



DATA NOTE

The genome sequence of the Bordered Straw, *Heliothis*

peltigera (Denis & Schiffermüller) 1775

[version 1; peer review: 2 approved]

Denise C. Wawman ¹, Liam M. Crowley ¹,

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

¹Department of Biology, University of Oxford, Oxford, England, UK

V1 First published: 15 Apr 2024, 9:195
<https://doi.org/10.12688/wellcomeopenres.21221.1>

Latest published: 15 Apr 2024, 9:195
<https://doi.org/10.12688/wellcomeopenres.21221.1>

Abstract

We present a genome assembly from an individual male *Heliothis peltigera* (the Bordered Straw; Arthropoda; Insecta; Lepidoptera; Noctuidae). The genome sequence is 332.8 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 15.3 kilobases in length. Gene annotation of this assembly on Ensembl identified 17,114 protein coding genes.

Keywords

Heliothis peltigera, Bordered Straw moth, genome sequence, chromosomal, Lepidoptera



This article is included in the [Tree of Life gateway](#).

Open Peer Review

Approval Status

	1	2
version 1		
15 Apr 2024	view	view

1. **Savarimuthu Ignacimuthu**, St. Xavier's College, Interdisciplinary Research Centre in Biology, Xavier Research Foundation, Manonmaniam Sundaranar University, Thirunelveli, India
Sivasankaran Kuppasamy, Loyola College, Chennai, India
Loyola College, Chennai, India
2. **Qiye Li**, BGI Research, Beijing, China
University of Chinese Academy of Sciences, Beijing, China

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: Wawman DC: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Crowley LM: 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.

Copyright: © 2024 Wawman DC *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Wawman DC, Crowley LM, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Bordered Straw, *Heliothis peltigera* (Denis & Schiffermüller) 1775 [version 1; peer review: 2 approved]** Wellcome Open Research 2024, 9:195 <https://doi.org/10.12688/wellcomeopenres.21221.1>

First published: 15 Apr 2024, 9:195 <https://doi.org/10.12688/wellcomeopenres.21221.1>

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; Noctuoidea; Noctuidae; Heliiothinae; *Heliothis*; *Heliothis peltigera* (Denis & Schiffermüller) 1775 (NCBI:txid542982).

Background

The Bordered Straw *Heliothis peltigera* is a Noctuid moth, but although it has the characteristic thick body and wing-shape seen in this family, it is brighter coloured than most of its dull brown cousins, with a ground colour ranging from straw-yellow, early in the season, to darker orange-brown, as increasing temperatures in the larval stages causes the adults to become darker (Waring *et al.*, 2017). It is usually more distinctly marked than similar species with a bold grey kidney mark and diagnostic dark diffuse blotches on the leading edge of the forewing near the kidney and adjacent to the darker band on the outer edge: the similar Scarce Bordered Straw or Old World Bollworm *Helicoverpa armigera* lacks both of these diagnostic markings, and Eastern Bordered Straw *Heliothis nubigera* has only a smaller more diffuse marking near the kidney, with a paler ground colour and three or more distinct small dots near the outer edge (Skinner & Wilson, 2009; Waring *et al.*, 2017).

Within the United Kingdom *Heliothis peltigera* is considered to be an immigrant species (Waring *et al.*, 2017), but it is common in tropical and subtropical areas of Africa, where it is a major agricultural pest and, with climate change, is spreading north into Europe (Ragionieri *et al.*, 2017). The larvae cause damage to crops such as safflower *Carthamus tinctorius*, tobacco *Nicotiana tabacum*, grapevine *Vitis* spp., cotton *Gossypium* spp., stone and citrus fruit trees, fodder crops, medicinal and ornamental plants (Dunkelblum & Kehat, 1992). Consequently, the synthesis and production of pheromones by *H. peltigera* (Altstein, 2004; Ragionieri *et al.*, 2017) and the moths' response to them (Dunkelblum & Kehat, 1992) have been the subject of intense research. More recently molecular methods have been used to identify a strain of *Helicoverpa armigera* single nucleopolyhedrovirus (HearNPV-TR), which infects *H. peltigera*, and is similar to other viruses which have been used for the biological control of other insect pests (Eroglu *et al.*, 2020).

