonomy*. 2023; [Accessed 3 June 2023].
Reference Source

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

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

Henwood B, Sterling P, Lewington R: **Field Guide to the Caterpillars of Great Britain and Ireland.** London: Bloomsbury, 2020; 448.
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

Randle Z, Evans-Hill LJ, Parsons MS, *et al.*: **Atlas of Britain & Ireland's Larger Moths.** Newbury: NatureBureau, 2019.
Reference Source

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): 1665–80.
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): 3210–2.
PubMed Abstract | 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

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads.** *BioRxiv.* 2022.
Publisher 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 Md, Misra S, Li H, *et al.*: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Waring P, Townsend M, Lewington R: **Field Guide to the Moths of Great Britain and Ireland: Third Edition.** Bloomsbury Wildlife Guides, 2017.
Reference Source

Wellcome Sanger Institute: **The genome sequence of the Dark Spectacle, *Abrostola triplasia* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB55021, 2022.

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 12 September 2023

https://doi.org/10.21956/wellcomeopenres.21740.r63622

✔ **William Hemstrom**
University of California, California, USA

Boyes *et al.* present here a chromosome-level genome assembly for *Abrostola triplasia*, the first such for the species. The reference individual was sequenced using a mix of PacBio long reads and assembled with the aid of Hi-C data. They then annotated the genome using transcriptomic data. The result is a very high quality genome, available from the ENA.

I have no major issues with this work or the author's methodology.

Minor Issues:

1. The y axis label for Figure 3 is uninformative, and should probably be replaced with something like "Coverage".

2. The grey line on Figure 4 is not visible behind the teal line for the most part. It should probably be re-plotted with the gray line as the top-most graphical layer.

3. Figure 5 needs axis labels and a color legend. While I understand that the axes denote chromosomes and positions along them, labels would be helpful.

4. There is at least one position shown on Figure 5 that seems to have a globally high interaction frequencies to the rest of the genome. This should probably be addressed briefly in the text.

5. The assembly section thoroughly describes the tools used, but should probably more fully describe the parameter choices and options used in each step. If this would make the text too long for publication here, these details could be relegated to a supplementary document. Alternatively (and probably preferably), the scripts used to conduct the analysis could be made available.

6. The second sentence of the Background section reads a bit awkwardly. Perhaps something like "... there are relatively few records from Scotland, mostly from southern parts of the

region."?
As a side-note, I very much appreciate the careful documentation of tool versions in Table 3.

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: I am a population and ecological geneticist specializing on methodological approaches. My major study species include Monarch Butterflies. I am not a specialist on genome assemblies, and so am not the best person to dive deeply into the technicalities of that aspect. However, the overall quality of the genome produced seems excellent in comparison to what I typically work with.

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 12 September 2023

https://doi.org/10.21956/wellcomeopenres.21740.r65939

Jesper Bechsgaard
Aarhus University, Aarhus, Denmark
Jeppe Bayer Pedersen
Aarhus University, Aarhus, Central Denmark Region, Denmark

This data note describes the reference genome of *Abrostola triplasia.* The analyses are sound and well described. I am sure that this data can be used for exciting analyses of decline and inter-specific competition with the *A. tripartita*. If the authors could outline some of the questions this data could be used to address in the Background section, instead of saying 'A genome sequence for *Abrostola triplasia* will contribute to a growing data set of resources for understanding lepidopteran biology', which does not really say anything, it would be very interesting.

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

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: Population genetics, molecular evolution

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