DATA NOTE

The genome sequence of the Box-tree Moth, *Cydalima perspectalis* (Walker, 1859) [version 1; peer review: awaiting peer review]

Gavin R. Broad[1], Douglas Boyes[2+], Riccardo Poloni[3],
Natural History Museum Genome Acquisition Lab,
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Natural History Museum, London, England, UK
[2]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[3]CIRB, Collège de France, Université PSL, Paris, France

[+] Deceased author

v1 **First published:** 16 May 2024, **9**:272
https://doi.org/10.12688/wellcomeopenres.21678.1
Latest published: 16 May 2024, **9**:272
https://doi.org/10.12688/wellcomeopenres.21678.1

Abstract

We present a genome assembly from an individual female *Cydalima perspectalis* (the Box-tree Moth; Arthropoda; Insecta; Lepidoptera; Crambidae). The genome sequence is 483.7 megabases in span. Most of the assembly is scaffolded into 32 chromosomal pseudomolecules, including the Z and W sex chromosomes. The mitochondrial genome has also been assembled and is 15.25 kilobases in length.

Keywords

Cydalima perspectalis, Box-tree 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: Broad GR: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Boyes D**: Investigation, Resources; **Poloni R**: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
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: Broad GR, Boyes D, Poloni R *et al.* **The genome sequence of the Box-tree Moth, *Cydalima perspectalis* (Walker, 1859) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:272 https://doi.org/10.12688/wellcomeopenres.21678.1

First published: 16 May 2024, **9**:272 https://doi.org/10.12688/wellcomeopenres.21678.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Pyraloidea; Crambidae; Spilomelinae; *Cydalima*; *Cydalima perspectalis* (Walker, 1859) (NCBI:txid1309588).

Background

Cydalima perspectalis, the Box-tree Moth, is a native of subtropical Asia which has rapidly become a very familiar species in Europe. This is a large and conspicuous moth of the family Crambidae, with a wingspan of around 40 mm and, in the typical colour form, a black border to shining white wings. A frequent melanic form is basically dark brown with a white discal spot on the forewings, and seems to have an advantage in the face of predation (Poloni *et al.*, 2024), whereas a less common form resembles the typical colour form but with a dark brown band on the anal side of the fore wings (Cook & Muggleton, 2023). Larvae eat Box (*Buxus* species), and have become pests, defoliating large quantities of *Buxus*. Unmistakable in its non-native range, *C. perspectalis* is one of nine known species of *Cydalima* found in Asia and Australia (Mally & Nuss, 2010).

The establishment and spread of *C. perspectalis* in Britain were summarised by Plant *et al.* (2019) up the end of 2018. In summary, not long after the first sightings in Europe, presumably a result of the horticulture trade (Bras *et al.*, 2019), the first British sighting was in Kent in 2007 followed by scattered records until 2014, after which there has been a spectacular increase in abundance and range. Although particularly numerous in south-east England, *C. perspectalis* can be found as far north as Scotland. In GRB's garden, *C. perspectalis* is one of the most frequently light-trapped moths. As in its native range, there are multiple, overlapping generations per year, with adult moths on the wing from May to October (Plant *et al.*, 2019).

Although in Britain the main effect of *C. perspectalis* is on ornamental Boxes, this does amount to significant economic impacts, there are native *Buxus* here, and damage to native *Buxus* forests elsewhere in Europe can be extensive, resulting in the death of many trees (Matsiakh *et al.*, 2018; Plant *et al.*, 2019). Canelles *et al.* (2021) concluded that much of Europe was suitable for *C. perspectalis* but that the likely defoliation impacts will vary depending on climatic suitability and dispersal of adults. There is much ongoing research into potential control of *C. perspectalis* and into the nature and limits of its invasive biology, for which a genome should prove timely.

Genome sequence report

The genome was sequenced from a female *Cydalima perspectalis* (Figure 1) collected from Tonbridge, Kent, UK (51.19, 0.29). A total of 57-fold coverage in Pacific

Figure 1. Photograph of the *Cydalima perspectalis* (ilCydPers1) specimen used for genome sequencing.

Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 248 missing joins or mis-joins and removed 20 haplotypic duplications, reducing the assembly length by 0.42% and the scaffold number by 39.21%, and increasing the scaffold N50 by 6.04%.

The final assembly has a total length of 483.7 Mb in 199 sequence scaffolds with a scaffold N50 of 16.9 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, 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 (99.29%) of the assembly sequence was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the Z and W sex chromosomes. 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 58.6 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 94.9% (single = 94.7%, duplicated = 0.2%), using the lepidoptera_odb10 reference set (n = 5,286).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/1309588.

Table 1. Genome data for *Cydalima perspectalis*, ilCydPers1.1.

Project accession data		
Assembly identifier	ilCydPers1.1	
Species	*Cydalima perspectalis*	
Specimen	ilCydPers1	
NCBI taxonomy ID	1309588	
BioProject	PRJEB61127	
BioSample ID	SAMEA7521513	
Isolate information	ilCydPers1: thorax (DNA and Hi-C sequencing) ilCydPers2: thorax (RNA sequencing)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	58.6	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:94.9%[S:94.7%,D:0.2%], F:0.9%,M:4.2%,n:5,286	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.29%	*≥ 95%*
Sex chromosomes	ZW	*localised homologous pairs*
Organelles	Mitochondrial genome: 15.25 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11242108, ERR11226478, ERR11242109	
Hi-C Illumina	ERR11217097	
PolyA RNA-Seq Illumina	ERR11217102	
Genome assembly		
Assembly accession	GCA_951394215.1	
Accession of alternate haplotype	GCA_951394235.1	
Span (Mb)	483.7	
Number of contigs	1315	
Contig N50 length (Mb)	1.0	
Number of scaffolds	199	
Scaffold N50 length (Mb)	16.9	
Longest scaffold (Mb)	23.68	

* 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 version 5.3.2. 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/ilCydPers1_1/dataset/ilCydPers1_1/busco.

Methods

Sample acquisition and nucleic acid extraction

A female *Cydalima perspectalis* (specimen ID NHMUK010634993, ToLID ilCydPers1) was collected from Tonbridge, Kent, UK (latitude 51.19, longitude 0.29) on 2020-06-26 using a light trap. The specimen was collected and identified by Gavin Broad (Natural History Museum) and preserved on dry ice.

