DATA NOTE

The genome sequence of the Diamondback Moth, *Plutella xylostella* (Linnaeus, 1758) [version 1; peer review: 3 approved]

Douglas Boyes[1+],
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

[+] Deceased author

v1 **First published:** 18 Sep 2023, **8**:404
https://doi.org/10.12688/wellcomeopenres.20006.1
Latest published: 18 Sep 2023, **8**:404
https://doi.org/10.12688/wellcomeopenres.20006.1

Abstract

We present a genome assembly from an individual male *Plutella xylostella* (the Diamondback Moth; Arthropoda; Insecta; Lepidoptera; Plutellidae). The genome sequence is 323.3 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 35.12 kilobases in length. Gene annotation of this assembly on Ensembl identified 17,190 protein coding genes.

Keywords

Plutella xylostella, diamondback moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 18 Sep 2023	✓ view	✓ view	✓ view

1. **Jerome H L Hui**, The Chinese University of Hong Kong, Hong Kong, Hong Kong
2. **Kyle M Benowitz**, Austin Peay State University (Ringgold ID: 2536), Clarksville, USA
3. **Muzafar Riyaz**, Xavier Research FoundationSt. Xavier's College, Palayamkottai, India

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;

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, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective *et al.* **The genome sequence of the Diamondback Moth, *Plutella xylostella* (Linnaeus, 1758) [version 1; peer review: 3 approved]** Wellcome Open Research 2023, **8**:404 https://doi.org/10.12688/wellcomeopenres.20006.1

First published: 18 Sep 2023, **8**:404 https://doi.org/10.12688/wellcomeopenres.20006.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Yponomeutoidea; Plutellidae; *Plutella*; *Plutella xylostella* (Linnaeus, 1758) (NCBI:txid51655).

Background

The Diamondback Moth, *Plutella xylostella*, is a micro-moth in the Plutellidae family, previously in the family Yponomeutidae. Members of this family are often characterised by elongated forewings with a distinctive shape (Sterling & Parsons, 2018). Although there are only seven species, the Plutellidae family has significant relevance in agriculture, as the Diamondback Moth is a notorious global pest. This moth is common in Britain and Ireland, arriving on the shores in great numbers. The adults fly by day and come to light. There are several broods each year, more in warmer areas (Sterling & Parsons, 2018).

The P. xylostella larva is the main pest of cruciferous crops worldwide (Zalucki *et al.*, 2012). It is the most widely distributed of all lepidopteran pests (Talekar & Shelton, 1993), and indeed has the widest distribution of all Lepidoptera (Furlong *et al.*, 2013). The annual cost of losses of crop production to *P. xylostella* infestations have been estimated at up to 5 billion USD (Furlong *et al.*, 2013). It has also rapidly evolved field resistance to all major classes of synthetic and biological insecticides (Furlong *et al.*, 2013) through mutations in insecticidal receptors, including Bt toxins (Baxter *et al.*, 2011), and overexpression of detoxification genes. Pesticide resistance might also be conferred by the gut microbiota (Xia *et al.*, 2013; Xia *et al.*, 2018).

To understand the capacity of this moth to respond to environmental stressors and develop resistance to insecticides, it has been the target of many molecular studies. In 2013, two parallel projects reported the first *P. xylostella* reference genomes, also establishing a genomic and transcriptomic database (Jouraku *et al.*, 2013; You *et al.*, 2013). A chromosome-level haploid genome assembly was generated by a trio binning strategy (Ward *et al.*, 2021). An analysis of 532 genomes through resequencing and variation analysis provided evidence that *P. xylostella* originated in South America and expanded throughout the world in three major expansions (You *et al.*, 2020).

The genome of *P. xylostella* was sequenced using the Darwin Tree of Life pipeline, a part of the broader collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. This methodological approach represents a standardised procedure ensuring quality and consistency in genomic analysis. Here we present a chromosomally complete genome sequence for *P. xylostella*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from one male *Plutella xylostella* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (51.77, –1.32). A total of 77-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 148-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 8 missing joins or mis-joins and removed 2 haplotypic duplications, reducing the assembly length by 0.11% and the scaffold number by 5.71%, and increasing the scaffold N50 by 1.97%.

