DATA NOTE

The genome sequence of the ringed china-mark, *Parapoynx stratiotata* (Linnaeus, 1758) [version 1; peer review: 1 approved]

Douglas Boyes[1+], Richard Chadd[2+],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Natural History Museum 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, Peter Mulhair[3],
Darwin Tree of Life Consortium

[1]UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK
[2]UK Environment Agency, Peterborough, UK
[3]Department of Zoology, University of Oxford, Oxford, UK

[+] Deceased author

v1 **First published:** 04 Apr 2022, **7**:121
https://doi.org/10.12688/wellcomeopenres.17808.1
Latest published: 04 Apr 2022, **7**:121
https://doi.org/10.12688/wellcomeopenres.17808.1

Abstract

We present a genome assembly from an individual male *Parapoynx stratiotata* (the ringed china-mark; Arthropoda; Insecta; Lepidoptera; Crambidae). The genome sequence is 478 megabases in span. The majority of the assembly (99.98%) is scaffolded into 30 chromosomal pseudomolecules, with the Z sex chromosome assembled. The mitochondrial genome was also assembled and is 15.4 kilobases in length.

Keywords

Parapoynx stratiotata, ringed china-mark, genome sequence, chromosomal, Lepidoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓

	1
version 1 04 Apr 2022	✓ view

1. **Clément Schneider**, LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany

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; **Chadd R**: Investigation, Resources; **Mulhair P**: 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, Chadd R, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the ringed china-mark, *Parapoynx stratiotata* (Linnaeus, 1758) [version 1; peer review: 1 approved]** Wellcome Open Research 2022, **7**:121 https://doi.org/10.12688/wellcomeopenres.17808.1

First published: 04 Apr 2022, **7**:121 https://doi.org/10.12688/wellcomeopenres.17808.1

Species taxonomy

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea; Crambidae; Nymphulinae; Parapoynx; *Parapoynx stratiotata* (Linnaeus, 1758) (NCBI:txid1594321).

Background

Parapoynx stratiotata (ringed china-mark) is a crambid moth with a recognisable dark ring with a white centre on the forewing. The female is larger (28–30 mm wingspan) and has a darker orange-brown forewing, while the males are smaller (20–24 mm wingspan) and have pale ochre forewings. *Paraponyx stratiotata* is a Palearctic species with a wide distribution across Europe. It is found across Britain and Ireland but is rarer in Scotland.

This species is part of the Acentropinae (Nymphulinae) subfamily, which have all adapted to life in freshwater environments (Léger *et al.*, 2021; Regier *et al.*, 2012). The larvae of this aquatic moth live completely submerged in water and rely on aquatic plants for food and habitat, while the adults are terrestrial. The female lays the eggs into water by placing her abdomen below the water surface. As such, the adults are found mostly near the margins of ponds and lakes, marshes and slow-flowing rivers. The larvae possess branched tracheal gills on all segments except the prothorax, which are used for breathing, and build a portable tube-like case (Vallenduuk & Cuppen, 2004), typical of the *Parapoynx*-type larva form. The larvae may be important components of freshwater habitats, altering water quality through stimulation of nutrient release, particularly phosphorus (Grutters *et al.*, 2016). The larvae are also phytophagous, like most lepidopterans, and were found to feed on both native and exotic plants pointing to their potential usefulness for biotic resistance to invading pest species (Grutters *et al.*, 2016; Habeck, 1983).

Paraponyx stratiotata is a key lepidopteran representative for the underrepresented aquatic insects (Hotaling *et al.*, 2020). Analysis of a genome sequence for this species will provide vital insight for understanding how insects, including Lepidoptera (Pabis, 2018), have adapted to live in aquatic environments.

The genome of *P. stratiotata* was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all of the named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *P. stratiotata*, based on one male specimen from Wytham Woods, Oxfordshire, UK.

Genome sequence report

The genome was sequenced from a single female *P. stratiotata* (ilParStra1; Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-country: Berkshire), UK (latitude 51.764, longitude -1.327). A total of 26-fold coverage in Pacific Biosciences single-molecule long reads and 101-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 23 missing/misjoins and removed 7 haplotypic duplications, reducing the assembly size by 1.02% and the scaffold number by 23.81%, and increasing the scaffold N50 by 3.54%.

