Check for updates

DATA NOTE

The genome sequence of the small elephant hawk moth, *Deilephila porcellus* (Linnaeus, 1758) [version 1; peer review: 2 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, Laura Sivess[2],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, UK
[2]Department of Life Sciences, Natural History Museum, London, UK

[+] Deceased author

v1 **First published:** 08 Mar 2022, **7**:80
https://doi.org/10.12688/wellcomeopenres.17740.1
Latest published: 08 Mar 2022, **7**:80
https://doi.org/10.12688/wellcomeopenres.17740.1

Abstract

We present a genome assembly from an individual male *Deilephila porcellus* (the small elephant hawk moth; Arthropoda; Insecta; Lepidoptera; Sphingidae). The genome sequence is 402 megabases in span. The majority of the assembly (99.99%) is scaffolded into 29 chromosomal pseudomolecules, with the Z sex chromosome assembled.

Keywords

Deilephila porcellus, small elephant hawk moth, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 08 Mar 2022	✓ view	✓ view

1. **Martin Pippel**, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
2. **Surya Saha**, Boyce Thompson Institute for Plant Research, Ithaca, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Boyes D: Investigation, Resources; **Sivess L**: 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) 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 small elephant hawk moth, *Deilephila porcellus* (Linnaeus, 1758) [version 1; peer review: 2 approved]** Wellcome Open Research 2022, **7**:80 https://doi.org/10.12688/wellcomeopenres.17740.1

First published: 08 Mar 2022, **7**:80 https://doi.org/10.12688/wellcomeopenres.17740.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Sphingidae; Macroglossinae; Macroglossini; Deilephila; *Deilephila porcellus* (Linnaues, 1758) (NCBI:txid644661).

Background

Deilephila porcellus (small elephant hawk-moth) is characterised by striking pink and sand markings and is distributed across Europe, reaching as far East as China. Often confused with *Deilephila elpenor* (elephant hawk-moth), *Deilephila porcellus* can be identified most easily by a slightly smaller wingspan (40–45mm), brighter colouration and lack of the longitudinal pink abdominal stripe, typical of *D. elpenor*.

Deilephila porcellus is widespread throughout Britain, of rather local distribution in Southern England and Wales and scarce in Scotland and Northern England. This species flies from May to July and can be found in a range of open habitats including grassland, heathland, sand dunes and shingle beaches (Waring *et al.*, 2017). Adults are generalists, nocturnally feeding on the nectar of numerous flowering plants, including Rhododendron and Honeysuckle. Orchids are frequently visited for nectar; the relative frequency of different hawk-moth pollinators, with their differing proboscis lengths, has been shown to select for different spur lengths in the lesser butterfly orchid (*Platanthera bifolia*). in open areas in Sweden, the relatively short-tongued *Deilephila porcellus* is the most frequent pollinator and the orchid's spurs are correspondingly short when compared to woodland populations, mainly pollinated by the long-tongued *Sphinx ligustri (Boberg et al., 2014)*. Caterpillars, which primarily feed on bedstraws (*Galium*), emerge from June to September, and vary in colouration from brown to grey-green with large eyespots situated towards the anterior end. Functionally, eyespots and behaviour act to deter avian predation; when threatened, larvae widen anterior segments of the body, adopting defensive postures thought to mimic snakes, thus reducing incidence of attacks (Hossie & Sherratt, 2013; Poulton, 1890). The full lifecycle takes one year to complete, with pupae over-wintering in cocoons beneath larval food plants or just below the surface of the leaf litter.

Here we present a genome sequence for *D. porcellus*, generated as part of the Darwin Tree of Life Project.

Genome sequence report

The genome was sequenced from a single male *D. porcellus* (Figure 1) collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.337). A total of 40-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 92-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 4 missing/misjoins and removed 1 haplotypic duplications, reducing the scaffold number by 9.09%.