We present a chromosomally complete genome sequence for *Heliothis peltigera*, based on one specimen collected using a mercury vapour light trap in a rural garden in the hamlet of Bratton, near Minehead, in Somerset, as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from a male *Heliothis peltigera* (Figure 1) collected from Bratton, Somerset, UK (51.20, -3.51). A total of 75-fold coverage in Pacific Biosciences



Figure 1. Photograph of the *Heliothis peltigera* (ilHelPelt1) specimen used for genome sequencing.

single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 4 missing joins or mis-joins, increasing the scaffold number by 3.23%.

The final assembly has a total length of 332.8 Mb in 31 sequence scaffolds with a scaffold N50 of 11.8 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 (99.99%) 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). The Z chromosome was identified based on synteny with *Craniophora ligustri* (GCA_905163465.1) (Boyes *et al.*, 2023). 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 68.1 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 99.1% (single = 98.9%, 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/542982>.

Genome annotation report

The *Heliothis peltigera* genome assembly (GCA_958496145.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 17,323 transcribed mRNAs from 17,114 protein-coding genes (Table 1; https://rapid.ensembl.org/Heliothis_peltigera_GCA_958496145.1/Info/Index).

Table 1. Genome data for *Heliothis peltigera*, ilHelPelt1.1.

Project accession data		
Assembly identifier	ilHelPelt1.1	
Species	<i>Heliothis peltigera</i>	
Specimen	ilHelPelt1	
NCBI taxonomy ID	542982	
BioProject	PRJEB62179	
BioSample ID	SAMEA112226458	
Isolate information	ilHelPelt1, male: thorax (DNA sequencing), head (Hi-C scaffolding)	
Assembly metrics*		Benchmark
Consensus quality (QV)	68.1	≥ 50
k-mer completeness	100.0%	≥ 95%
BUSCO**	C:99.1%[S:98.9%,D:0.2%], F:0.2%,M:0.7%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.99%	≥ 95%
Sex chromosomes	ZZ	localised homologous pairs
Organelles	Mitochondrial genome: 15.3 kb	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11458819	
Hi-C Illumina	ERR11468748	
Genome assembly		
Assembly accession	GCA_958496145.1	
Accession of alternate haplotype	GCA_958496135.1	
Span (Mb)	332.8	
Number of contigs	56	
Contig N50 length (Mb)	9.2	
Number of scaffolds	31	
Scaffold N50 length (Mb)	11.8	
Longest scaffold (Mb)	15.08	
Genome annotation		
Number of protein-coding genes	17,114	
Number of gene transcripts	17,323	

* 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/ilHelPelt1_1/dataset/ilHelPelt1_1/busco.

Methods

Sample acquisition and nucleic acid extraction

A male *Heliothis peltigera* (specimen ID Ox002232, ToLID ilHelPelt1) was collected from Bratton, Somerset, UK (latitude

51.20, longitude -3.51) on 2022-06-20, using a light trap. The specimen was collected by Denise Wawman (University of Oxford) and identified by Liam Crowley (University of Oxford), and then snap-frozen on dry ice.