The final assembly has a total length of 323.3 Mb in 33 sequence scaffolds with a scaffold N50 of 11.3 Mb (Table 1). Most (99.98%) 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 according to synteny with *P. xylostella* genome assembly GCA_019096205.1 (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 63.2 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.0% (single = 97.5%, duplicated = 0.5%), 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/51655.

Genome annotation report

The *Plutella xylostella* genome assembly (GCA_932276165.1) was annotated using the Ensembl rapid annotation pipeline

Figure 1. Photograph of the *Plutella xylostella* (ilPluXylo3) specimen used for genome sequencing.

Table 1. **Genome data for *Plutella xylostella*, ilPluXylo3.1.**

Project accession data		
Assembly identifier	ilPluXylo3.1	
Species	*Plutella xylostella*	
Specimen	ilPluXylo3	
NCBI taxonomy ID	51655	
BioProject	PRJEB48401	
BioSample ID	SAMEA7520369	
Isolate information	ilPluXylo3, male: whole organism (DNA sequencing) ilPluXylo4: whole organism (Hi-C scaffolding)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	63.2	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:98.0%[S:97.5%,D:0.5%], F:0.8%,M:1.2%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.98%	*≥ 95%*
Sex chromosomes	Z chromosome	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR7224286	
10X Genomics Illumina	ERR7220498, ERR7220499, ERR7220500, ERR7220501 ERR7220505, ERR7220502, ERR7220503, ERR7220504	
Hi-C Illumina	ERR7220506	
Genome assembly		
Assembly accession	GCA_932276165.1	
Accession of alternate haplotype	GCA_932276175.1	
Span (Mb)	323.3	
Number of contigs	40	
Contig N50 length (Mb)	11.0	
Number of scaffolds	33	
Scaffold N50 length (Mb)	11.3	
Longest scaffold (Mb)	16.2	
Genome annotation		
Number of protein-coding genes	17,190	
Number of non-coding genes	12,680	
Number of gene transcripts	49,308	

* 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/Plutella%20xylostella/dataset/CAKNZZ01/busco.