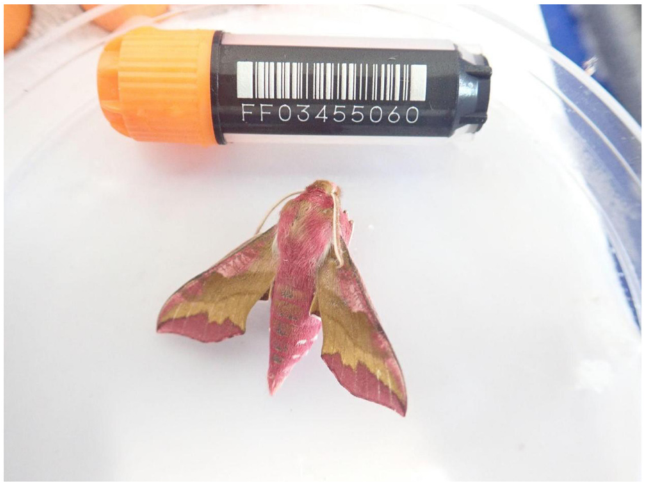

Figure 1. Image of the *Deilephila porcellus* specimen taken prior to preservation and processing.

The final assembly has a total length of 402 Mb in 30 sequence scaffolds with a scaffold N50 of 15.1 Mb (Table 1). Of the assembly sequence, 99.99% was assigned to 29 chromosomal-level scaffolds, representing 28 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 5; Table 2). The assembly has a BUSCO (Simão *et al.*, 2015) completeness of 98.8% (single, 98.5%, duplicated 0.2%) using the lepidoptera_odb10 reference set (n=5286). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods

Sample acquisition and nucleic acid extraction

A male *D. porcellus* (ilDeiPorc1) was collected from Wytham Woods, Oxfordshire, UK (latitude 51.772, longitude -1.337) by Douglas Boyes, University of Oxford, using a light trap. The specimens were identified by the same individual and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The ilDeiPorc1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Abdomen tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMpure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible

Table 1. Genome data for *Deilephila porcellus*, ilDeiPorc1.2.

Project accession data	
Assembly identifier	ilDeiPorc1.2
Species	*Deilephila porcellus*
Specimen	ilDeiPorc1
NCBI taxonomy ID	644661
BioProject	PRJEB42950
BioSample ID	SAMEA7520522
Isolate information	Male, head/thorax (Hi-C), abdomen (genome assembly)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6406201, ERR6412027
10X Genomics Illumina	ERR6054401-ERR6054404
Hi-C Illumina	ERR6054400
Genome assembly	
Assembly accession	GCA_905220455.2
Accession of alternate haplotype	GCA_905220465.1
Span (Mb)	402
Number of contigs	35
Contig N50 length (Mb)	14.9
Number of scaffolds	30
Scaffold N50 length (Mb)	15.1
Longest scaffold (Mb)	20.4
BUSCO* genome score	C:98.8%[S:98.5%,D:0.2%],F:0.3%, M:0.9%,n:5286

*BUSCO scores based on the lepidoptera_odb10 BUSCO set using v5.1.2. 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/ilDeiPorc1.2/dataset/CAJMZX02/busco.

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 Chromium read cloud sequencing libraries were constructed according to the manufacturers' instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina HiSeq X (10X) instruments. Hi-C data were generated from head/thorax tissue of ilDeiPorc1 using the Arima v2 kit and sequenced on HiSeq X.

Genome assembly

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021). Haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing

