






DATA NOTE

The genome sequence of the Emperor moth, *Saturnia pavonia* (Linnaeus, 1758) [version 1; peer review: 2 approved, 2 approved with reservations]

Liam M. Crowley ¹, Ellen Baker ¹, Peter W. H. Holland 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

¹University of Oxford, Oxford, England, UK

v1 First published: 19 Feb 2024, 9:48
<https://doi.org/10.12688/wellcomeopenres.20652.1>
Latest published: 19 Feb 2024, 9:48
<https://doi.org/10.12688/wellcomeopenres.20652.1>

Abstract

We present a genome assembly from an individual male *Saturnia pavonia* (the Emperor moth; Arthropoda; Insecta; Lepidoptera; Saturniidae). The genome sequence is 489.9 megabases in span. Most of the assembly is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. The mitochondrial genome has also been assembled and is 15.29 kilobases in length. Gene annotation of this assembly on Ensembl identified 11,903 protein coding genes.

Keywords

Saturnia pavonia, Emperor moth, genome sequence, chromosomal, Lepidoptera






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

Open Peer Review

Approval Status

	1	2	3	4
version 1				
19 Feb 2024	view	view	view	view

1. **Christopher B Cunningham**, University of Georgia, Athens, USA
2. **Jaakko Pohjoismäki** , University of Eastern Finland, Joensuu, Finland
3. **Peter Dearden** , University of Otago, Dunedin, New Zealand
4. **Marco Gerdol** , University of Trieste, Trieste, Italy

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: **Crowley LM:** Investigation, Resources; **Baker E:** Investigation, Resources; **Holland PWH:** Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, <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.*

Copyright: © 2024 Crowley LM *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: Crowley LM, Baker E, Holland PWH *et al.* **The genome sequence of the Emperor moth, *Saturnia pavonia* (Linnaeus, 1758) [version 1; peer review: 2 approved, 2 approved with reservations]** Wellcome Open Research 2024, 9:48 <https://doi.org/10.12688/wellcomeopenres.20652.1>

First published: 19 Feb 2024, 9:48 <https://doi.org/10.12688/wellcomeopenres.20652.1>

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphimesenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Bombycoidea; Saturniidae; Saturniinae; Saturniini; *Saturnia*; *Eudia*; *Saturnia pavonia* (Linnaeus, 1758) (NCBI:txid332931).

Background

The family Saturniidae, named after the ringed planet Saturn, includes over 2000 species of moth worldwide and includes emperor moths and giant silk moths. Two species of Saturniidae have been recorded in Britain: *Saturnia pavonia* (the Emperor moth or Small Emperor moth, a resident breeding species) and *S. pyri* (the Giant peacock moth, the few records likely to be escapees from captive breeding). As befits the family and genus name, *S. pavonia* has large eyespots in the centre of all four wings, each with a black centre surrounded by a buff yellow ring and arcs formed from blue or red scales. The four coloured eyespots give the moth a striking appearance, especially as the moth is one of the largest in Britain or Northern Europe with a wingspan of 60 mm in males or 80 mm in females. Ford (1967) speculated that a hypothesised role of the eye-spots in predator deterrence could be accentuated immediately after eclosion from the pupa, when the wings are still puckered.

The Emperor moth *S. pavonia* is found widely across Europe, from the north of Finland and Norway to the south of Spain and Italy. The range also extends further east in Eurasia to Russia, Kazakhstan and Mongolia (GBIF Secretariat, 2023). In Britain and Ireland, the moth has been recorded from the south coast of England to the north of Scotland, but is commonest on acid heathlands, moorlands and coastal sand dunes (NBN Atlas Partnership, 2023).