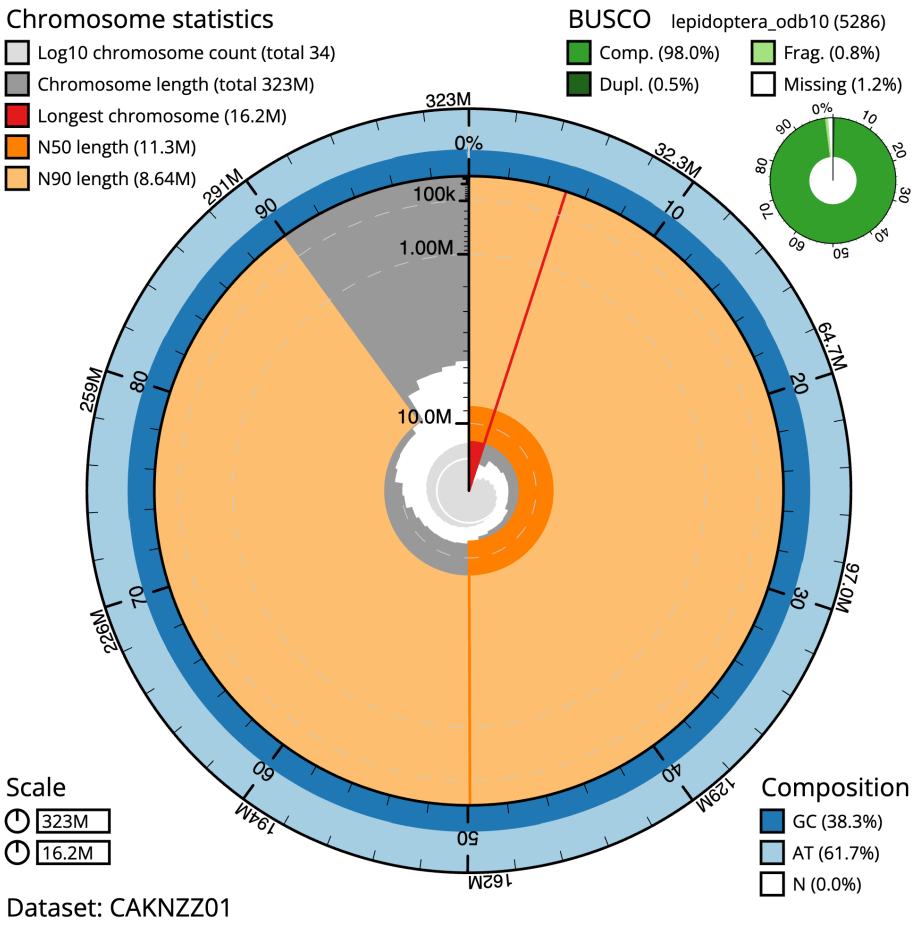

Figure 2. Genome assembly of *Plutella xylostella*, ilPluXylo3.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 323,337,879 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (16,174,618 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (11,339,310 and 8,635,333 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/Plutella%20xylostella/dataset/CAKNZZ01/snail.

(Table 1; https://rapid.ensembl.org/Plutella_xylostella_GCA_932276165.1/Info/Index). The resulting annotation includes 49,308 transcribed mRNAs from 17,190 protein-coding and 12,680 non-coding genes.

Methods

Sample acquisition and nucleic acid extraction

The specimens used in this study were collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire), UK (latitude 51.77, longitude –1.32) on 2019-09-21, using a light trap. Douglas Boyes (University of Oxford) collected and identified the specimens. The specimens were snap-frozen on dry ice. The specimen used for genome sequencing was a male *Plutella xylostella* (specimen ID Ox000293, ToLID ilPluXylo3) while the specimen used for Hi-C sequencing had specimen ID Ox000294 (ToLID ilPluXylo4).

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The ilPluXylo3 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