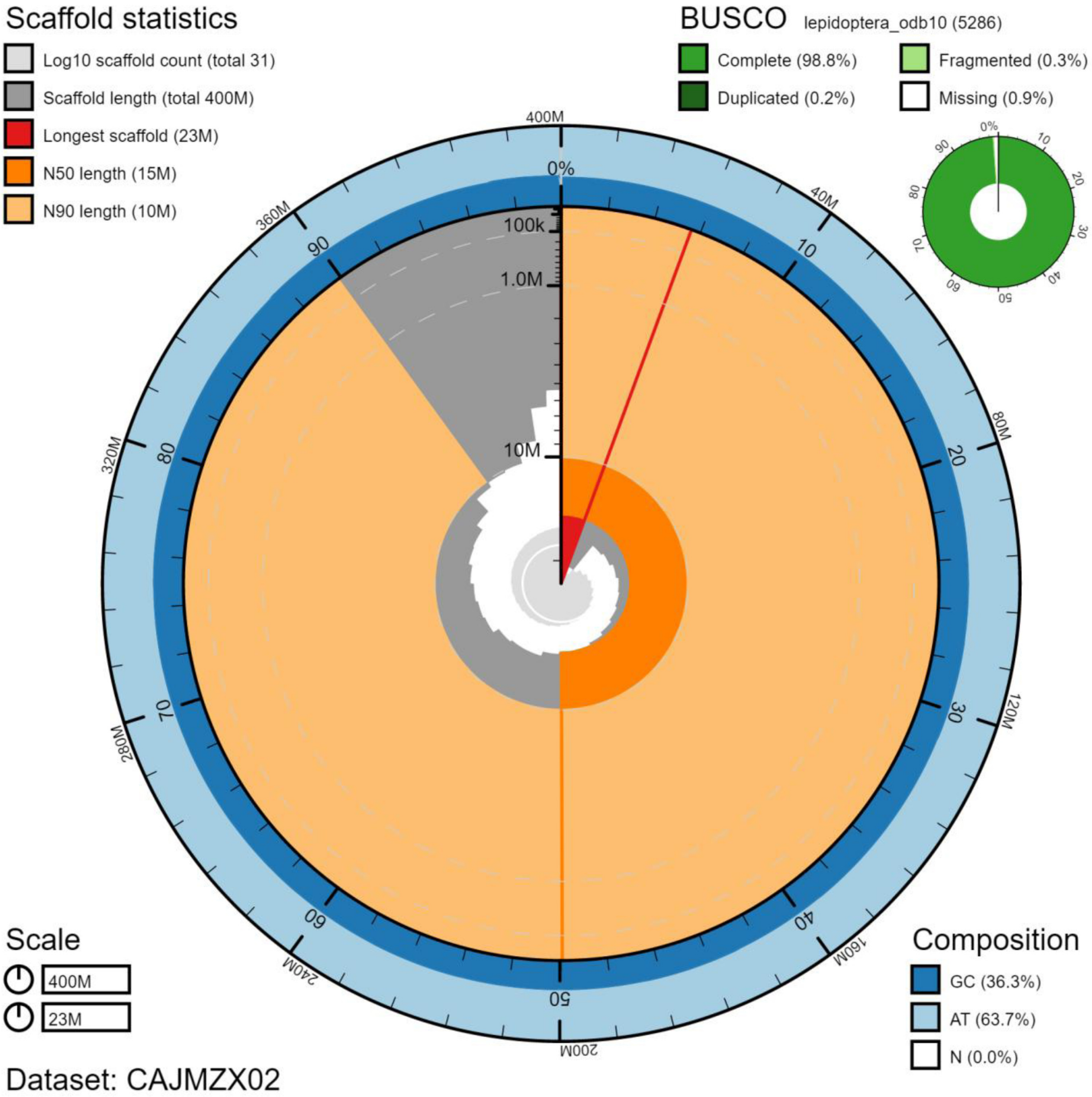

Figure 2. Genome assembly of *Deilephila porcellus*, ilDeiPorc1.2: metrics. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 402,071,895 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (22,662,151 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (15,067,504 and 10,104,753 bp), respectively. The pale grey spiral shows the cumulative chromosome 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/ilDeiPorc1.2/dataset/CAJMZX02/snail.

was performed by aligning 10X Genomics read data to the assembly with longranger 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)