Adult males are day-flying and particularly active on bright warm days in April and May, when they may be seen flying in search of scent trails released by receptive females. Much has been written about the potency of the *S. pavonia* female sex pheromones and the distances over which males detect females, but many estimates are based conjecture rather than controlled experimentation. After mating, eggs are laid on the foodplant and larvae develop rapidly through summer. On heathlands, the preferred larval foodplant is heather, *Calluna vulgaris*, but the larvae will also eat leaves of hawthorn, apple, bramble and many other shrubs and trees (South, 1961). Late instar larvae are bright green with black hoops studded with yellow globular outgrowths (scoli); a genetic variant with pink-coloured scoli has also been reported (Laussmann *et al.*, 2012). When physically stimulated, hollow spines on the scoli secrete a blend of proteins and aromatic compounds thought to deter pathogens or predators; benzonitrile is the dominant small molecule but the secreted proteins have not been identified (Deml & Dettner, 1990; Laussmann *et al.*, 2012). Larvae pupate inside a tough silken cocoon attached to stems and twigs of the foodplant near the ground. The

species overwinters at the pupal stage, sometimes for two or more winters (Laussmann *et al.*, 2012).

Here we report a complete genome sequence for the Emperor moth *Saturnia pavonia* determined as part of the Darwin Tree of Life project. The genome sequence of *S. pavonia* will facilitate research into the biochemical basis of chemical defence and pheromone communication in insects, and contribute to the growing set of resources for studying molecular evolution in the Lepidoptera.

Genome sequence report

The genome was sequenced from one male *Saturnia pavonia* (Figure 1) reared from a larva collected in Wytham Woods, Oxfordshire, UK (51.77, -1.33). A total of 50-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 6 missing joins or mis-joins and removed 2 haplotypic duplications, reducing the scaffold number by 13.51%.

The final assembly has a total length of 489.9 Mb in 31 sequence scaffolds with a scaffold N50 of 17.7 Mb (Table 1). The snailplot 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 (99.98%) of the assembly sequence was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not

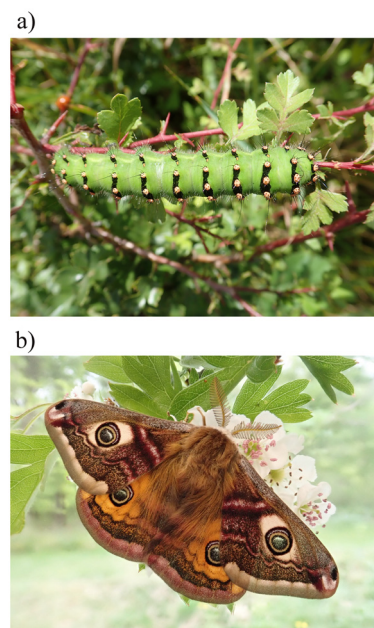


Figure 1. Photograph of the *Saturnia pavonia* (IlSatPavo1) specimen used for genome sequencing; **a)** Larva, **b)** Adult.

Table 1. Genome data for *Saturnia pavonia*, ilSatPavo1.1.

Project accession data		
Assembly identifier	ilSatPavo1.1	
Species	Saturnia pavonia	
Specimen	ilSatPavo1	
NCBI taxonomy ID	332931	
BioProject	PRJEB57274	
BioSample ID	SAMEA110451586	
Isolate information	ilSatPavo1, male: thorax (DNA and Hi-C sequencing), abdomen (RNA sequencing)	
Assembly metrics*		Benchmark
Consensus quality (QV)	68.6	≥ 50
k-mer completeness	100.0%	≥ 95%
BUSCO**	C:98.6%[S:98.4%,D:0.2%] ,F:0.4%,M:1.0%,n:5,286	C ≥ 95%
Percentage of assembly mapped to chromosomes	99.98%	≥ 95%
Sex chromosomes	Z	localised homologous pairs
Organelles	Mitochondrial genome: 15.29 kb	complete single alleles
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10462077	
Hi-C Illumina	ERR10466811	
PolyA RNA-Seq Illumina	ERR11606290	
Genome assembly		
Assembly accession	GCA_947532125.1	
Accession of alternate haplotype	GCA_947532135.1	
Span (Mb)	489.9	
Number of contigs	72	
Contig N50 length (Mb)	13.2	
Number of scaffolds	31	
Scaffold N50 length (Mb)	17.7	
Longest scaffold (Mb)	22.76	
Genome annotation		
Number of protein-coding genes	11,903	
Number of non-coding genes	1,739	
Number of gene transcripts	22,289	

* 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/CANNWM01/dataset/CANNWM01/busco>.