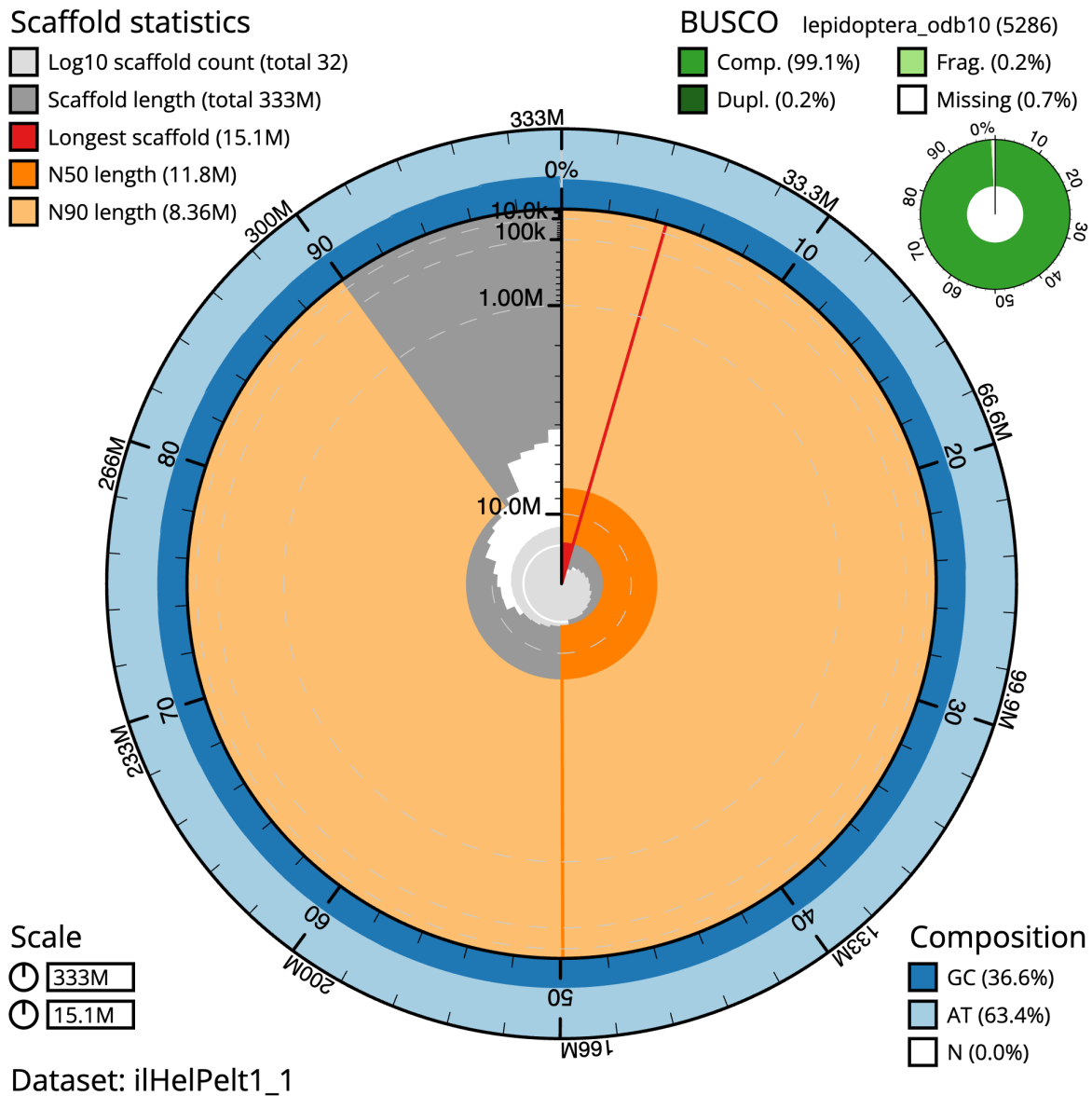


Figure 2. Genome assembly of *Heliothis peltigera*, ilHelPelt1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 332,843,978 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (15,076,996 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (11,813,164 and 8,357,070 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/ilHelPelt1_1/dataset/ilHelPelt1_1/snail.