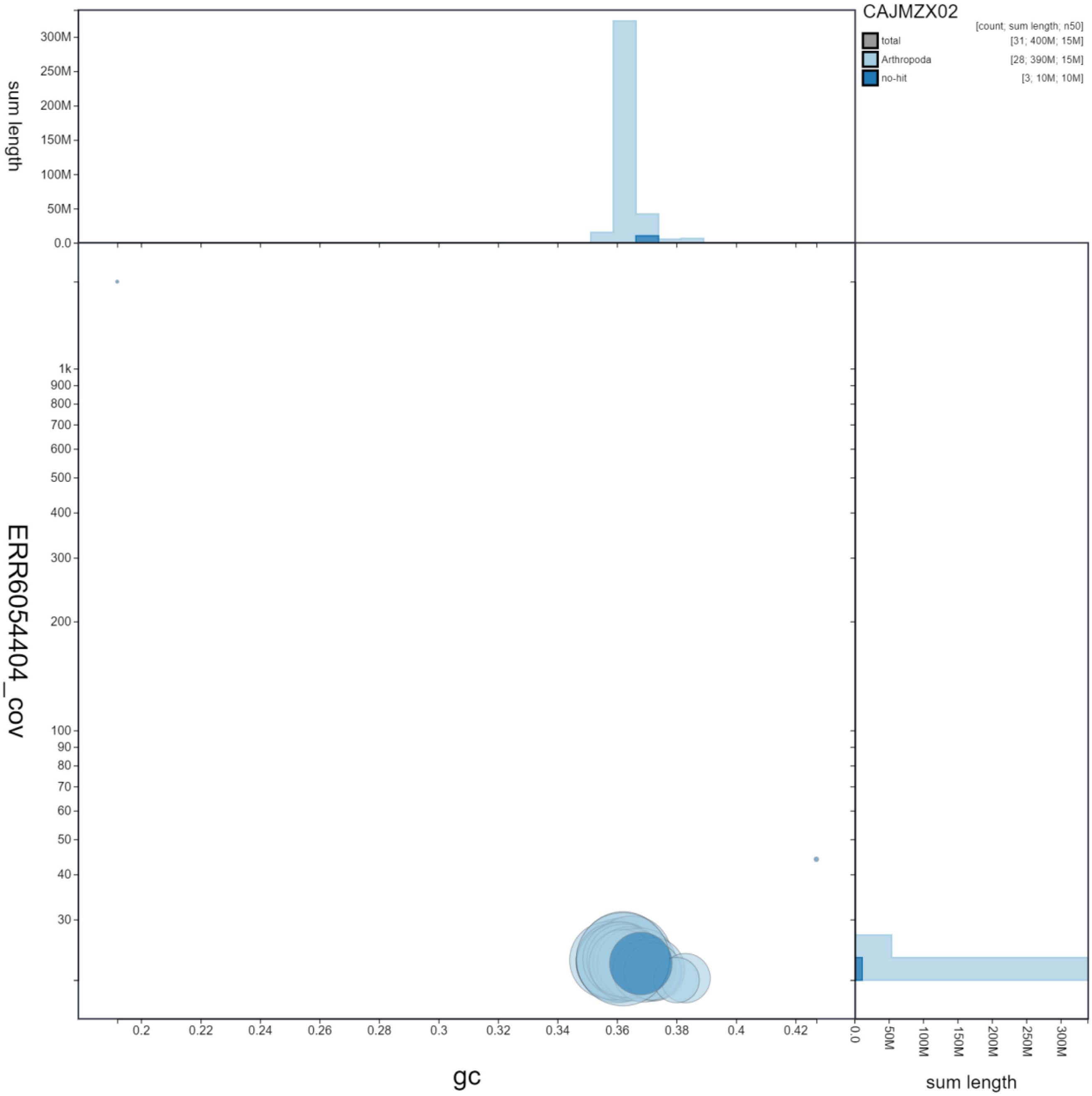

Figure 3. Genome assembly of *Deilephila porcellus*, ilDeiPorc1.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilDeiPorc1.2/dataset/CAJMZX02/blob.

as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2021), which performed annotation using MitoFinder (Allio *et al.*, 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis *et al.*, 2020). Table 3 contains a list of all software tool versions used, where appropriate.