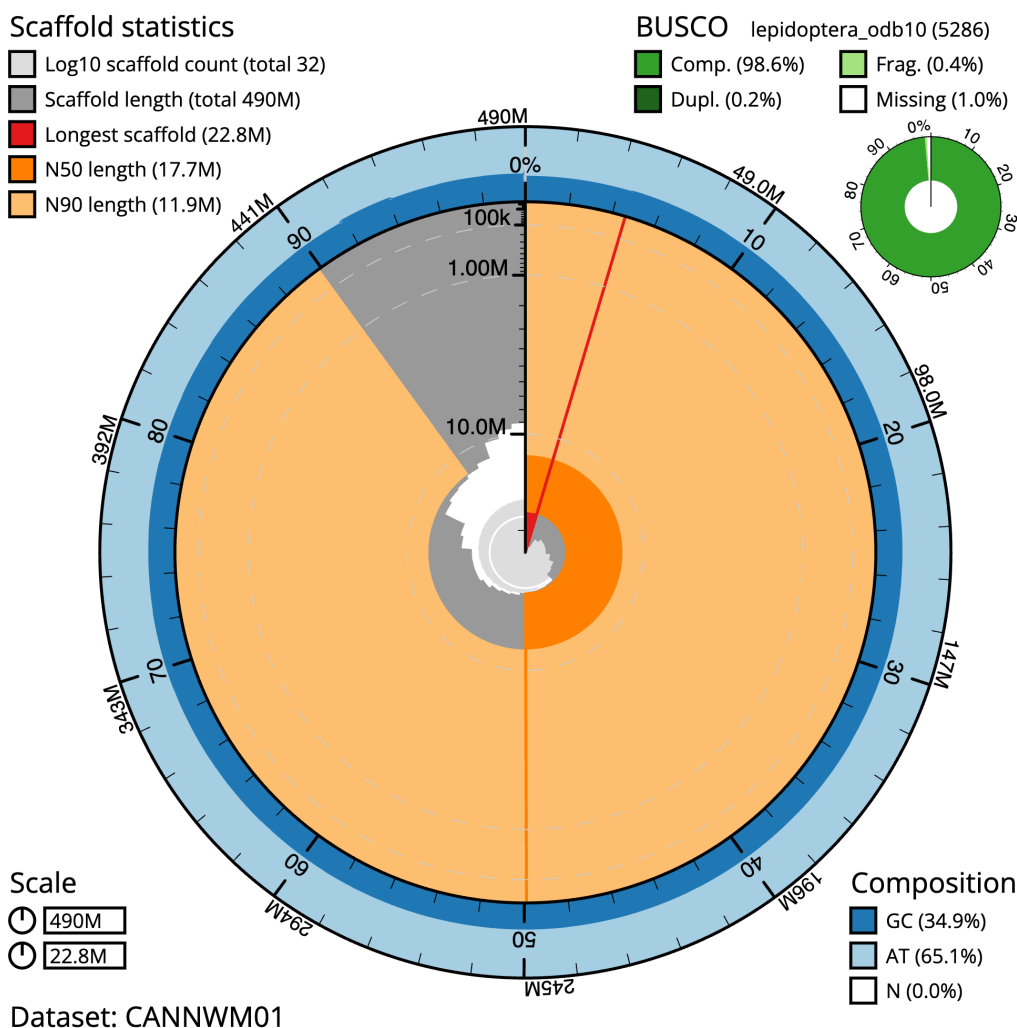


Figure 2. Genome assembly of *Saturnia pavonia*, ilSatPavo1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 489,898,868 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (22,760,314 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (17,680,408 and 11,900,431 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/CANNWM01/dataset/CANNWM01/snail>.

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 68.6 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 98.6% (single = 98.4%, 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/332931>.

Genome annotation report

The *Saturnia pavonia* genome assembly (GCA_947532125.1) was annotated using the Ensembl rapid annotation pipeline (Table 1; https://rapid.ensembl.org/Saturnia_pavonia_GCA_947532125.1/Info/Index). The resulting annotation includes 22,289 transcribed mRNAs from 11,903 protein-coding and 1,739 non-coding genes.