The final assembly has a total length of 478 Mb in 32 sequence scaffolds with a scaffold N50 of 12.2 Mb (Table 1). Of the assembly sequence, 99.98% was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes (numbered by sequence length), and the Z sex chromosome (Figure 2–Figure 4; Table 2). The assembly has a BUSCO (Manni *et al.*, 2021) completeness of 98.4% (single 98.0%, duplicated 0.4%) 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

One male *P. stratiotata* (ilParStra1) and one *P. stratiotata* of unknown sex (ilParStra3) were collected from Wytham Woods, Oxfordshire, UK (ilParStra1: latitude 51.765, longitude -1.335; ilParStra3: latitude 51.764, longitude -1.327) by Douglas Boyes, UKCEH, from woodland using a light trap. The specimens were identified by the same individual and preserved on dry ice. A third (larval) specimen of unknown sex was collected from Snakeholm Pit, Lincolnshire, UK (latitude 53.230, longitude -0.330) by Richard Chadd, UK Environment Agency, from freshwater using a kicknet. The specimen was identified by the same individual and snap-frozen on dry ice.

DNA was extracted from whole organism tissue of ilParStra1 at the Wellcome Sanger Institute (WSI) Scientific Operations core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer's instructions. RNA was extracted from whole organism larval tissue of ilParStra2 in the Tree of Life Laboratory at the WSI using TRIzol (Invitrogen), according to the manufacturer's instructions. RNA was then eluted in 50 μl RNAse-free water and its concentration RNA assessed using a Nanodrop spectrophotometer and Qubit

Figure 1. Image of the *Parapoynx stratiotata* (ilParStra1) specimen taken prior to preservation and processing.

Table 1. Genome data for *Parapoynx stratiotata*, ilParStra1.1.

Project accession data	
Assembly identifier	ilParStra1.1
Species	*Parapoynx stratiotata*
Specimen	ilParStra1
NCBI taxonomy ID	NCBI:txid254363
BioProject	PRJEB45129
BioSample ID	SAMEA7519920
Isolate information	ilParStra1: female, adult whole organism (genome assembly); ilParStra2: unknown sex, larval whole organism (RNA-Seq); ilParStra3: unknown sex, adult whole organism (Hi-C)
Raw data accessions	
PacificBiosciences SEQUEL II	ERR6594500
10X Genomics Illumina	ERR6054811-ERR6054814
Hi-C Illumina	ERR6054816
PolyA RNA-Seq Illumina	ERR6054815
Genome assembly	
Assembly accession	GCA_910589355.1
Accession of alternate haplotype	GCA_910589245.1
Span (Mb)	478
Number of contigs	62
Contig N50 length (Mb)	13.8
Number of scaffolds	32
Scaffold N50 length (Mb)	17.1
Longest scaffold (Mb)	19.7
BUSCO* genome score	C:98.4%[S:98.0%,D:0.4%],F:0.4%,M:1.2%,n:5286

*BUSCO scores based on the lepidopetra_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/ilParStra1.1/dataset/CAJUUF01/busco.

Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud 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. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq X (10X) and Illumina HiSeq 4000 (RNA-Seq) instruments. Hi-C data were generated from abdomen tissue of ilParStra3 using the Arima v2 Hi-C kit and sequenced on an Illumina NovaSeq 6000.

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 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) as described

