DATA NOTE

The genome sequence of the Cabbage Moth, *Mamestra brassicae* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Amaia Green Etxabe[1], Helen Hesketh[1], Stephen Short[1], Alex Robinson[1], David Spurgeon[1], 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

v1 **First published:** 23 Oct 2023, **8**:486
https://doi.org/10.12688/wellcomeopenres.20165.1
Latest published: 23 Oct 2023, **8**:486
https://doi.org/10.12688/wellcomeopenres.20165.1

Abstract

We present a genome assembly from an individual male *Mamestra brassicae* (the Cabbage Moth; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 576.2 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 15.38 kilobases in length. Gene annotation of this assembly on Ensembl identified 12,891 protein coding genes.

Keywords

Mamestra brassicae, cabbage moth, 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: Green Etxabe A: Investigation, Resources; **Hesketh H**: Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Short S**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Robinson A**: Investigation, Resources; **Spurgeon D**: Resources, 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 work was also supported by research grants NE/M01438X/1, NE/S000135/1 and NE/S000224/2 from the UK Natural Environment Research Council (part of UK Research and Innovation), and Defra (Department for Environment, Food and Rural affairs, UK) as project HH3101TX.
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: Green Etxabe A, Hesketh H, Short S *et al.* **The genome sequence of the Cabbage Moth, *Mamestra brassicae* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2023, **8**:486 https://doi.org/10.12688/wellcomeopenres.20165.1

First published: 23 Oct 2023, **8**:486 https://doi.org/10.12688/wellcomeopenres.20165.1

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; Hadeninae; *Mamestra*; *Mamestra brassicae* (Linnaeus, 1758).

Background

The Cabbage Moth, *Mamestra brassicae* (Linnaeus, 1758) is a common moth, found throughout most of Britain and the Channel Islands. It is recorded less frequently in western Scotland, and in Ireland. In the UK, the moths are best known as a pest of cultivated brassicas, hence the common name. However, the larvae will feed on leaves of many wild or cultivated herbaceous plants, and have also been found on woody species such as sallow and oak (Waring *et al.*, 2017). Larvae develop through six instars, the first five of which have a coppercoloured head, with a light green abdomen and a white stripe along the stigmata (Devetak *et al.*, 2010). The final instar larvae have a dusky dorsal stripe, speckled with white and a yellow/green or dusky brown strip down both sides, still with the copper-coloured head.

The adult moth is a dark brownish grey, although sometimes is paler or even blackish, often with brown blotches. The forewings have a span of 35–50 mm. A distinguishable characteristic on the forewing is a chalky-white mark that resembles a kidney shape, surrounded by a white border (Figure 1). Adults fly from May to October in the UK and there are one to two generations of larvae, per year. *Mamestra brassicae* has an annual life cycle. The moth overwinters underground in the pupal stage, and sometimes as a larva. This is mainly during the later summer and autumn. The species is not considered as being under threat, and is not listed on the International Union for Conservation of Nature (IUCN) red list (Freyhof, 2014).

Mamestra brassicae has been cultured for over 40 years, and is used as an important research species for the ecological study of host-parasite interactions (Burden *et al.*, 2003). A laboratory population maintained at UKCEH (UK Centre for Ecology & Hydrology) is known to harbour persistent, covert viruses (Hughes *et al.*, 1993). As the moth is also a prevalent pest of cultivated crops, other studies have focussed on the opportunity for biological control, using baculoviruses as biopesticides (Hesketh & Hails, 2015). More recently, *M. brassicae* has been used for developing high through-put bioassay methods for toxicity testing in Lepidoptera (Badder *et al.*, 2023).

Figure 1. Photograph of adult *Mamestra brassicae* by Olaf Leillinger (not the specimen used for genome sequencing).

This genome will provide further insights into lepidopterans, a group that play important ecosystem roles and are recognised as an ecologically relevant indicator of environmental health (Goldstein, 2017). More specifically, the current research groups that utilise *M. brassicae* will benefit, leading to a greater understanding of the molecular biology underlying both host-pathogen interactions and the sensitivity of beneficial non-target species to anthropogenic contamination.

We present a chromosomally complete genome sequence for the moth, *Mamestra brassicae*, generated using a specimen taken from a laboratory culture kept at the UKCEH (Wallingford), 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 *Mamestra brassicae* grown in culture at UKCEH, Wallingford, Oxfordshire, UK. A total of 35-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 61-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 18 missing joins or mis-joins, reducing the scaffold number by 23.64%.