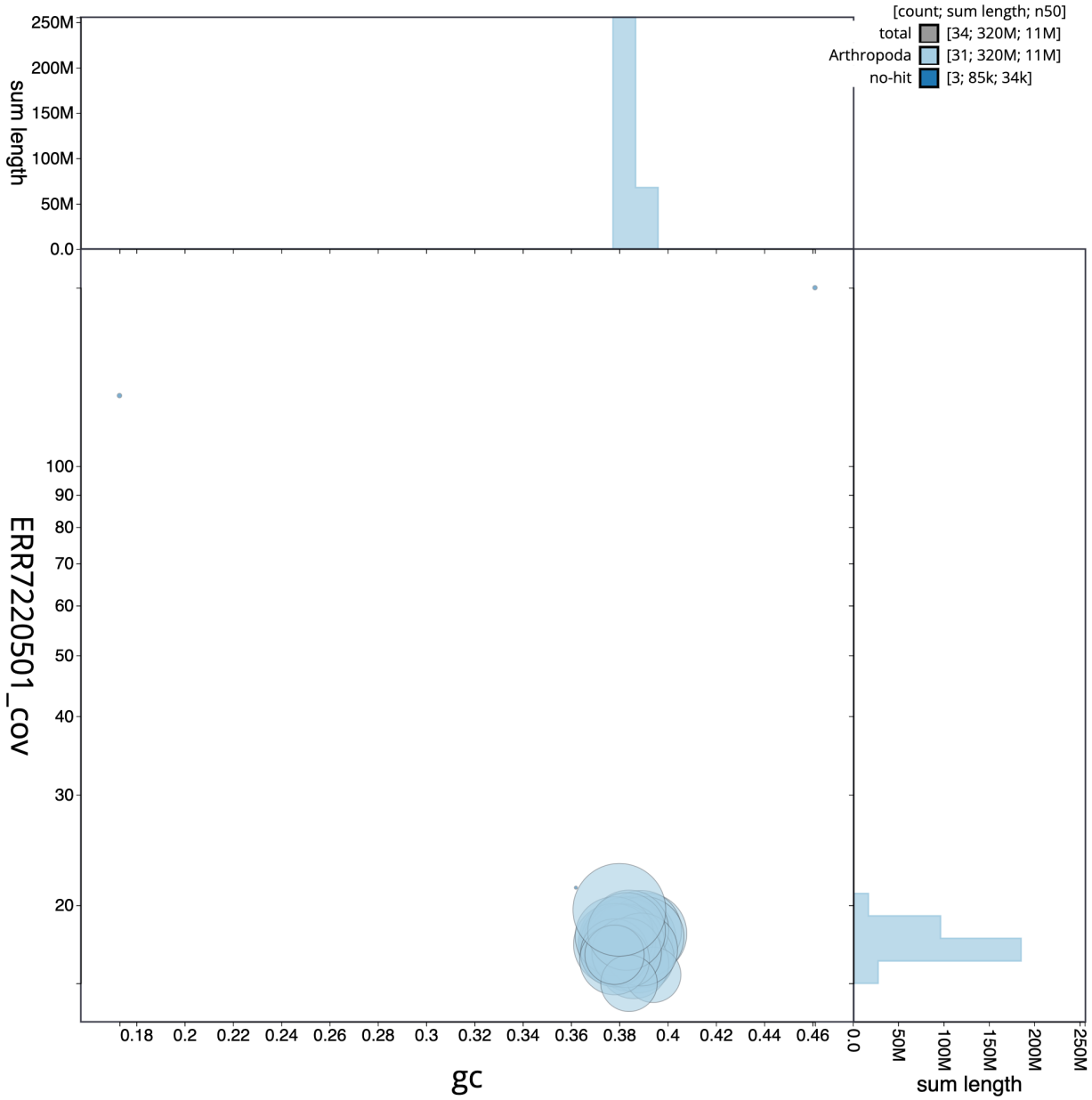

Figure 3. Genome assembly of *Plutella xylostella*, ilPluXylo3.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Plutella%20xylostella/dataset/CAKNZZ01/blob.

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 HiSeq X Ten (10X) instruments. Hi-C data were also generated from the whole organism tissue of ilPluXylo4 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