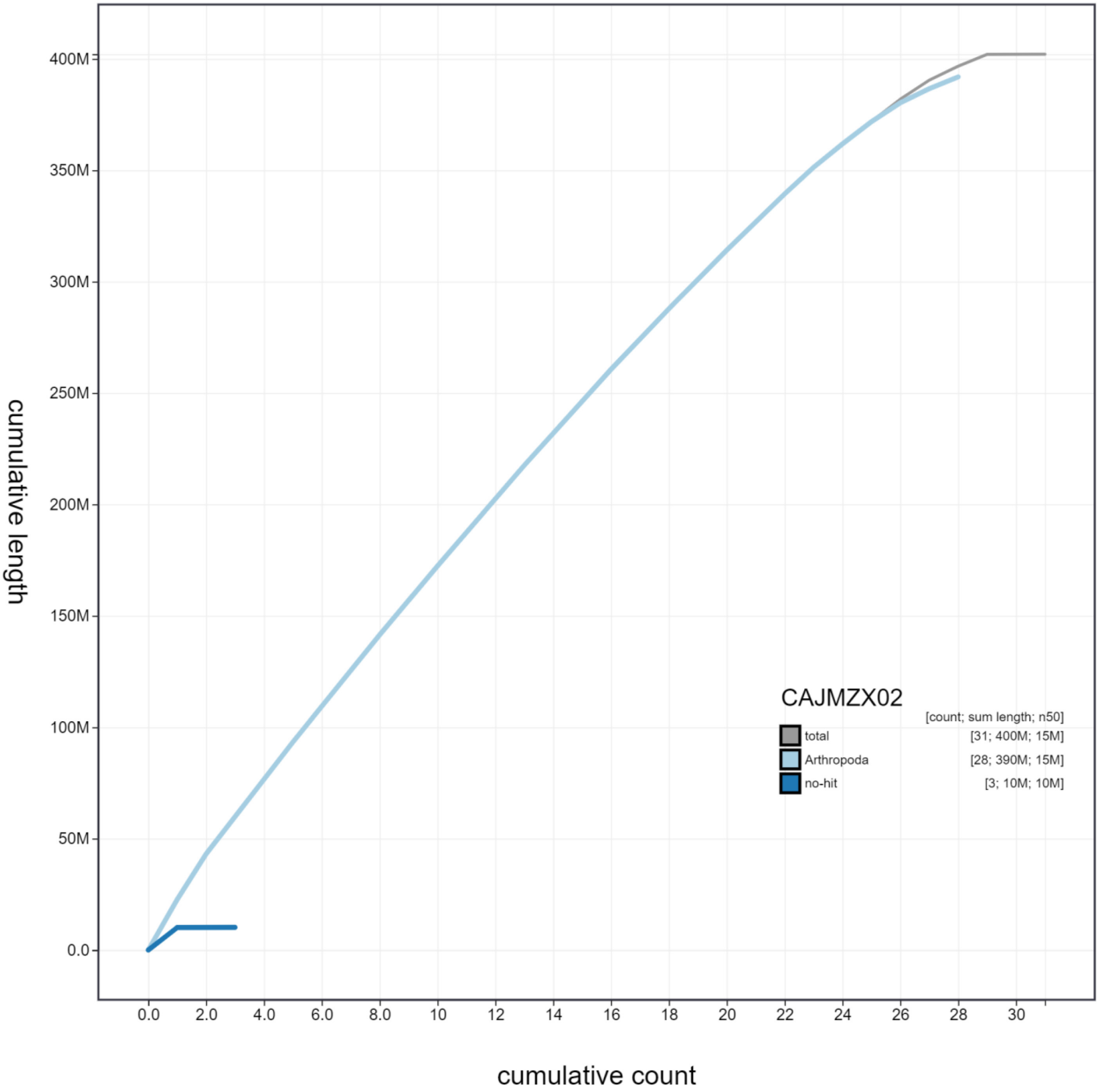

Figure 4. Genome assembly of *Deilephila porcellus*, ilDeiPorc1.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilDeiPorc1.2/dataset/CAJMZX02/cumulative.