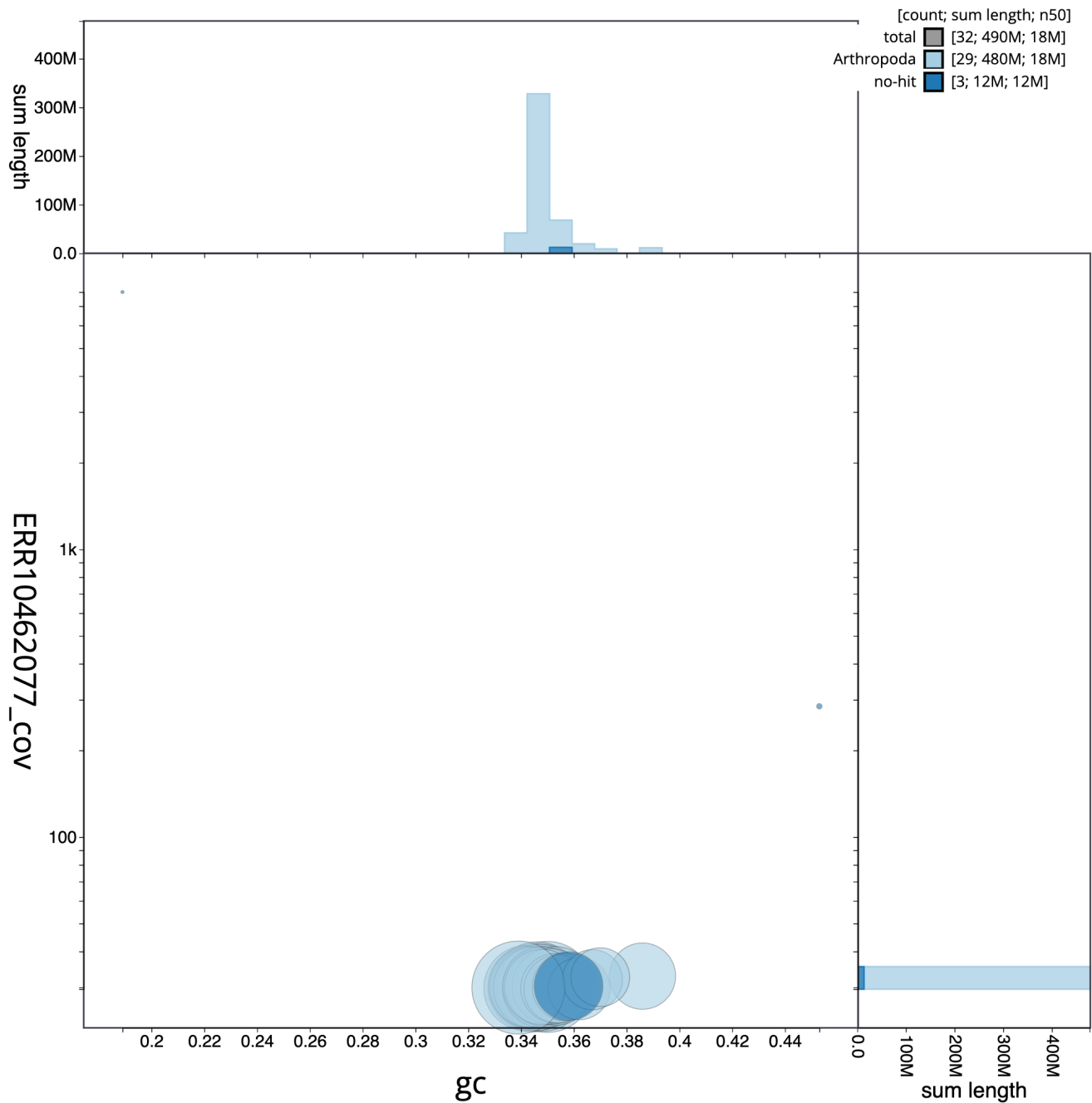


Figure 3. Genome assembly of *Saturnia pavonia*, ilSatPavo1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CANNWM01/dataset/CANNWM01/blob>.

Methods

Sample acquisition and nucleic acid extraction

The *Saturnia pavonia* specimen used for genome sequencing and Hi-C data (specimen ID Ox002139, ToLID ilSatPavo1) was collected as a larva from Wytham Woods, Oxfordshire

(biological vice-county Berkshire), UK (latitude 51.77, longitude -1.33) on 09/08/2021 by Ellen Baker (University of Oxford). The larva was reared by Liam Crowley (University of Oxford). The adult moth eclosed on 02/05/2022 and was preserved on dry ice.