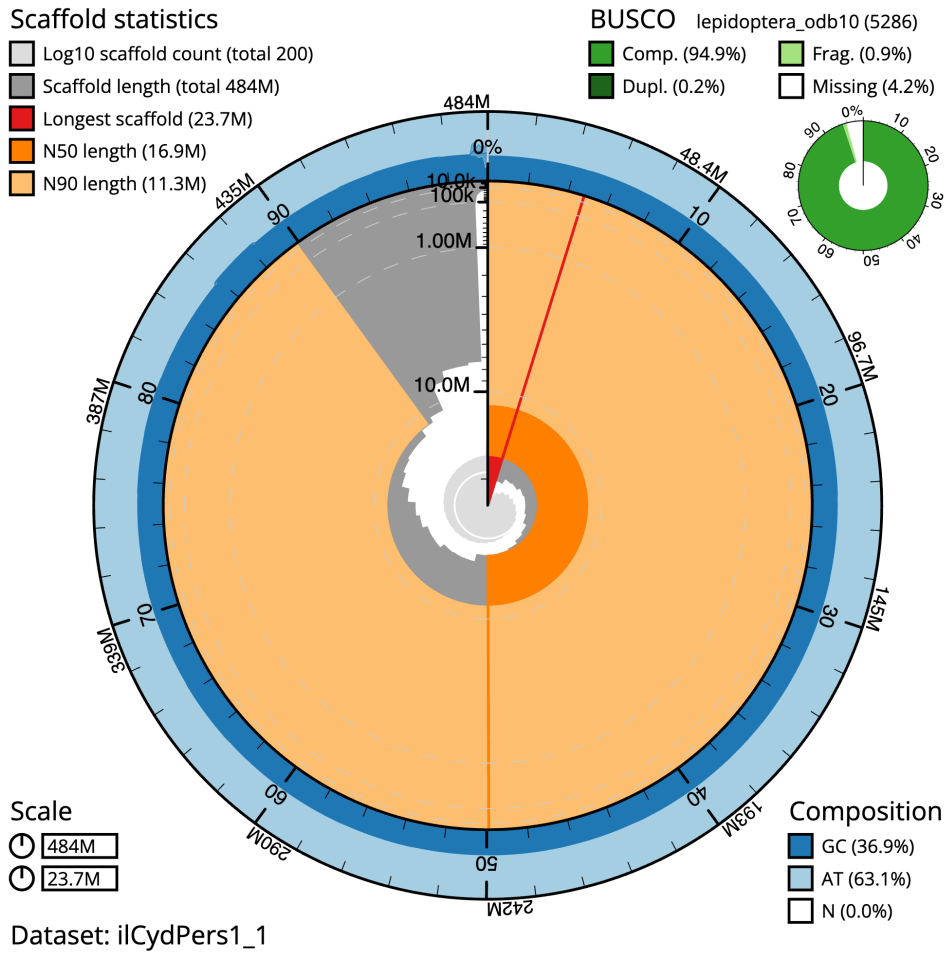

Figure 2. Genome assembly of *Cydalima perspectalis*, ilCydPers1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 483,694,489 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (23,684,527 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (16,856,657 and 11,297,898 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/ilCydPers1_1/dataset/ilCydPers1_1/snail.

The specimen used for RNA sequencing (specimen ID Ox000578, ToLID ilCydPers2) was collected from Wytham Woods, Oxfordshire (biological vice-county Berkshire) (latitude 51.77, longitude –1.34) on 2020-07-05. The specimen was collected and identified by Douglas Boyes (University of Oxford) and preserved on dry ice.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilCydPers1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the thorax was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Manual MagAttract v1 protocol (Strickland *et al.*, 2023b). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023a): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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 thorax tissue of ilCydPer2 in the Tree of Life Laboratory at the WSI using the RNA Extraction:

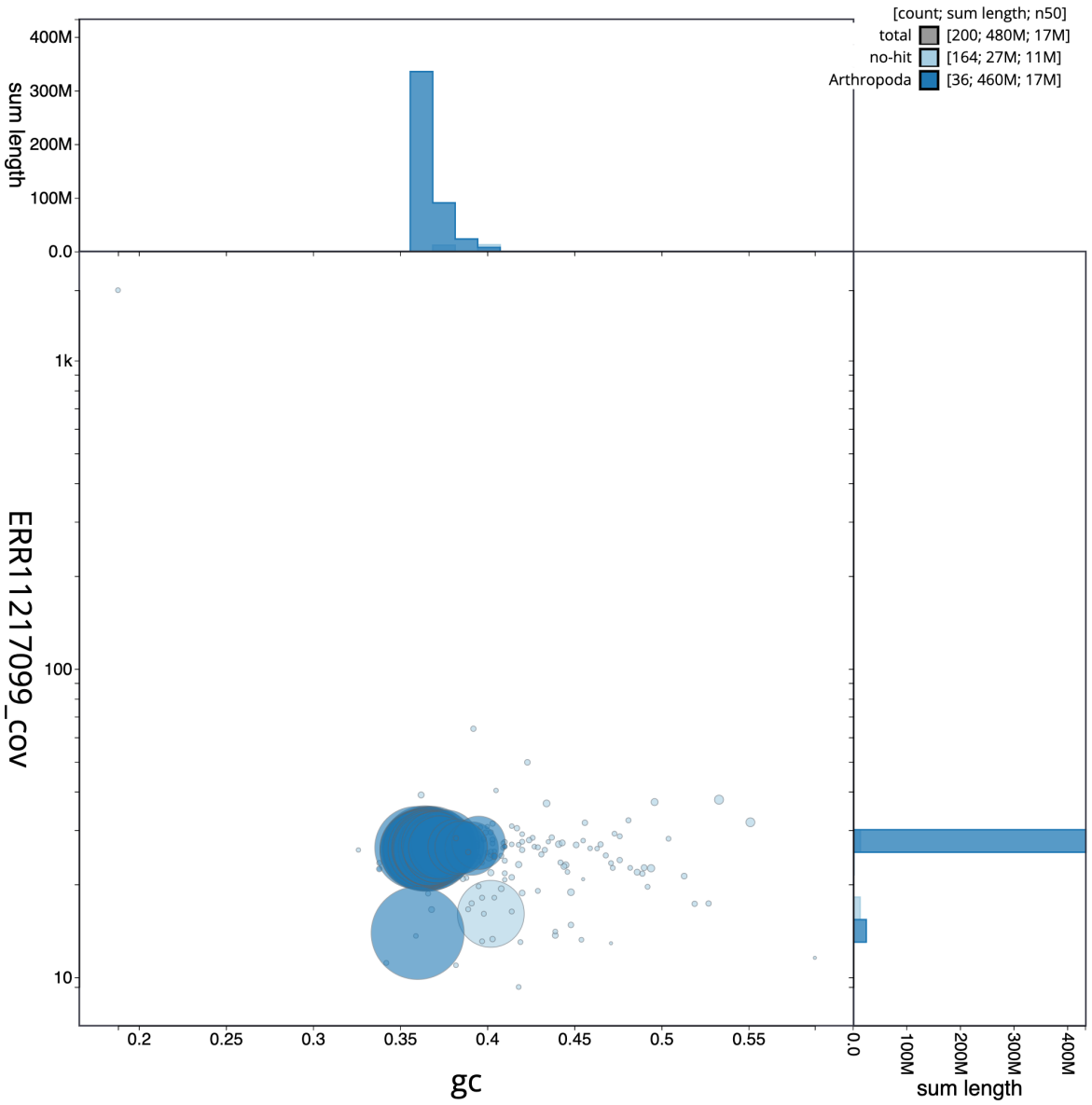

Figure 3. Genome assembly of *Cydalima perspectalis*, ilCydPers1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilCydPers1_1/dataset/ilCydPers1_1/blob.

Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

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 HiSeq 4000 (RNA-Seq) instruments. Hi-C data were also generated from remaining thorax tissue of ilCydPers1 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