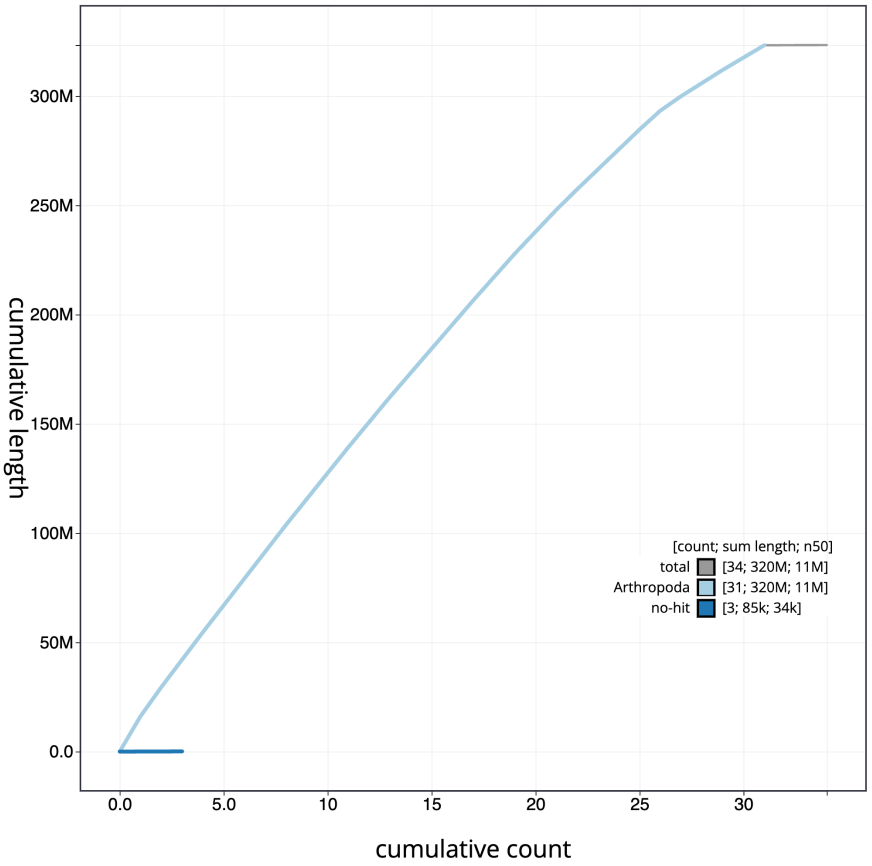

Figure 4. Genome assembly of *Plutella xylostella*, ilPluXylo3.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Plutella%20xylostella/dataset/CAKNZZ01/cumulative.

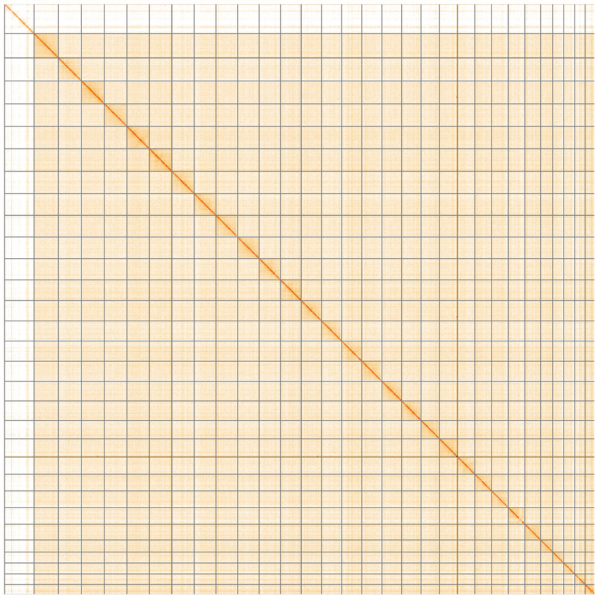

Figure 5. Genome assembly of *Plutella xylostella*, ilPluXylo3.1: Hi-C contact map of the ilPluXylo3.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=TaoX1aegQfituIk_tZJ_8Q.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Plutella xylostella*, ilPluXylo3.

INSDC accession	Chromosome	Length (Mb)	GC%
OW026580.1	2	6.02	38.5
OW026569.1	3	10.81	38.0
OW026557.1	4	12.54	39.0
OW026554.1	5	13.32	38.5
OW026560.1	6	12.23	38.0
OW026571.1	7	10.13	38.0
OW026565.1	8	11.12	38.5
OW026558.1	9	12.34	38.0
OW026559.1	10	12.28	39.0
OW026583.1	11	5.53	39.0
OW026561.1	12	11.9	38.0
OW026562.1	13	11.82	38.5
OW026567.1	14	11.0	38.0
OW026555.1	15	12.29	39.0
OW026574.1	16	9.35	39.0
OW026572.1	17	11.82	38.5
OW026566.1	18	11.05	38.0
OW026570.1	19	10.6	38.5
OW026575.1	20	9.17	38.5
OW026568.1	21	10.96	38.0
OW026556.1	22	12.6	38.0
OW026563.1	23	11.65	38.0
OW026582.1	24	5.75	39.5
OW026573.1	25	9.98	39.0
OW026577.1	26	8.64	38.0
OW026578.1	27	9.08	38.5
OW026576.1	28	9.1	38.0
OW026564.1	29	11.34	38.5
OW026581.1	30	5.99	38.5
OW026579.1	31	6.62	38.0
OW026553.1	Z	16.17	38.0
OW026584.1	MT	0.04	17.5

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 *Plutella xylostella* assembly (GCA_932276165.1). 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.