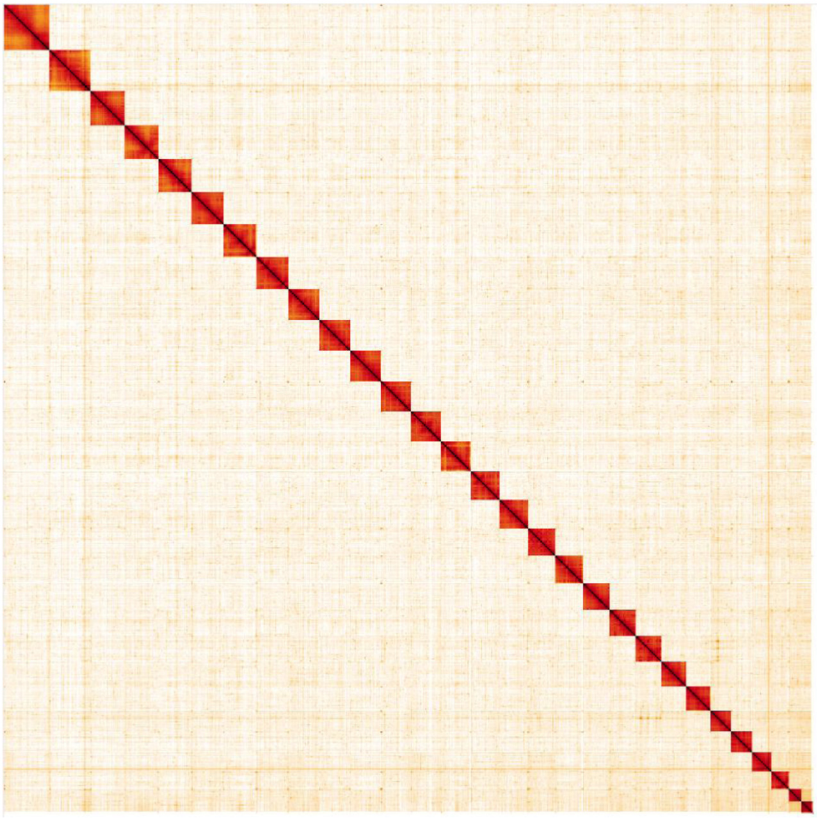

Figure 5. Genome assembly of *Deilephila porcellus*, ilDeiPorc1.2: Hi-C contact map. Hi-C contact map of the ilDeiPorc1.2 assembly, visualised in HiGlass. Chromosomes are given in size order from left to right and top to bottom.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Deilephila porcellus* ilDeiPorc1.2.

INSDC accession	Chromosome	Size (Mb)	GC%
LR999971.1	1	20.41	36.2
LR999972.1	2	17.03	36.1
LR999973.1	3	16.78	36.4
LR999974.1	4	16.43	36.4
LR999975.1	5	16.18	36.2
LR999976.1	6	16.07	36.5
LR999977.1	7	15.87	36.0
LR999978.1	8	15.64	36.1
LR999979.1	9	15.35	35.7
LR999980.1	10	15.14	36.2
LR999981.1	11	15.07	35.9
LR999982.1	12	14.89	35.9
LR999983.1	13	14.45	36.1
LR999984.1	14	14.44	35.9
LR999985.1	15	14.25	36.1
LR999986.1	16	13.66	36.3
LR999987.1	17	13.52	36.2
LR999988.1	18	13.31	36.4
LR999989.1	19	12.99	36.8
LR999990.1	20	12.67	36.6
LR999991.1	21	12.58	36.5
LR999992.1	22	11.95	36.2
LR999993.1	23	10.52	37.2
LR999994.1	24	10.10	36.8
LR999995.1	25	9.95	36.9
LR999996.1	26	8.52	37.2
LR999997.1	27	6.29	38.3
LR999998.1	28	5.31	38.0
LR999970.1	Z	22.66	36.2
LR999999.2	MT	0.02	19.3
-	Unplaced	0.04	42.7

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.12	Cheng *et al.*, 2021
purge_dups	1.2.3	Guan *et al.*, 2020
SALSA2	2.2	Ghurye *et al.*, 2019
longranger align	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
freebayes	1.3.1-17-gaa2ace8	Garrison & Marth, 2012
MitoHiFi	1.0	Uliano-Silva *et al.*, 2021
gEVAL	N/A	Chow *et al.*, 2016
PretextView	0.1.x	https://github.com/wtsi-hpag/PretextView
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
BlobToolKit	2.6.4	Challis *et al.*, 2020

Data availability

European Nucleotide Archive: Deilephila porcellus (small elephant hawk-moth). Accession number PRJEB42950; https://identifiers.org/ena.embl/PRJEB42950.

The genome sequence is released openly for reuse. The *D. porcellus* 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 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 University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.5746938.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.5744972.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.6125027.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.5746904.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.6125046.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.5638618.

References

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**

Boberg E, Alexandersson R, Jonsson M, *et al.*: **Pollinator Shifts and the Evolution of Spur Length in the Moth-Pollinated Orchid *Platanthera Bifolia*.** *Ann Bot.* 2014; **113**(2): 267–75.
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): 1361–74.
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): 170–75.
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**

Garrison E, Marth G: **Haplotype-Based Variant Detection from Short-Read Sequencing.** arXiv: 1207.3907. 2012.
Reference Source

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): 2896–98.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Hossie TJ, Sherratt TN: **Defensive Posture and Eyespots Deter Avian Predators**

from Attacking Caterpillar Models. *Anim Behav.* 2013; **86**(2): 383–89.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | Publisher Full Text | Free Full Text

Poulton EB: **The Colours of Animals: Their Meaning and Use, Especially Considered in the Case of Insects.** D. Appleton. 1890.
Publisher Full Text