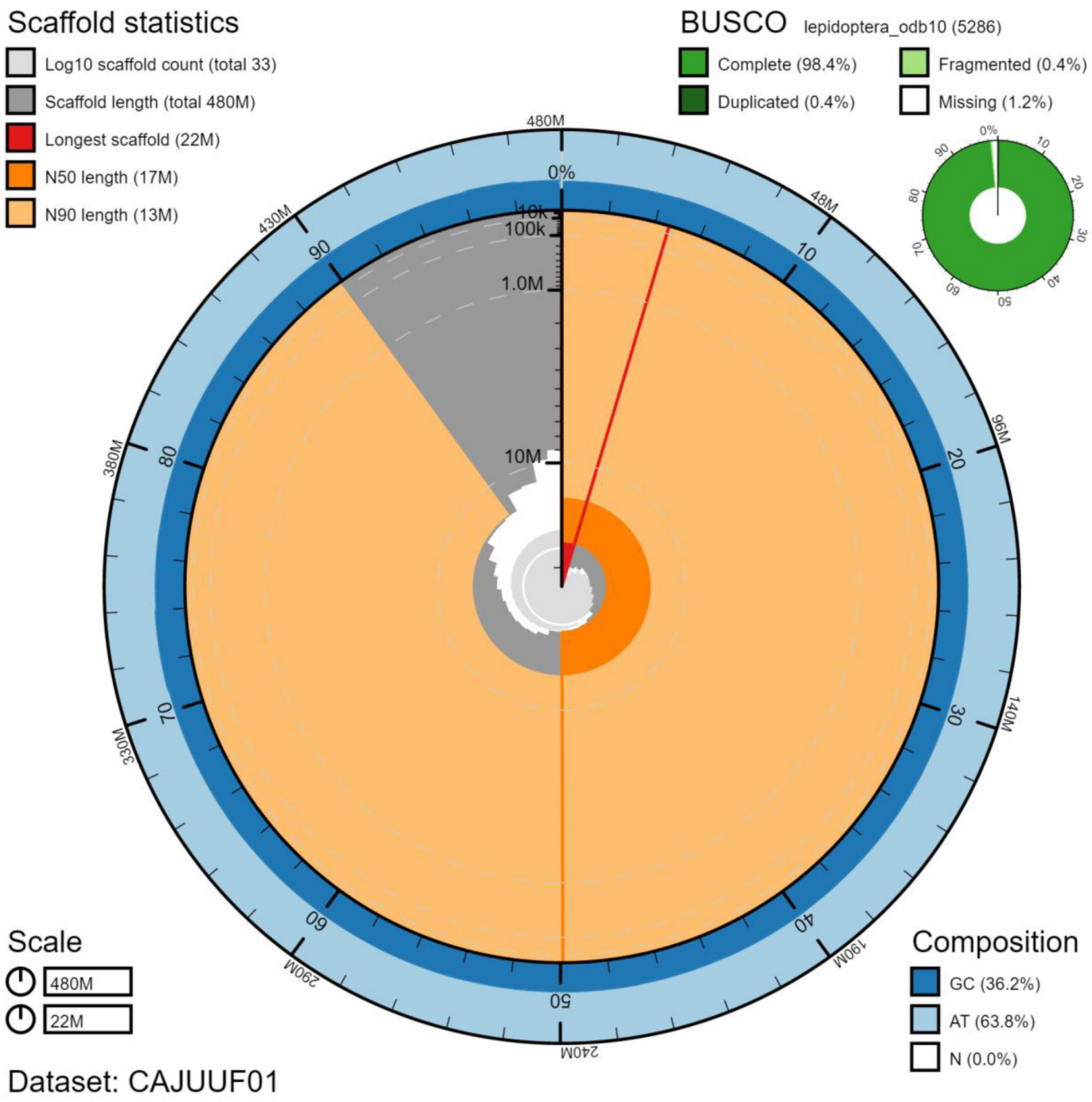

Figure 2. Genome assembly of *Parapoynx stratiotata*, ilParStra1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 478,248,748 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (22,161,367 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (17,072,597 and 12,933,045 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/ilParStra1.1/dataset/CAJUUF01/snail.