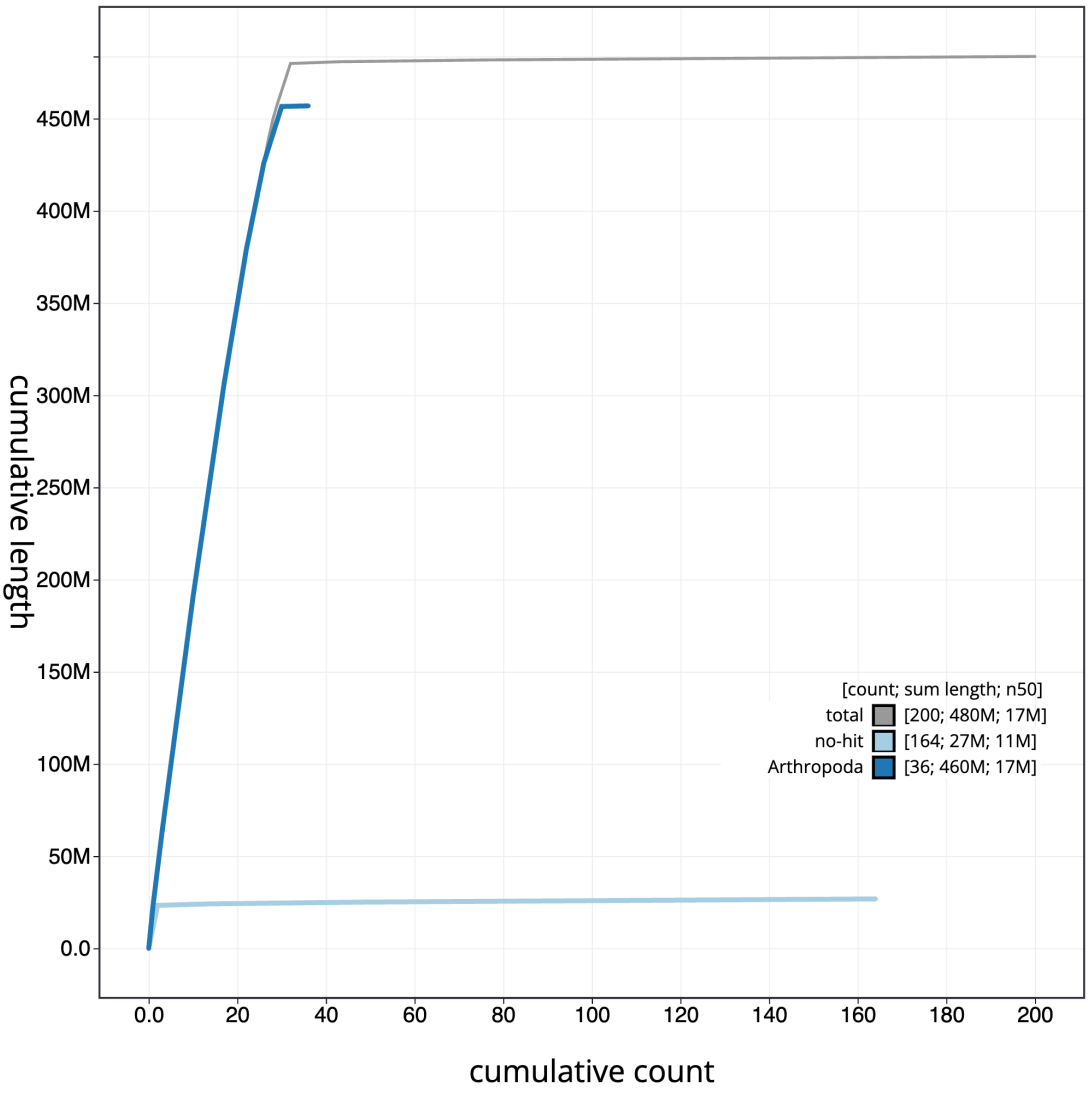

Figure 4. Genome assembly of *Cydalima perspectalis*, ilCydPers1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilCydPers1_1/dataset/ilCydPers1_1/cumulative.

scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected using the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was performed using JBrowse2 (Diesh *et al.*, 2023), HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (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.