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	0.15.3	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	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

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: *Plutella xylostella* (diamondback moth). Accession number PRJEB48401; https://identifiers.org/ena.embl/PRJEB48401. (Wellcome Sanger Institute, 2021)

The genome sequence is released openly for reuse. The *Plutella xylostella* 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

Baxter SW, Badenes-Pérez FR, Morrison A, *et al.*: **Parallel Evolution of *Bacillus thuringiensis* Toxin Resistance in Lepidoptera.** *Genetics.* 2011; **189**(2): 675–679.
PubMed Abstract | Publisher Full Text | Free Full Text

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
PubMed Abstract | Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL - a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–10.
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

Furlong MJ, Wright DJ, Dosdall LM: **Diamondback Moth Ecology and Management: Problems, Progress, and Prospects.** *Annu Rev Entomol.* 2013; **58**(1): 517–541.
PubMed Abstract | Publisher Full Text

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

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

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

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

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

Jouraku A, Yamamoto K, Kuwazaki S, *et al.*: **KONAGAbase: a genomic and transcriptomic database for the diamondback moth, *Plutella xylostella.*** *BMC Genomics.* 2013; **14**(1): 464.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

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

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

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | Publisher Full Text

Sterling P, Parsons M: **Field Guide to the Micro-moths of Great Britain and Ireland.** London: Bloomsbury, 2018.

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a; [Accessed 21 July 2023].
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023].
Reference Source

Talekar NS, Shelton AM: **Biology, Ecology, and Management of the Diamondback Moth.** *Annu Rev Entomol.* 1993; **38**(1): 275–301.
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 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

Ward CM, Perry KD, Baker G, *et al.*: **A haploid diamondback moth (*Plutella xylostella* L.) genome assembly resolves 31 chromosomes and identifies a diamide resistance mutation.** *Insect Biochem Mol Biol.* 2021; **138**: 103622.
PubMed Abstract | Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the Diamondback Moth, *Plutella xylostella* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB48401, 2021.

Xia X, Sun B, Gurr GM, *et al.*: **Gut Microbiota Mediate Insecticide Resistance in the Diamondback Moth, *Plutella xylostella* (L.).** *Front Microbiol.* 2018; **9**: 25.
PubMed Abstract | Publisher Full Text | Free Full Text

Xia X, Zheng D, Zhong H, *et al.*: **DNA Sequencing Reveals the Midgut Microbiota of Diamondback Moth, *Plutella xylostella* (L.) and a Possible Relationship with Insecticide Resistance.** *PLoS One.* 2013; **8**(7): e68852.
PubMed Abstract | Publisher Full Text | Free Full Text

You M, Ke F, You S, *et al.*: **Variation among 532 genomes unveils the origin and evolutionary history of a global insect herbivore.** *Nat Commun.* 2020; **11**(1): 2321.
PubMed Abstract | Publisher Full Text | Free Full Text

You M, Yue Z, He W, *et al.*: **A heterozygous moth genome provides insights into herbivory and detoxification.** *Nat Genet.* 2013; **45**(2): 220–225.
PubMed Abstract | Publisher Full Text

Zalucki MP, Shabbir A, Silva R, *et al.*: **Estimating the Economic Cost of One of the World's Major Insect Pests, *Plutella xylostella* (Lepidoptera: Plutellidae): Just How Long Is a Piece of String?** *J Econ Entomol.* 2012; **105**(4): 1115–1129.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 1

Reviewer Report 10 May 2024

https://doi.org/10.21956/wellcomeopenres.22152.r78120

✔ **Muzafar Riyaz**
Xavier Research FoundationSt. Xavier's College, Palayamkottai, Tamil Nadu, India

The paper presents a comprehensive account of the genomic sequencing and annotation of the Diamondback Moth, Plutella xylostella (Linnaeus, 1758), a significant agricultural pest. Through detailed descriptions of sample collection, sequencing methodologies, assembly techniques, and annotation processes, the study achieves a high-quality genome assembly with chromosome-level scaffolding and accurate gene annotations. The background section provides context on the pest's agricultural significance, referencing previous molecular studies, while the methods section outlines experimental procedures and software tools utilized, ensuring reproducibility and transparency. Figures and tables have aided in visualizing assembly metrics, and the data availability statement provides accession numbers for raw sequence data and the genome assembly, facilitating accessibility for further research. The extensive list of references demonstrates a thorough review of relevant literature, supporting the study's findings within the broader scientific context. Overall, this paper contributes valuable genomic resources for P. xylostella research, aiding in pest management strategies and advancing understanding in insect genomics.