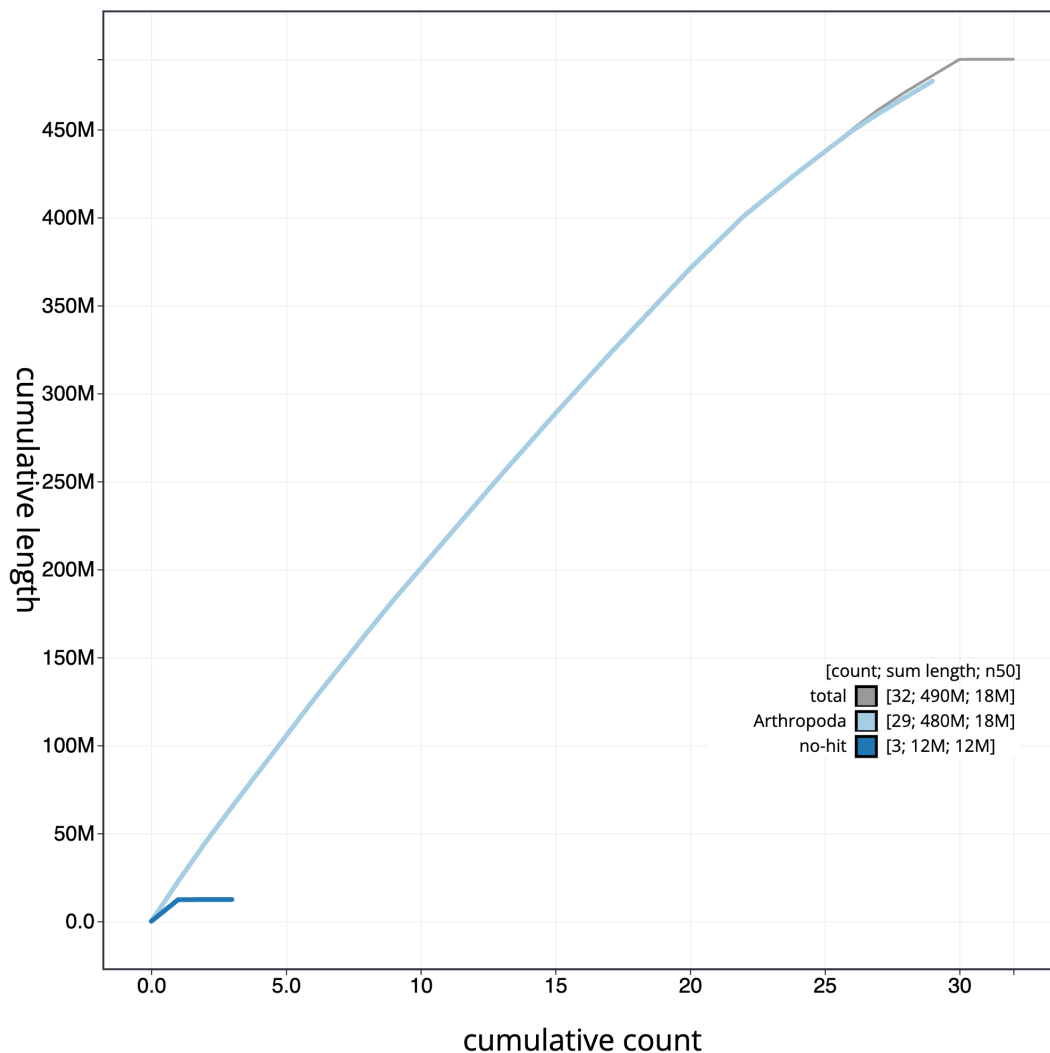


Figure 4. Genome assembly of *Saturnia pavonia*, ilSatPavo1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CANNWM01/dataset/CANNWM01/cumulative>.