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 ilHelPelt1 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 Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). DNA was sheared into an average fragment size of 12–20 kb in a

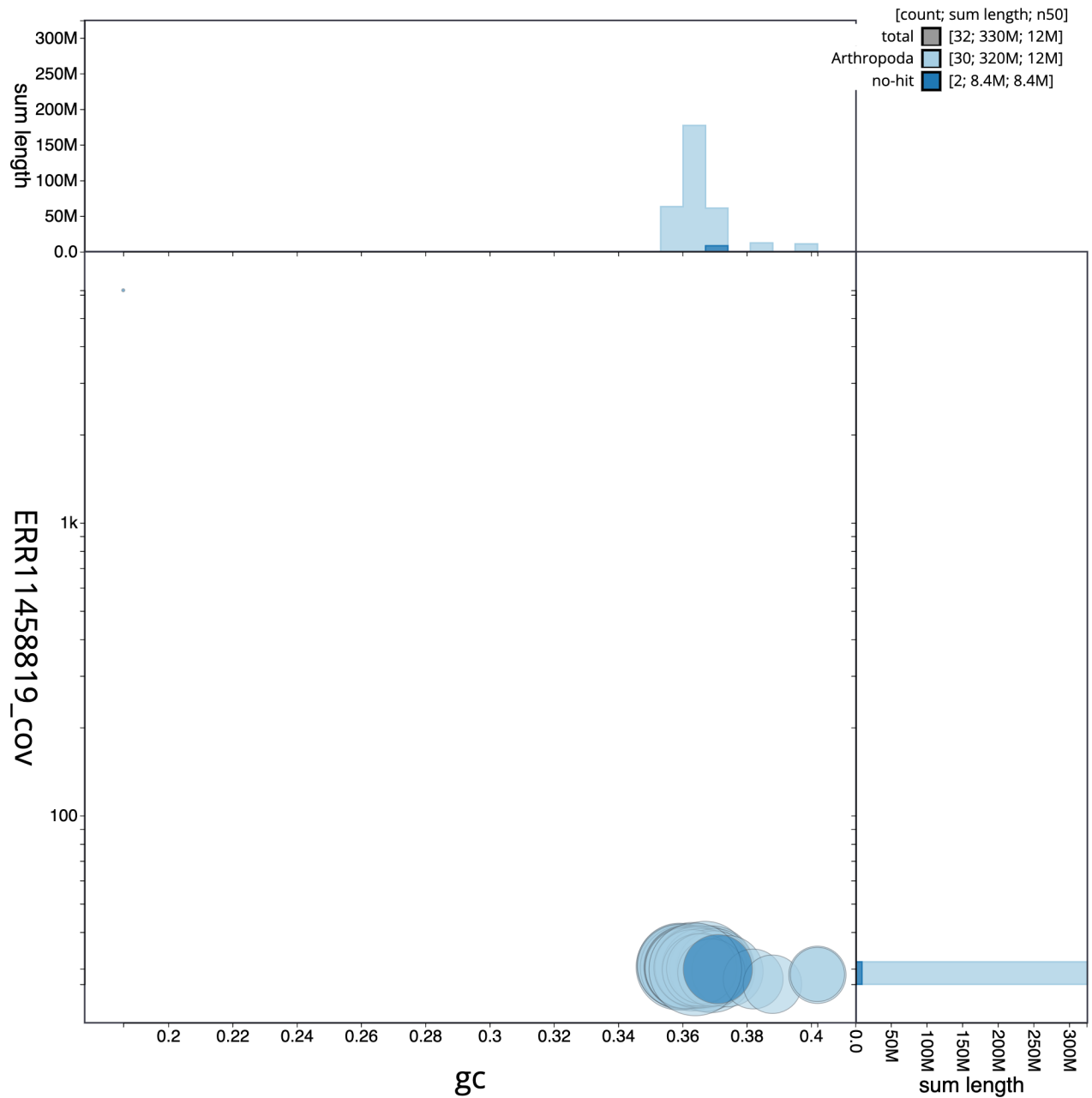


Figure 3. Genome assembly of *Heliothis peltigera*, ilHelPelt1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilHelPelt1_1/dataset/ilHelPelt1_1/blob.