The final assembly has a total length of 576.2 Mb in 42 sequence scaffolds with a scaffold N50 of 19.4 Mb (Table 1). A summary of the assembly statistics is shown in Figure 2, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.95%) 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 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 56.9 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 98.9% (single = 98.5%, duplicated = 0.5%), using the lepidoptera_odb10 reference set (n = 5,286).

Table 1. Genome data for *Mamestra brassicae*, ilMamBras1.1.

Project accession data		
Assembly identifier	ilMamBras1.1	
Assembly release date	2021-01-18	
Species	*Mamestra brassicae*	
Specimen	ilMamBras1	
NCBI taxonomy ID	55057	
BioProject	PRJEB42134	
BioSample ID	SAMEA7524129	
Isolate information	ilMamBras1, male: mid-body (DNA sequencing and Hi-C data) ilMamBras2 and ilMamBras4: mid-body (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	56.9	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:98.9%[S:98.5%,D:0.5%], F:0.2%,M:0.8%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.95%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembles	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6412357	
10X Genomics Illumina	ERR6002683, ERR6002681, ERR6002682, ERR6002684	
Hi-C Illumina	ERR6002679, ERR6002680, ERR6002678	
PolyA RNA-Seq Illumina	ERR6002685, ERR6787417	
Genome assembly		
Assembly accession	GCA_905163435.1	
Accession of alternate haplotype	GCA_905163405.1	
Span (Mb)	576.2	
Number of contigs	70	
Contig N50 length (Mb)	17.8	
Number of scaffolds	42	
Scaffold N50 length (Mb)	19.4	
Longest scaffold (Mb)	35.5	
Genome annotation		
Number of protein-coding genes	12,891	
Number of non-coding genes	1,528	
Number of gene transcripts	21,647	

* 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/Mamestra%20brassicae/dataset/CAJHZQ01.1/busco.