Protocols developed by the Wellcome Sanger Institute (WSI) Tree of Life core laboratory have been published on protocols.io (Denton *et al.*, 2023b). The workflow for high molecular weight (HMW) DNA extraction at the WSI includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the ilSatPavo1 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 in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 31 (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): 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 abdomen tissue of ilSatPavo1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana 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.

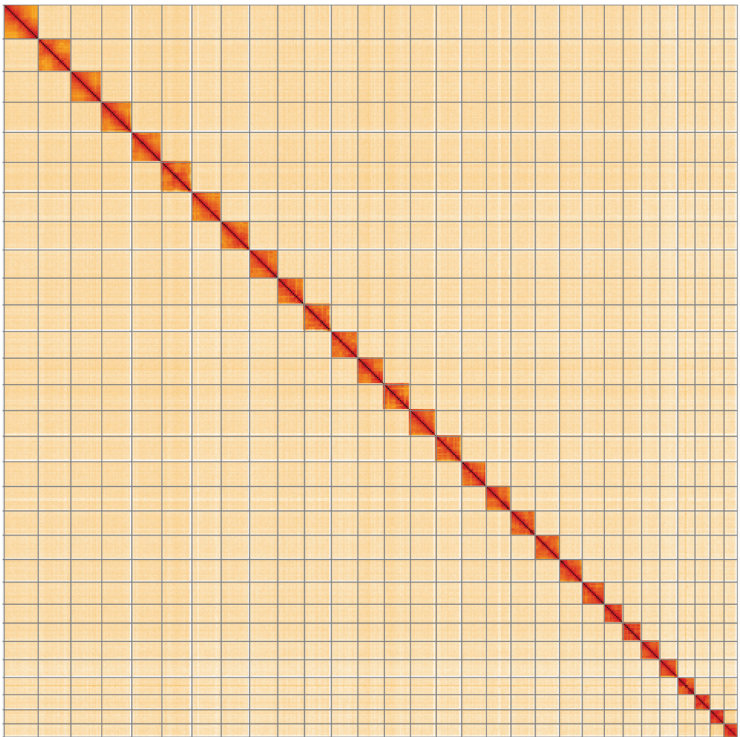


Figure 5. Genome assembly of *Saturnia pavonia*, ilSatPavo1.1: Hi-C contact map of the ilSatPavo1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at <https://genome-note-higlass.tol.sanger.ac.uk/I/?d=SljOrrzkSG67snu-sBNiBg>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Saturnia pavonia*, ilSatPavo1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX383895.1	1	21.77	35.0
OX383896.1	2	20.59	34.5
OX383897.1	3	20.11	34.5
OX383898.1	4	20.09	34.5
OX383899.1	5	20.08	34.5
OX383900.1	6	19.53	34.5
OX383901.1	7	18.97	34.0
OX383902.1	8	18.82	34.5
OX383903.1	9	17.89	34.5
OX383904.1	10	17.84	34.5
OX383905.1	11	17.8	34.5
OX383906.1	12	17.68	34.5
OX383907.1	13	17.4	34.5
OX383908.1	14	17.22	34.5

INSDC accession	Chromosome	Length (Mb)	GC%
OX383909.1	15	17.0	34.5
OX383910.1	16	16.6	35.0
OX383911.1	17	16.28	35.0
OX383912.1	18	16.26	35.5
OX383913.1	19	16.23	35.0
OX383914.1	20	15.11	35.0
OX383915.1	21	14.8	35.0
OX383916.1	22	12.6	35.5
OX383917.1	23	12.32	36.0
OX383918.1	24	12.19	35.5
OX383919.1	25	11.9	35.5
OX383920.1	26	11.49	38.5
OX383921.1	27	10.06	36.0
OX383922.1	28	9.4	36.5
OX383923.1	29	8.99	37.0
OX383894.1	Z	22.76	34.0
OX383924.1	MT	0.02	19.0

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 NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from remaining thorax tissue of *ilSatPavo1* 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 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 *Saturnia pavonia* assembly (GCA_947532125.1). 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.