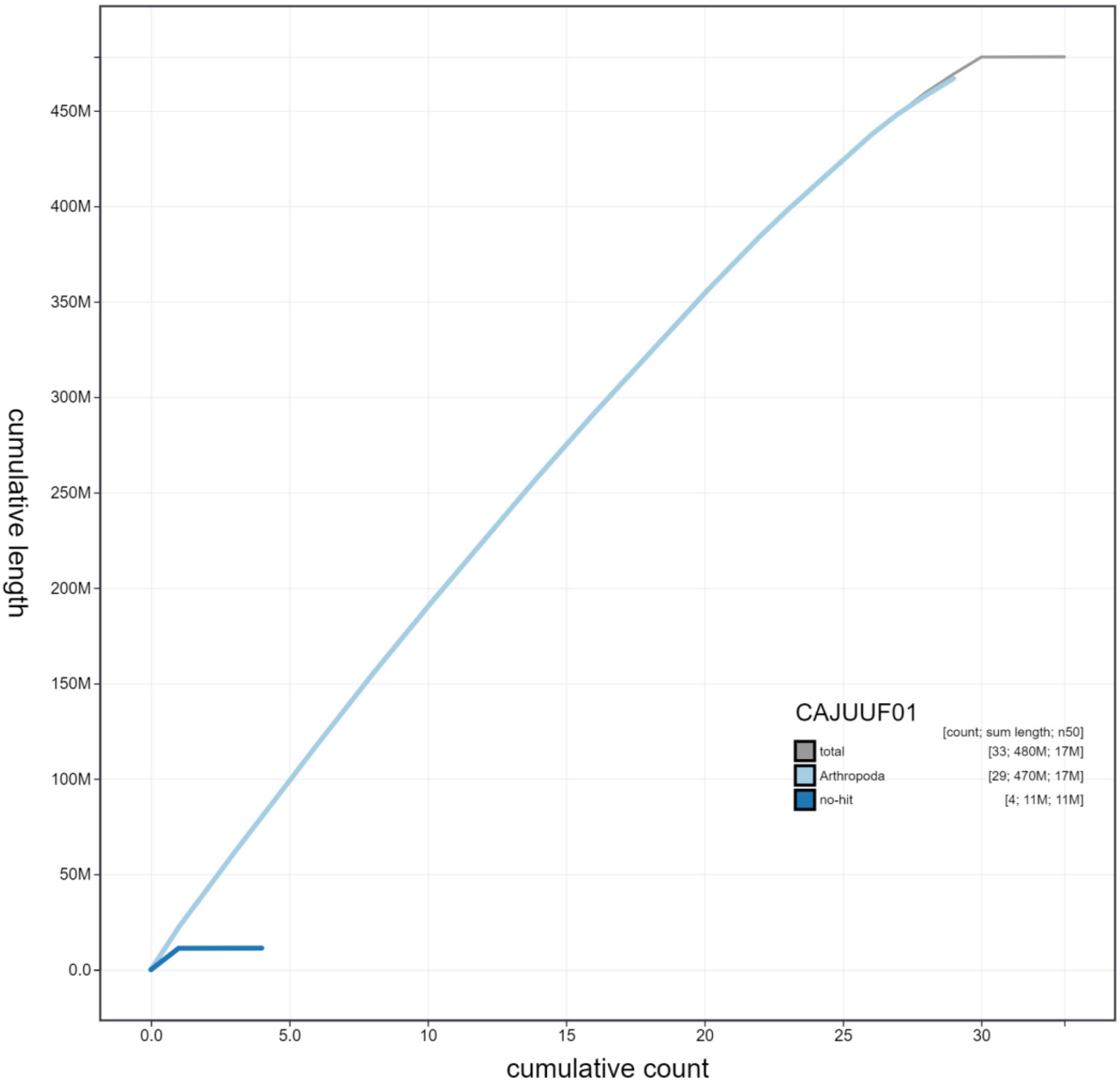

Figure 3. Genome assembly of *Parapoynx stratiotata*, ilParStra1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/ilParStra1.1/dataset/CAJUUF01/cumulative.

previously (Howe *et al.*, 2021). Manual curation (Howe *et al.*, 2021) 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.

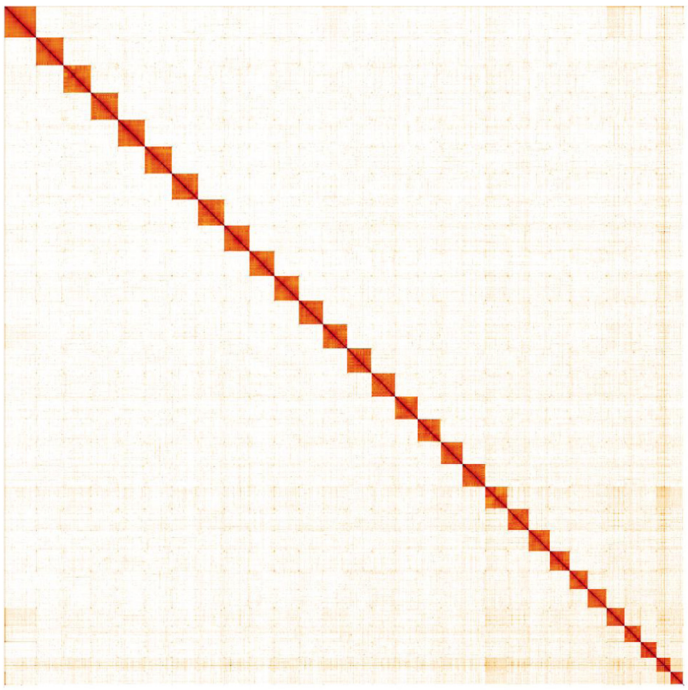

Figure 4. Genome assembly of *Parapoynx stratiotata*, ilParStra1.1: Hi-C contact map. Hi-C contact map of the ilParStra1.1 assembly, visualised in HiGlass. Chromosomes are shown in size order from left to right and top to bottom. The interactive Hi-C map can be viewed here.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Parapoynx stratiotata*, ilParStra1.1.