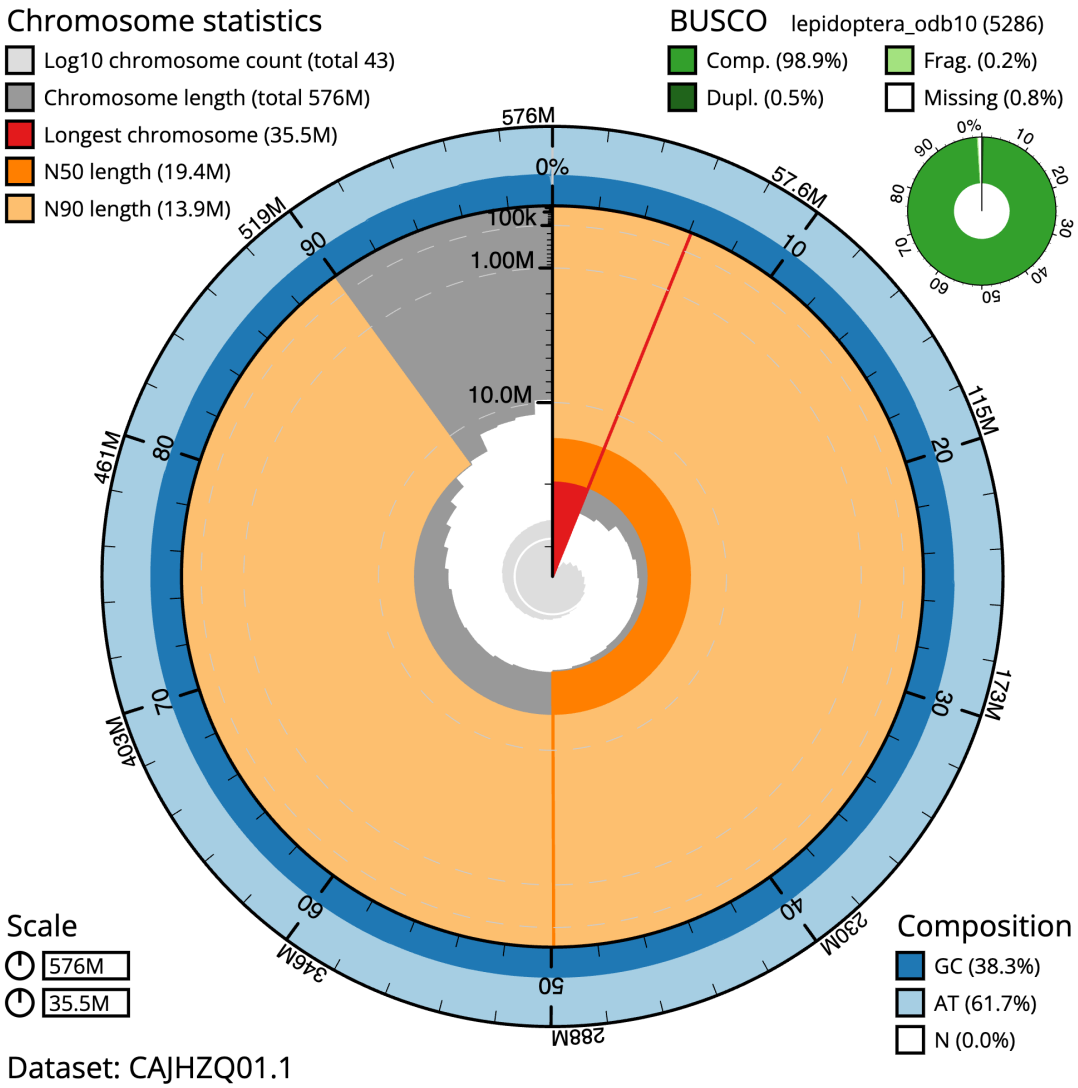

Figure 2. Genome assembly of *Mamestra brassicae*, ilMamBras1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 576,184,143 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (35,467,487 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (19,384,296 and 13,925,230 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/Mamestra%20brassicae/dataset/CAJHZQ01.1/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/55057.

Genome annotation report

The *Mamestra brassicae* genome assembly (GCA_905163435.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Mamestra_brassicae_GCA_905163435.1/Info/Index). The resulting annotation includes 21,647 transcribed mRNAs from 12,891 protein-coding and 1,528 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

Mamestra brassicae specimens were harvested from a culture at UKCEH, Wallingford, Oxfordshire, UK on 2020-03-17. The specimens were cultured by Helen Hesketh and Alex Robinson (UKCEH) and collected by Stephen Short and Amaia