Megaruptor 3 system with speed setting 31 (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Oatley *et al.*, 2023b): 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.

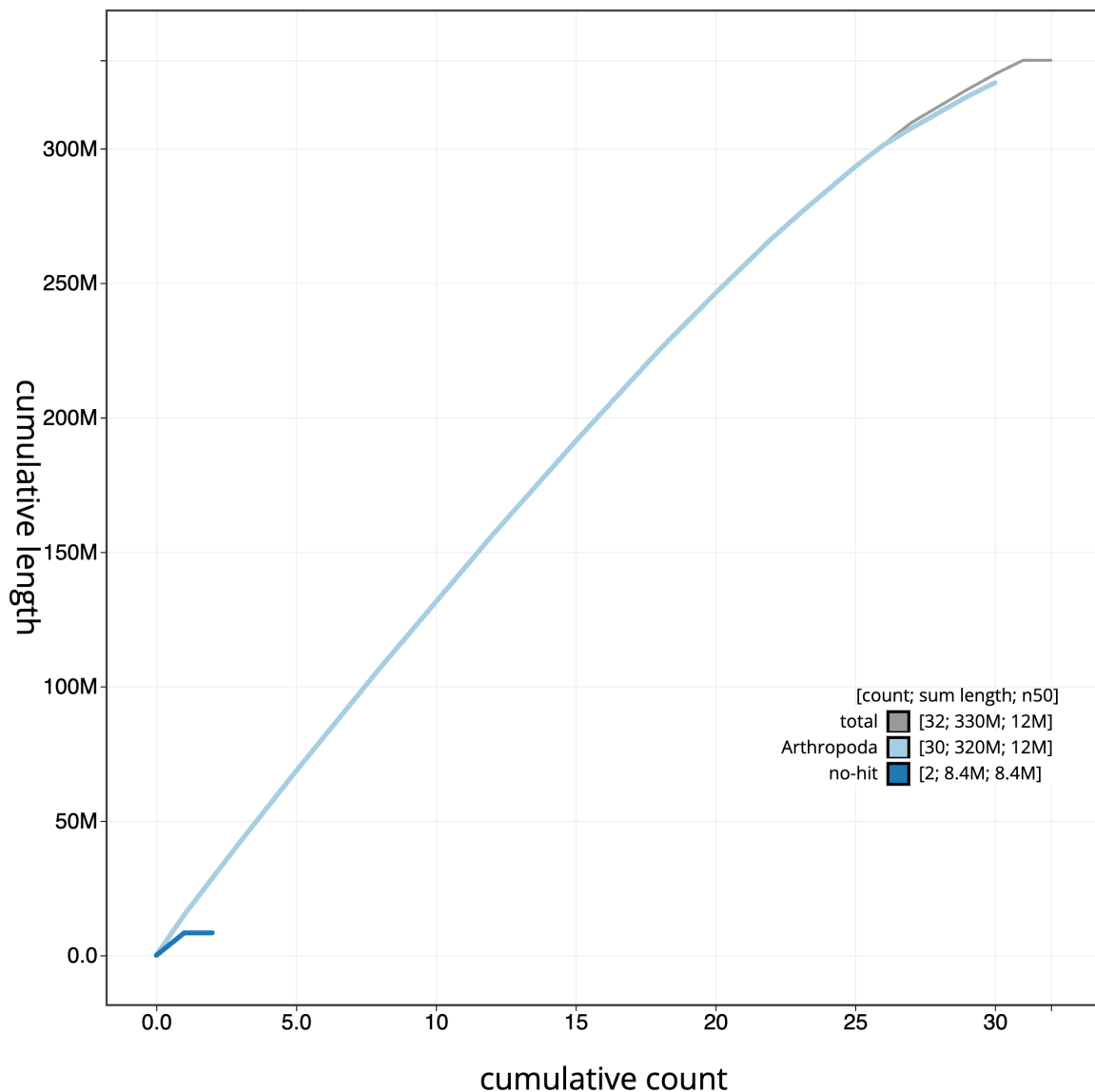


Figure 4. Genome assembly of *Heliothis peltigera*, ilHelPelt1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilHelPelt1_1/dataset/ilHelPelt1_1/cumulative.