INSDC accession	Chromosome	Size (Mb)	GC%
LR989850.1	1	14.66	35.8
LR989851.1	2	13.96	35.9
LR989852.1	3	13.80	36.0
LR989853.1	4	13.71	35.0
LR989854.1	5	13.41	35.9
LR989855.1	6	13.21	35.4
LR989856.1	7	13.11	35.2
LR989857.1	8	13.03	34.8
LR989858.1	9	12.89	35.0
LR989859.1	10	12.78	35.3
LR989860.1	11	12.76	35.4
LR989861.1	12	12.18	35.4
LR989862.1	13	12.17	35.6
LR989863.1	14	12.03	34.9
LR989864.1	15	11.79	35.0
LR989865.1	16	11.78	35.5
LR989866.1	17	11.55	35.3
LR989867.1	18	11.41	35.6
LR989868.1	19	10.76	36.1
LR989869.1	20	10.65	35.1
LR989870.1	21	10.47	35.8
LR989871.1	22	9.66	35.0
LR989872.1	23	9.56	35.9
LR989873.1	24	9.48	34.9
LR989874.1	25	8.49	35.4
LR989875.1	26	8.14	34.8
LR989876.1	27	7.98	38.1
LR989877.1	28	6.89	36.1
LR989878.1	29	6.78	36.7
LR989879.1	30	5.62	37.4
LR989880.1	W	3.29	37.8
LR989849.1	Z	19.12	35.9
LR989881.1	MT	0.02	19.9
-	Unplaced	15.89	37.4

Table 3. Software tools used.

Software tool	Version	Source
Hifiasm	0.12-r304	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	2.0	Uliano-Silva *et al.*, 2021
gEVAL	0.2.x	Chow *et al.*, 2016
HiGlass	1.11.6	Kerpedjiev *et al.*, 2018
BlobToolKit	2.6.4	Challis *et al.*, 2020

Ethics/compliance issues

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. 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. 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: Parapoynx stratiotata (ringed china-mark). Accession number PRJEB45129; https://identifiers.org/ena.embl/PRJEB45129.

The genome sequence is released openly for reuse. The *P. stratiotata* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using the RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.5746938.

Members of the Natural History Museum Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.5746819.

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

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

Grutters BMC, Gross EM, Bakker ES: **Insect Herbivory on Native and Exotic Aquatic Plants: Phosphorus and Nitrogen Drive Insect Growth and Nutrient Release.** *Hydrobiologia.* 2016; **778**(1): 209–20.
Publisher Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98.
PubMed Abstract | Publisher Full Text | Free Full Text

Habeck DH: **Japm-21-01-026.pdf.** *J Aquat Plant Manage.* 1983.
Reference Source

Hotaling S, Kelley JL, Frandsen PB: **Aquatic Insects Are Dramatically Underrepresented in Genomic Research.** *Insects.* 2020; **11**(9): 601.
PubMed Abstract | Publisher Full Text | Free 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

Léger T, Mally R, Neinhuis C, *et al.*: **Refining the Phylogeny of Crambidae with Complete Sampling of Subfamilies (Lepidoptera, Pyraloidea).** *Zoologica Scripta.* 2021; **50**(1): 84–99.
Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–54.
PubMed Abstract | Publisher Full Text | Free Full Text

Pabis K: **What Is a Moth Doing under Water? Ecology of Aquatic and Semi-Aquatic Lepidoptera.** *Knowl Manag Aquat Ecosyst.* 2018; (419): 42.
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

Regier JC, Mitter C, Alma Solis M, *et al.*: **A Molecular Phylogeny for the Pyraloid Moths (Lepidoptera: Pyraloidea) and Its Implications for Higher-level Classification.** *Syst Entomol.* 2012; **37**(4): 635–56.
Publisher Full Text

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

Vallenduuk HJ, Cuppen HPJ: **The Aquatic Living Caterpillars (Lepidoptera: Pyraloidea: Crambidae) of Central Europe. A Key to the Larvae and Autecology.** 2004; **45**: 1–17.
Reference Source

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 22 April 2022

https://doi.org/10.21956/wellcomeopenres.19715.r49662

✔ **Clément Schneider**
LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany

The authors present a de novo chromosome-level assembly for a Lepidoptera. Species biology is briefly summarized, and genome assembly statistics are provided. The method is sound and up-to-date, the genome can be considered nearly complete and of high quality, making it a valuable and reliable resource. No debatable discussion is involved.

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: Entomology

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.