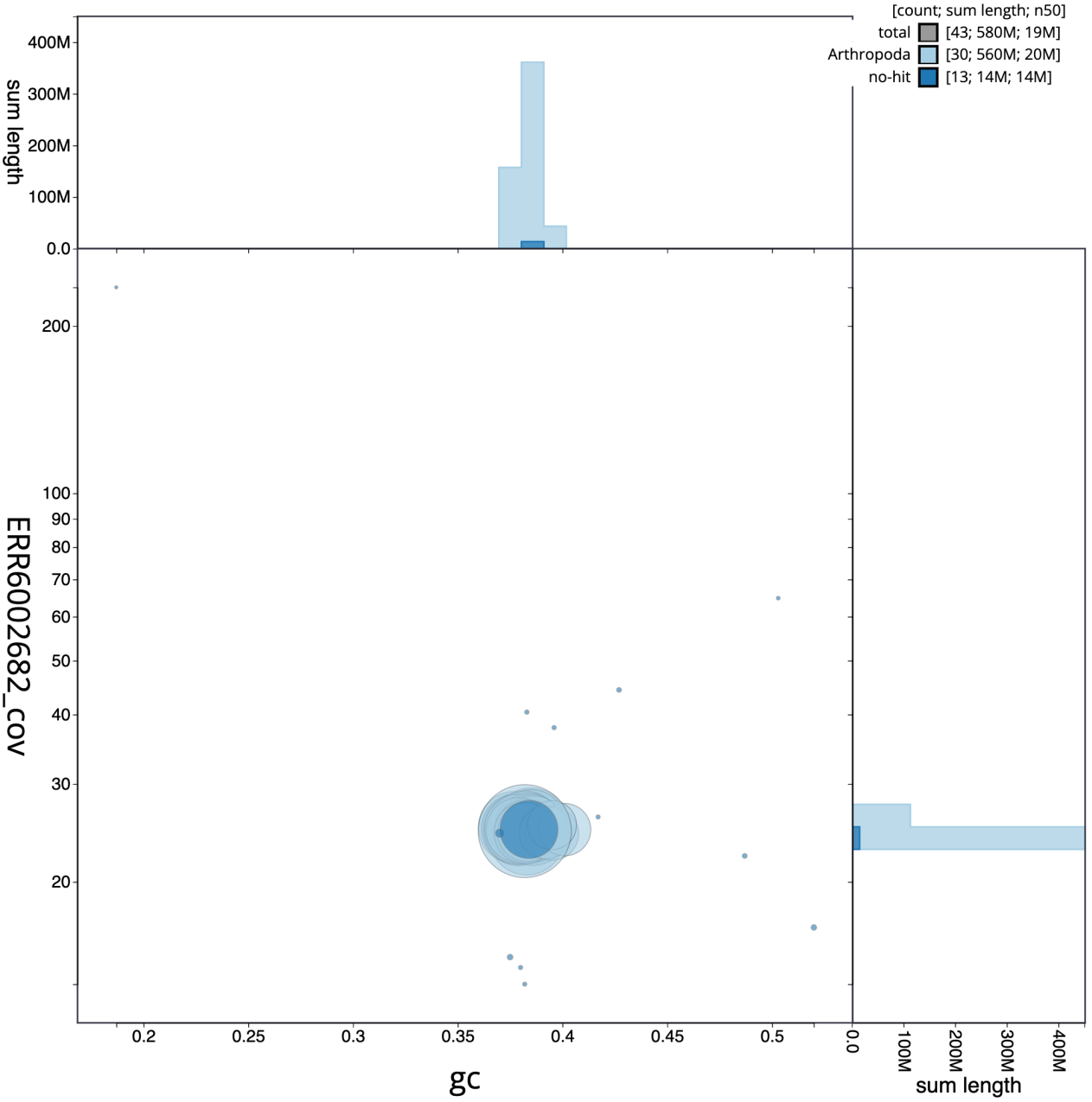

Figure 3. Genome assembly of *Mamestra brassicae*, ilMamBras1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Mamestra%20brassicae/dataset/CAJHZQ01.1/blob.

Green Etxabe (UKCEH). The specimens were identified by Amaia Green Etxabe, and flash frozen in liquid nitrogen. The specimen used for DNA sequencing and Hi-C data was a male with specimen ID SAN0001209 (ToLID ilMamBras1), while specimens with IDs SAN0001210 (ToLID ilMamBras2) and SAN0001212 (ToLID ilMamBras4) were used for RNA sequencing.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilMamBras1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Mid-body 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