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. DNA sequencing was performed by the Scientific Operations core at the WSI on a Pacific Biosciences SEQUEL II instrument. Hi-C data were also generated from head tissue of ilHelPelt1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using gHiGlass (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

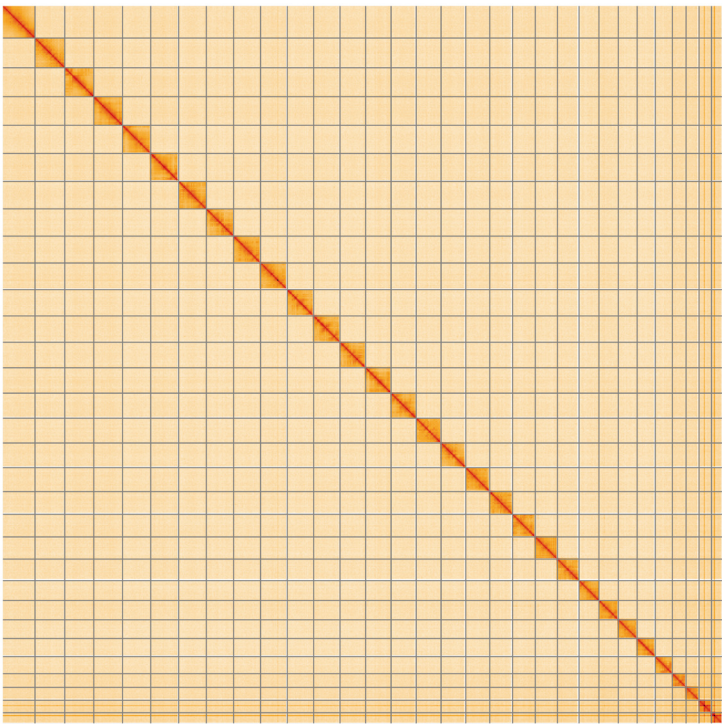


Figure 5. Genome assembly of *Heliothis peltigera*, ilHelPelt1.1: Hi-C contact map of the ilHelPelt1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=YWF6i8BXRhmu_mJj5vyPDg.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Heliothis peltigera*, ilHelPelt1.

INSDC accession	Chromosome	Length (Mb)	GC%
OY292322.1	1	13.75	36.5
OY292323.1	2	13.4	36.5
OY292324.1	3	13.25	37.0
OY292325.1	4	13.06	36.0
OY292326.1	5	12.89	36.5
OY292327.1	6	12.8	36.0
OY292328.1	7	12.6	36.0
OY292329.1	8	12.45	36.0
OY292330.1	9	12.33	36.5
OY292331.1	10	12.2	36.0
OY292332.1	11	12.2	36.0
OY292333.1	12	11.81	36.0
OY292334.1	13	11.77	36.5
OY292335.1	14	11.6	36.5
OY292336.1	15	11.49	36.0

INSDC accession	Chromosome	Length (Mb)	GC%
OY292337.1	16	11.42	36.5
OY292338.1	17	11.09	36.5
OY292339.1	18	10.54	37.0
OY292340.1	19	10.49	37.0
OY292341.1	20	10.26	36.5
OY292342.1	21	9.96	37.0
OY292343.1	22	9.22	36.5
OY292344.1	23	8.94	37.5
OY292345.1	24	8.69	36.5
OY292346.1	25	8.36	37.0
OY292347.1	26	7.95	37.0
OY292348.1	27	6.3	38.0
OY292349.1	28	6.01	39.0
OY292350.1	29	5.77	40.0
OY292351.1	30	5.16	40.0
OY292321.1	Z	15.08	36.5
OY292352.1	MT	0.02	19.0

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 BRAKER2 pipeline (Brůna *et al.*, 2021) was used in the default protein mode to generate annotation for the *Heliothis peltigera* assembly (GCA_958496145.1) in Ensembl Rapid Release at the EBI.