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

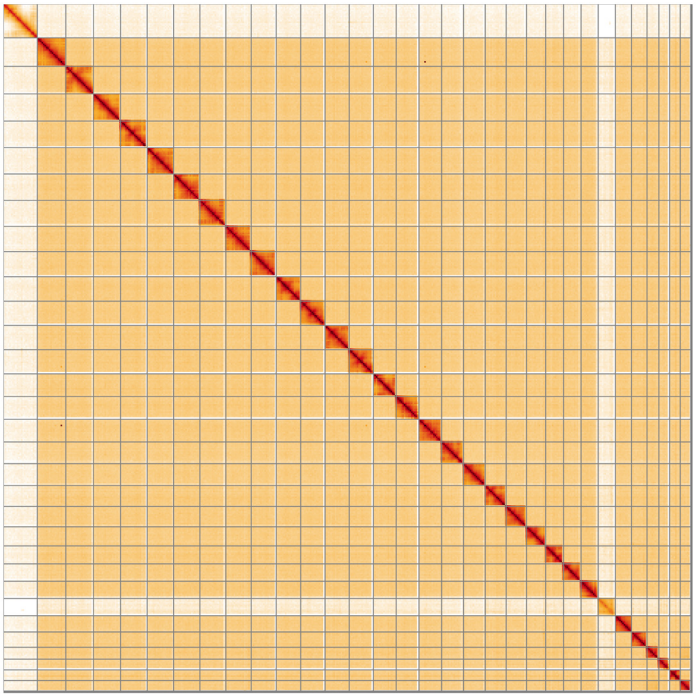

Figure 5. Genome assembly of *Cydalima perspectalis*, ilCydPers1.1: Hi-C contact map of the ilCydPers1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=bsaRcVplS0yOkvPaHOu0Ng.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Cydalima perspectalis*, ilCydPers1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX596200.1	1	19.94	36.5
OX596201.1	2	19.2	36.5
OX596202.1	3	19.01	37.0
OX596203.1	4	18.55	36.5
OX596204.1	5	18.4	36.5
OX596205.1	6	18.4	36.0
OX596206.1	7	18.17	36.5
OX596207.1	8	17.69	36.5
OX596208.1	9	17.47	36.5
OX596209.1	10	17.11	36.5
OX596210.1	11	16.97	36.0
OX596211.1	12	16.89	36.5
OX596212.1	13	16.86	36.0
OX596213.1	14	15.93	36.5
OX596214.1	15	15.9	37.0
OX596215.1	16	15.82	36.5
OX596216.1	17	15.23	37.0
OX596217.1	18	15.2	36.5
OX596218.1	19	14.62	37.0
OX596219.1	20	14.22	36.5
OX596220.1	21	13.34	37.0
OX596221.1	22	12.51	37.5
OX596222.1	23	12.21	37.0
OX596223.1	24	12.02	37.0
OX596225.1	25	11.3	37.5
OX596226.1	26	10.75	37.5
OX596227.1	27	8.3	38.0
OX596228.1	28	7.47	39.0
OX596229.1	29	7.46	39.5
OX596230.1	30	7.27	38.5
OX596224.1	W	12.0	40.0
OX596199.1	Z	23.68	36.0
OX596231.1	MT	0.02	19.0

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.2.1	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	3	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.5	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
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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: *Cydalima perspectalis* (box-tree moth). Accession number PRJEB61127; https://identifiers.org/ena.embl/PRJEB61127 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Cydalima perspectalis* 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. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.7139035.

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.7125292.

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 Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

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

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Bras A, Avtzis DN, Kenis M, *et al.*: **A complex invasion story underlies the fast spread of the invasive box tree moth (*Cydalima perspectalis*) across Europe.** *J Pest Sci.* 2019; **92**(3): 1187–1202.
Publisher Full Text

Canelles Q, Bassols E, Vayreda J, *et al.*: **Predicting the potential distribution and forest impact of the invasive species *Cydalima perspectalis* in Europe.** *Ecol Evol.* 2021; **11**(10): 5713–5727.
PubMed Abstract | Publisher Full Text | Free 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

Cook L, Muggleton J: **Encounters with the colour forms of the box-tree moth *Cydalima perspectalis* (Walker, 1859) (Lep.: Crambidae).** *Entomologist's Rec J Var.* 2023; **135**(2): 83.
Reference Source

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
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

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free Full Text

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
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

Mally R, Nuss M: **Phylogeny and nomenclature of the box tree moth, *Cydalima perspectalis* (Walker, 1859) comb. n., which was recently introduced into Europe (Lepidoptera: Pyraloidea: Crambidae: Spilomelinae).** *Eur J Entomol.* 2010; **107**(3): 393–400.
Publisher 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

Matsiakh I, Kramarets V, Mamadashvili G: **Box tree moth *Cydalima perspectalis* as a threat to the native populations of *Buxus colchica* in Republic of Georgia.** *J Entomol Res Soc.* 2018; **20**: 29–42.
Reference Source

Plant CW, Poole C, Salisbury A, *et al.*: **The Box-tree Moth *Cydalima perspectalis* (Walker, 1859) in Britain: an overview of its spread and current status.** *Entomologist's Rec J Var.* 2019; **131**: 122–147.
Reference Source

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 - Ancient Atlantis**. 2023.
Publisher Full Text

Poloni R, Dhennin M, Mappes J, *et al.*: **Exploring polymorphism in a palatable prey: predation risk and frequency dependence in relation to distinct levels of conspicuousness.** *Evol Lett.* 2024; qrad071.
Publisher 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

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023a.
Publisher Full Text

Strickland M, Moll R, Cornwell C, *et al.*: **Sanger Tree of Life HMW DNA extraction: manual MagAttract.** *protocols.io.* 2023b.
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

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 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

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

Wellcome Sanger Institute: **The genome sequence of the Box-tree Moth, *Cydalima perspectalis* (Walker, 1859).** European Nucleotide Archive. [dataset], accession number PRJEB61127. 2023.

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