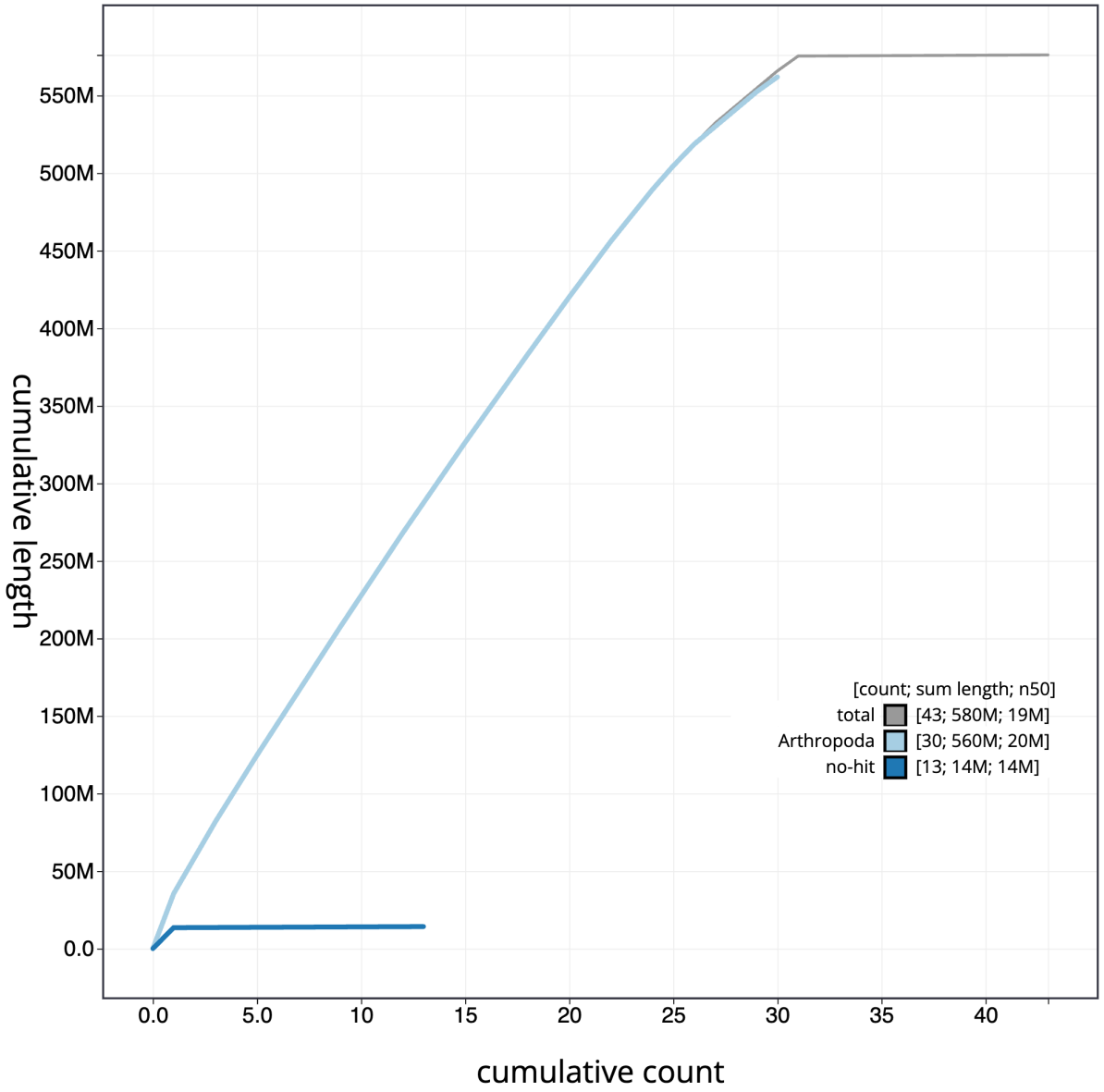

Figure 4. Genome assembly of *Mamestra brassicae*, ilMamBras1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Mamestra%20brassicae/dataset/CAJHZQ01.1/cumulative.

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.

RNA was extracted from mid-body tissue of ilMamBras2 and lMamBras4 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

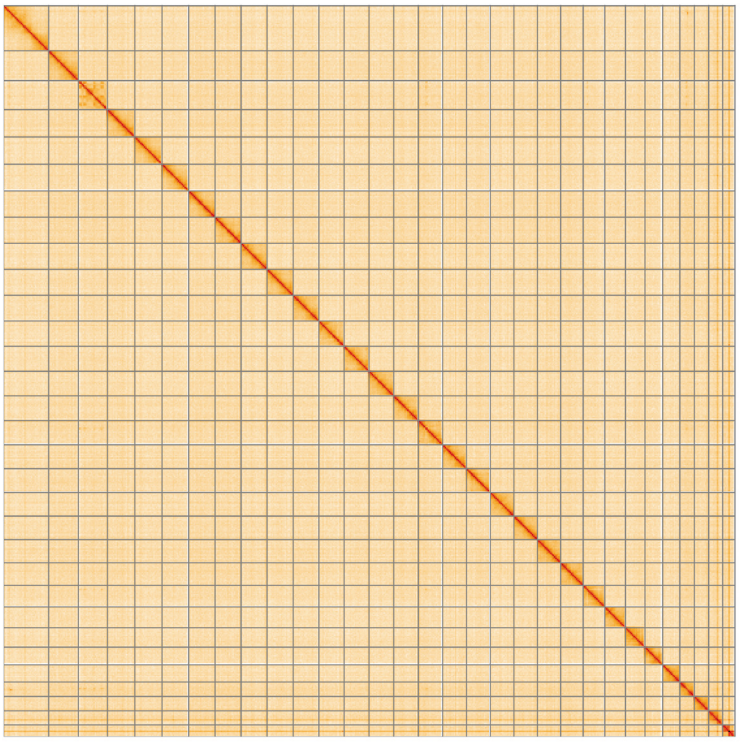

Figure 5. Genome assembly of *Mamestra brassicae*, ilMamBras1.1: Hi-C contact map of the ilMamBras1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=KZHVM3MXSLWhyUNeHYKxfw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Mamestra brassicae*, ilMamBras1.