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.

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
YaHS	1.2a.2	https://github.com/c-zhou/yahs

Data availability

European Nucleotide Archive: *Heliothis peltigera* (bordered straw). Accession number PRJEB62179; <https://identifiers.org/ena.embl/PRJEB62179> (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Heliothis peltigera* 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.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](#)
- Altstein M: **Novel insect control agents based on neuropeptide antagonists: The PK/PBAN family as a case study.** *J Mol Neurosci*. 2004; **22**(1–2): 147–157. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor[®]3 for LI PacBio.** *protocols.io*. 2023. [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](#)
- Boyes D, Boyes C, University of Oxford and Wytham Woods Genome Acquisition Lab, et al.: **The genome sequence of the Coronet, *Craniophora ligustri* (Denis and Schiffermüller, 1775) [version 1; peer review: 1 approved, 1 approved with reservations].** *Wellcome Open Res*. 2023; **8**: 81. [Publisher Full Text](#)
- Brůna T, Hoff KJ, Lomsadze A, et al.: **BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.** *NAR Genom Bioinform*. 2021; **3**(1): lqaa108. [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](#)
- 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](#)
- Dunkelblum E, Kehat M: **Sexual behaviour of *Heliothis peltigera* (Lepidoptera: Noctuidae).** *Bulletin of Entomological Research*. 1992; **82**(1): 13–17. [Publisher Full Text](#)
- Eroglu GB, Inan C, Nalcacioglu R, et al.: **Genome sequence analysis of a *Helicoverpa armigera* single nucleopolyhedrovirus (HearNPV-TR) isolated from *Heliothis peltigera* in Turkey.** *PLoS One*. 2020; **15**(6): e0234635. [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*. 2021; **10**(1): gaa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Manni M, Berkeley MR, Seppely 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](#)
- Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2.** *protocols.io*. 2023a. [Publisher Full Text](#)
- Oatley G, Sampaio F, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Automated SPRI.** *protocols.io*. 2023b. [Publisher Full Text](#)
- Ragionieri L, Özbacı B, Neupert S, et al.: **Identification of mature peptides from *pban* and *capa* genes of the moths *Heliothis peltigera* and *Spodoptera littoralis*.** *Peptides*. 2017; **94**: 1–9. [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.: **Merquy: 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](#)

Skinner B, Wilson D: **Colour Identification Guide to the Moths of the British Isles**. Leiden, Netherlands: Brill, 2009.
[Reference Source](#)

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

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev)**. *Zenodo*. 2023b.
[Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashennikova 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 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, *et al.*: **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 Bordered Straw, *Heliothis peltigera* (Denis & Schiffermüller) 1775**. European Nucleotide Archive, [dataset], accession number PRJEB62179, 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](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 09 August 2024

<https://doi.org/10.21956/wellcomeopenres.23468.r90165>

© 2024 Li Q. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Qiye Li

¹ BGI Research, Beijing, China

² College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

In this study, Wawman *et al.* present the first chromosome-scale genome assembly of the Bordered Straw (*Heliothis peltigera*) based on PacBio HiFi long-read and Hi-C long-range sequencing data. The contiguity, completeness and accuracy of the genome assembly are excellent. Together with the protein-coding genes annotated for the genome, this study provides a valuable dataset for the community. However, I do have some suggestions that may help improve the manuscript.

1. In the last paragraph of Background, the authors claim that they present a “chromosomally complete genome sequence” for *Heliothis peltigera*. Given that there are still some gaps inside the chromosomes, and that the presence of telomeres on chromosomal ends has not been ascertained, I would suggest avoiding emphasizing the genome sequence is complete. Instead, “near-complete” is a more cautious statement.