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

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhyllp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
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
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	1.1a.2	https://github.com/c-zhou/yahs

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: *Saturnia pavonia* (emperor moth). Accession number PRJEB57274; <https://identifiers.org/ena.embl/PRJEB57274> (Wellcome Sanger Institute, 2022). The genome sequence is released openly for reuse. The *Saturnia pavonia* 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](#)
- 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](#)
- 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](#)
- 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](#)
- Deml R, Dettner K: **Chemical defense of *Eudia (Saturnia) pavonia* caterpillars.** *Naturwissenschaften.* 1990; **77**: 588–590. [Publisher 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](#)
- do Amaral RJV, Bates A, Denton A, et al.: **Sanger Tree of Life RNA Extraction: Automated MagMax[™] mirVana.** *protocols.io.* 2023. [Publisher Full Text](#)
- Ford EB: **Moths.** London: New Naturalist, Collins, 1967.
- GBIF Secretariat: ***Saturnia pavonia* Linnaeus, 1758.** *GBIF Backbone Taxonomy.* 2023; [Accessed 20 November 2023]. [Reference Source](#)
- 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): g1aa153. [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](#)
- Laussmann T, Bossems J, Metzger S, et al.: **Beiträge zur Biologie des Kleinen**

Nachtpfauenauges, *Saturnia pavonia* (Linnaeus, 1758). *Entomologie Heute*. 2012; **24**: 137–157.

[Reference Source](#)

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](#)

NBN Atlas Partnership: *Saturnia pavonia* (Linnaeus, 1758) Emperor Moth. *NBN Atlas*. 2023; [Accessed 20 November 2023].

[Reference Source](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2.** *Protocols.io*. 2023.

[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](#)

South R: **Moths of the British Isles**. New edition. London: Frederick Warne and Co, 1961.

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Manual SPRI.** *protocols.io*. 2023.

[Publisher Full Text](#)

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

[Publisher Full Text](#)

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

[Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, 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](#)

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res*. 2019; **47**(D1): D506–D515.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.

[Publisher Full Text](#)

Wellcome Sanger Institute: **The genome sequence of the Emperor moth, *Saturnia pavonia* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB57274, 2022.

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 21 May 2024

<https://doi.org/10.21956/wellcomeopenres.22857.r82138>

© 2024 Gerdol M. 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.



Marco Gerdol 

University of Trieste, Trieste, Italy

The manuscript by Crowley and colleagues reports the genome assembly of the Emperor moth, *Saturnia pavonia*. This resource was obtained as a part of the Darwin Tree of Life project, through the use of a highly standardized and reproducible pipeline. Therefore, there are no significant methodological concerns, and the quality of the assembled genome appears to be very high. I just have a very few minor comments dealing with cosmetic changes that the authors may consider including in a revised version.

"(51.77, -1.33)" -> I am not sure this is the standard or most common way of reporting geographical coordinates. Please add "latitude" and "longitude".

Genome annotation report: this was obtained through the Ensembl rapid annotation pipeline, which in other phyla (I am not really sure about arthropods) performs rather poorly. Do the BUSCO metrics obtained by analyzing the annotated gene set match those reported in the previous paragraph by analyzing the genome assembly?

Data availability: consider adding here the link to the Ensembl rapid annotation report.

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: non-model species 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 17 May 2024

<https://doi.org/10.21956/wellcomeopenres.22857.r78300>

© 2024 Dearden P. 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.



Peter Dearden 

University of Otago, Dunedin, Otago, New Zealand

This report describes the genome sequencing of a large day-flying moth; the Emperor moth (*Saturnia pavonia*). The report describes the key biological details of the moth well, including the relationships between the sampled British populations, and those in the rest of Europe. Photos are provided of the sampled animal from both larval and adult stages.

The genome report indicates that this is a very high-quality, contiguous, uncontaminated genome. the BUSCO scores show the genome is highly complete, and indicates the high quality of the assembly.

The data generated is available, and the tools used to carry out the work, are well described and available to others.

I strongly support this report. I would ask that in the section where the authors are describing the correction of scaffolds that they state the numbers of scaffolds affected, rather than just a percentage.

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

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 17 May 2024

<https://doi.org/10.21956/wellcomeopenres.22857.r82144>

© 2024 Pohjoismäki J. 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.



Jaakko Pohjoismäki

University of Eastern Finland, Joensuu, Finland

The genome report by Liam M Crowley, Ellen Baker, Peter WH Holland and the DToL consortium presents the reference genome assembly of the Emperor moth. It is a nice addition of an iconic species to the growing number of Lepidopteran genomes from the DToL.