INSDC accession	Chromosome	Length (Mb)	GC%
LR990988.1	1	23.41	38.5
LR990989.1	2	22.88	38.5
LR990990.1	3	21.53	38.5
LR990991.1	4	21.4	38.0
LR990992.1	5	20.95	37.5
LR990993.1	6	20.72	38.5
LR990994.1	7	20.52	38.5
LR990995.1	8	20.49	38.0
LR990996.1	9	20.45	38.0
LR990997.1	10	20.25	38.0
LR990998.1	11	19.89	38.0
LR990999.1	12	19.68	38.0
LR991000.1	13	19.38	38.0
LR991001.1	14	19.34	38.0
LR991002.1	15	19.12	38.5
LR991003.1	16	18.8	38.5
LR991004.1	17	18.76	38.0
LR991005.1	18	18.55	38.0
LR991006.1	19	18.54	38.5
LR991007.1	20	18.21	38.5
LR991008.1	21	17.7	38.5
LR991009.1	22	17.06	38.0
LR991010.1	23	16.32	38.5
LR991011.1	24	15.29	38.0
LR991012.1	25	13.93	38.5
LR991013.1	26	13.5	38.5
LR991014.1	27	11.43	39.5
LR991015.1	28	11.28	39.0
LR991016.1	29	11.04	40.0
LR991017.1	30	9.6	39.5
LR990987.1	Z	35.47	38.0
LR991018.1	MT	0.02	18.5

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 and 10X Genomics read cloud 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), Illumina HiSeq 4000 (RNA-Seq) and HiSeq X Ten (10X) instruments. Hi-C data were also generated from remaining mid-body tissue of ilMamBras1 using the Arima2 kit and sequenced on the HiSeq X Ten 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). 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 gene annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Mamestra brassicae* assembly (GCA_905163435.1). 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).

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/
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

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.

Data availability

European Nucleotide Archive: *Mamestra brassicae* (cabbage moth). Accession number PRJEB42134; https://identifiers.org/ena.embl/PRJEB42134. (Wellcome Sanger Institute, 2021) The genome sequence is released openly for reuse. The *Mamestra brassicae* 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 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

Badder C, Bart S, Robinson A, *et al.*: **A novel Lepidoptera bioassay analysed using a reduced GUTS model.** *Ecotoxicol Environ Saf.* 2023; **251**: 114504.
PubMed Abstract | Publisher 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

Burden JP, Nixon CP, Hodgkinson AE, *et al.*: **Covert infections as a mechanism for long-term persistence of baculoviruses.** *Ecology Letters.* 2003; **6**(6): 524–531.
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): 2508–10.
PubMed Abstract | Publisher Full Text | Free Full Text

Devetak M, Vidrih M, Trdan S: **Cabbage moth (*Mamestra brassicae* [L.]) and bright-line brown-eyes moth (*Mamestra oleracea* [L.]) - presentation of the species, their monitoring and control measures.** *Acta Agric Slov.* 2010; **95**(2).
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

Freyhof J: **Pleuronectes platessa.** The IUCN Red List of Threatened Species 2014: e.T135690A50018800. 2014; [Accessed 4 April 2023].
Reference Source

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.
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

Goldstein PZ: **Diversity and Significance of Lepidoptera: A Phylogenetic Perspective.** In: *Insect Biodiversity.* Wiley, 2017; 463–495.
Publisher Full Text

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

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

Hesketh H, Hails RS: ***Bacillus thuringiensis* impacts on primary and secondary baculovirus transmission dynamics in Lepidoptera.** *J Invertebr Pathol.* 2015; **132**: 171–181.
PubMed Abstract | Publisher Full Text

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

Hughes DS, Possee RD, King LA: **Activation and detection of a latent baculovirus resembling *Mamestra brassicae* nuclear polyhedrosis virus in *M. brassicae* insects.** *Virology.* 1993; **194**(2): 608–615.
PubMed Abstract | Publisher Full Text

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

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

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | Publisher Full Text | Free Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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].
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

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 Cabbage Moth, *Mamestra brassicae* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB42134, 2021.