References

1. Boyesh D: The genome sequence of the Diamondback Moth, Plutella xylostella. *researchgate.net*. 2023.

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: Insect genomics, Moths, Phylogenetics

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 09 May 2024

https://doi.org/10.21956/wellcomeopenres.22152.r82027

Kyle M Benowitz
Austin Peay State University (Ringgold ID: 2536), Clarksville, Tennessee, USA

The paper reports a new chromosome-level genome assembly for the lepidopteran pest *Plutella xylostella*. The assembly and annotation are of high quality, and the genome should be useful for other researchers in the field. The methods are current and easy to follow. Overall, the paper is well written, and I just have a few minor comments.

Minor comments:

1. There's a small confusion in the methods, as it is stated in the 2nd paragraph of the methods that tissue from ilPluXylo3 was 'set aside for Hi-C sequencing'. However, in all other parts of the report it is made clear that tissue from a separate individual, ilPluXylo4, was used for Hi-C.

2. In the methods for genome annotation, what transcriptomic data was used? What tissues/life stages were it from, and where can that data be found? This is important for the replicability of the work.

3. Given that a chromosome-level assembly for this species already exists, a brief sentence comparing the two assemblies could be added. Since this paper is just a data note, a large amount of comparative analysis or interpretation would be inappropriate; however, a quick statement noting the similarity in size/structure would be useful for readers.

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?

Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I work in the field of insect evolutionary genomics. I have been involved with several genome sequencing projects, including that of the lepidopteran pest Helicoverpa zea.

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 08 May 2024

https://doi.org/10.21956/wellcomeopenres.22152.r80234

Jerome H L Hui
The Chinese University of Hong Kong, Hong Kong, Hong Kong

Boyes and colleagues report the genome sequence of a male diamondblack/cabbage moth *Plutella xylostella* (Linnaeus, 1758). As one can judge from its common name, it is an important pest of brassicaceous crops such as cabbage. This species is well known for its migratory tendencies and now become a cosmopolitan species. In Britain, the immediate sources of this species have been suggested to be coming from countries in the western part of continental Europe (Wainwright et al 2020 Insects). As the authors have pointed out, certain level of molecular data of this species have already been obtained previously, including transcriptomes and draft genomes (N50 = 2.2 kb and 737kb from two studies, Jouraku et al 2013 BMC Genomics [Ref 2]; You et al 2013 Nature Genetics [Ref 3]). Nevertheless, it is necessary to point out that this report contains a much improved high-quality genome resource. It will be useful for a range of further studies, including from not limited to, revealing their population structures, dissecting host plant-insect relationships, to understanding their evolution with other lepidopterans, which have important values in both basic knowledge and applications.

This genome resource is excellent from the summary statistics, with high BUSCO numbers, high sequence continuity (scaffold N50), and majority of sequences contained on the 32 pseudochromosomes (plus mitochondrion). All in all, this is another valuable contribution by the Darwin Tree of Life Consortium.

References

1. Jouraku A, Yamamoto K, Kuwazaki S, Urio M, et al.: KONAGAbase: a genomic and transcriptomic database for the diamondback moth, Plutella xylostella.*BMC Genomics*. 2013; **14**: 464 PubMed Abstract | Publisher Full Text

2. You M, Yue Z, He W, Yang X, et al.: A heterozygous moth genome provides insights into herbivory and detoxification.*Nat Genet*. 2013; **45** (2): 220-5 PubMed Abstract | Publisher Full Text

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: I have published with Peter Holland more than three years ago, and confirm that this potential conflict of interest did not affect my ability to write an objective and unbiased review of the article.

Reviewer Expertise: Genomics, evolution, invertebrates

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.