Background

The background is well written and concise. One could add that *pavonia* is replaced by *pavoniella* in the east and that the two species have been separated only recently. This close relatedness might then deserve a comment regarding the utility of the genome assembly? Also, the pavonia is fairly early spring species, so I would imagine that May is too late for the flight time in the UK? As of note, there are many old studies on the pheromone attracting marked males, but these will require a bit of digging as they are difficult to find. Main point is that not all estimates are based on "conjecture".

The larvae are also sexually dimorphic, but this is maybe an unnecessary detail.

The genome assembly, annotation and completeness looks very good, representing the DToL quality standards. The methods are adequately described and following the DToL genome production pipeline without any issues.

My main complaint is that the genome assembly is from a male and thus lacks the W sex chromosome. I think this is a missed opportunity to collect valuable genome data on the sex chromosome evolution among insects. For *Saturnia pavonia* missing this opportunity is especially unnecessary, as the females are easily available from light traps, collecting as larvae or bought from a (British) web shop (<https://www.wwb.co.uk/emperor-moth-pavonia-cocoons>).

However, this detail does not otherwise affect the quality of the work. So far the use of males for the genome assemblies has been a recurrent issue with the Lepidopteran genomes. I hope this will be avoided in the future and females will be used for the coming species.

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: taxonomy; genetics; genomics; insects; mammals; molecular biology

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, however I have significant reservations, as outlined above.

Reviewer Report 10 May 2024

<https://doi.org/10.21956/wellcomeopenres.22857.r80533>

© 2024 Cunningham C. 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.



Christopher B Cunningham

University of Georgia, Athens, Georgia, USA

I think this is a good general resource for the genomics community. The methods are sound and relatively well documented. The article needs a few more details to be reasonably reproducible.

Abstract. There is no justification within the abstract, just results. At least a sentence that this was part of a large survey of insects is needed. And one sentence for its possible value.

Keywords. Not present.

Background. There is little true motivation for this study in the introduction. Nothing wrong is said, but the reader has little idea why the work was done. It's perfectly fine that the work was done in the context of a large survey, but tell the reader that. Almost all the details of its life history are irrelevant to the reader. The paper is not about its ecology or morphology, but about its genetics. That is not mentioned once.

The end paragraph mentions it is an existing or potential model for chemical defenses and pheromone communication but not direct justification or citation of why this species is particularly useful for these is given.

Figure 2/3/4/5 titles. These should be informative and tell the reader what you would like them to understand about the data; e.g., Fig 3. Little contamination was found in the final genome assembly.

Figure 4. has very little information and can easily be summarized as a sentence in the results. The

same is true of Fig 5

All parameter setting if they were different from default of each piece of bioinformatic software needs to be specified. Just put a sentence to begin that everything was default unless specified otherwise. What assessment was done to ensure that default was adequate?

A reasonable amount of genes is predicted. However, the reader has very little understanding of how these were evaluated beyond their initial generation. What QC beyond an automated annotation was done? One annotation run with default parameters should come with a strong disclaimer that any further work on specific gene families requires QC on the part of the user because the current annotation is a guide only.

What is the BUSCO percentage of the predicted gene set? That is an absolute minimum for a reader to be able to assess the minimum quality of the annotation.

Table 3 is of little to no value and contains much duplicated information. Just add the version numbers of software used inline at the appropriate places in the methods section.

Is the Ethics statement needed? It just says the provider of the sample should meet some agreed upon standard, but does not actually say that they did in this case. Either drop the statement or actually say the standard was met for this sample. This is not a paper about the standards of collection.

Add something to the report that contextualize the research for the reader. Is this species the first of a taxon to be assembled, is it an outgroup to some established model species, is it now possible to investigate some interesting aspect of the species biology, etc? This does not to be extensive or highly directed, but there is currently no justification of this work outside of the Data Availability Statement. Even Data Note should contain rationale for the work.

Is the rationale for creating the dataset(s) clearly described?

Partly

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: Genomics, genetics, epigenetics, behavior, reproduction

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, however I have significant reservations, as outlined above.