Rao SS, Huntley MH, Durand NC, *et al.*: **A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.** *Cell.* 2014; **159**(7): 1665–80.
PubMed Abstract | Publisher Full Text | Free Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–12.
PubMed Abstract | Publisher Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, *et al.*: **marcelauliano/MitoHiFi: mitohifi_v2.0**. 2021.
Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ✓ ✓

Version 1

Reviewer Report 11 April 2022

https://doi.org/10.21956/wellcomeopenres.19631.r49246

✓ **Surya Saha**
Boyce Thompson Institute for Plant Research, Ithaca, NY, USA

The rapid release of data for use by the community is commendable and the Wellcome Open Research Data Notes are an excellent practice adopted by the DToL project. This manuscript is well-written and presents a high quality and contiguous arthropod genome assembly with DNA sourced from a single individual which is the gold standard for genome assembly. The NCBI Bioproject is well structured and provides access to both the primary and alternate haplotype assemblies.

I only have minor comments:

1. Although the versions and tools are mentioned, the methods lack enough details for reasonable reproducibility. The review from Martin already describes these lacunae in sufficient detail but I will reiterate them here for the sake of completeness. Parameters used for execution (even if they were the defaults) need to be reported for readers. Sharing a script with commands to reproduce the major steps in the assembly using a github repository release with a zenodo DOI might be a solution worth considering.

2. Figures 3 and 4 don't provide a lot of insights for genomes of this quality.

3. Using a kmer mining toolkit like KAT (https://github.com/TGAC/KAT) can help uncover the heterozygosity and repetitiveness in the assembly considering this is not a completely phased assembly. This information might have already been generated with Meryl (https://github.com/VGP/vgp-assembly/tree/master/pipeline/meryl).

4. Was RepeatMasker run on the assembly as a part of the VGP pipeline (https://github.com/VGP/vgp-assembly/tree/master/pipeline). A summary table of repeat elements can be helpful for understanding the architecture of this genome.

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: genome assembly and annotation, i5k, Ag100Pest, AgriVectors, Hemipteran 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 22 March 2022

https://doi.org/10.21956/wellcomeopenres.19631.r49245

✓ **Martin Pippel**
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

“The genome sequence of the small elephant hawk moth, *Deilephila porcellus*” from Boyes *et al.* describes a high-quality genome assembly based on PacBio HiFi reads, 10X Genomics read clouds and HiC data.

The data note is well structured and based on the described methods and the publicly available sequencing data also reproducible by the scientific community. Already the contig assembly shows a very high contiguity, indeed 25 out 29 chromosomal-level scaffolds are assembled into single contigs.

Methods:
Genome assembly methods are a bit short and it could be hard to reproduce the assembly without further documentation of the used program arguments. Even if all tools were run in default mode, it would be worth mentioning. Some of the used tools, such as purge_dups depend itself on other programs like minimap2. But those could not be found in Table 3. I could not find any information if all variants from Freebayes were used for the error polishing or if a variant filtering step was applied beforehand.

I do only have some minor comments and suggestions:

- I could not find the HiC read coverage in the Data Note.

- The plots from Figures 3 and 4 are for such a high-contiguous assembly not very informative.
- Additional statistics about the sequencing data (e.g. read length N50 of HiFi reads and 10X read clouds, kmer based genome size and heterozygosity estimates) and the final assembly (e.g. merqury QV scores, repeat content) would be nice to see as well.
- Error-polishing strategies of HiFi-based assemblies are currently under debate. I assume you applied bcftools consensus on the filtered Freebayes VCF files similar to the VGP assembly pipeline (https://github.com/VGP/vgp-assembly/tree/master/pipeline). How big was the improvement based in the QV-score? Were the alternate contigs included in the longranger alignment step?
- More recent versions of hifiasm offer a greatly improved phasing module based on HiC reads. Due to the high HiFi coverage of 40X it would be interesting to see if this could create fully phased chromosomes.
- I was surprised that only 5 scaffolds contain the telomer motifs at both ends, 18 scaffolds (plus 1 contig) show one telomer sequence, and 6 scaffolds do not show the telomer sequence. Could it be possible that purge_dups trimmed off those sequence motifs from the end of the contigs or even removed those contigs?

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: genome assembly, VGP, Bat1K, ERGA, eurofish, bioinformatics

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.