2. The authors state that “The Z chromosome was identified based on synteny with *Craniophora ligustri*(GCA_905163465.1) (Boyce et al., 2023).” Given that sex chromosome transitions are quite common in some animal lineages, could the authors explain a little further in the manuscript why this is workable? For example, is that because the sex chromosomes are highly conserved across Noctuidae species? This information is not clear to the readers.

3. All the software tools used for building and assessing the genome assembly have been provided in methods and table 3. But to allow replication by others, the parameters of each tool are required as well. Even if default parameters are used, it is worthy to mention somewhere in the manuscript. I suggest that the authors may provide the comments of parameters used for each tool in table 3, especially when some tools are run with default settings, but some are not.

4. The protein-coding genes are annotated for the genome assembly, which are valuable for future studies. However, the completeness of the gene set is not assessed. At least, presenting the assessment for the gene set using BUSCO will be informative.

Some minor comments:

1. The first sentence of Background, "The Bordered Straw *Heliothis peltigera* is a Noctuid moth, but although..... to become darker (Waring et al., 2017)", is quit long and a bit hard to follow. Perhaps, the authors can break it into two or more sentences?
2. The Latin name of *Heliothis peltigera* has been spelled out for multiple times in the main text. Please consider using "*H. peltigera*" after the first time of spelling out.

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: Biodiversity genomics

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 07 August 2024

<https://doi.org/10.21956/wellcomeopenres.23468.r90157>

© 2024 Ignacimuthu S et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Savarimuthu Ignacimuthu**

St. Xavier's College, Interdisciplinary Research Centre in Biology, Xavier Research Foundation, Manonmaniam Sundaranar University, Thirunelveli, TN, India

Sivasankaran Kuppusamy

¹ Entomology Research Institute, Loyola College, Chennai, Tamil Nadu, India

² Entomology Research Institute, Loyola College, Chennai, Tamil Nadu, India

Review Report

The authors have sequenced the genome of *Heliothis peltigera* (Denis & Schiffermüller) 1775 with

Next Generation Sequence techniques. A total of 332.8 megabases in span were assembled through the genome assemble. Totally 17,114 protein-coding genes and 17,323 number of gene transcripts were observed through the annotation. Authors have utilized appropriate techniques for DNA isolation, library construction for sequencing and used proper software for assembly and annotation.

Comment 1:

The author's name of the species with year can be included in the first sentence of the background. "The Bordered Straw *Heliothis peltigera* (Denis & Schiffermüller) 1775 is a Noctuid moth..."

There is confusion in this sentence "but although it has the characteristic thick body and wing-shape seen in this family". Please delete the sentence or can be modified.

Comments 2:

Authors have used the word kidney instead of reniform in the background of the species. Actually, it's a kidney shaped or renal shaped (reniform) marking not kidney. So, the second sentence of the background can be changed as

"It is usually more distinctly marked than similar species with a bold grey reniform mark and diagnostic dark diffuse blotches on the leading edge of the forewing near the reniform and adjacent to the darker band on the outer edge: the similar Scarce Bordered Straw or Old World Bollworm *Helicoverpa armigera* lacks both of these diagnostic markings, and Eastern Bordered Straw *Heliothis nubigera* has only a smaller more diffuse marking near the reniform, with a paler ground colour and three or more distinct small dots near the outer edge (Skinner & Wilson, 2009; Waring et al., 2017)".

Comment 3:

Authors have given the genus name *Heliothis peltigera* full form throughout the article. For the first time the genus name can be given in full form then subsequently can be given in short form like *H. peltigera*

Comment 4:

The second paragraph first sentence fourth line "remove the comma (,) after and...."

Comment 5:

Total length of the mitochondrial genome sequence must be provided in the text.

Above all, I confirm that the manuscript meets the necessary scientific standard and is suitable for indexing"

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: Molecular biology

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