DATA NOTE

The genome sequence of the Suspected, *Parastichtis suspecta* (Hübner, 1809) (Lepidoptera: Noctuidae)

[version 1; peer review: 3 approved, 1 approved with reservations]

William B. V. Langdon[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

[1]University of Oxford, Oxford, England, UK

v1 **First published:** 18 May 2026, **11**:288
https://doi.org/10.12688/wellcomeopenres.26587.1
Latest published: 18 May 2026, **11**:288
https://doi.org/10.12688/wellcomeopenres.26587.1

Abstract

We present a genome assembly from an individual male *Parastichtis suspecta* (Suspected; Arthropoda; Insecta; Lepidoptera; Noctuidae). The assembly contains two haplotypes with total lengths of 868.84 megabases and 869.52 megabases. Most of haplotype 1 (99.8%) is scaffolded into 30 chromosomal pseudomolecules, including the Z sex chromosome. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.38 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces genomes for eukaryotic species found in Britain and Ireland.

Keywords

Parastichtis suspecta, Suspected, 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 18 May 2026	✓ view	? view	✓ view	✓ view

1. **Gulab D Khedkar**, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, India
2. **Anwesha Das**, Sanaka Educational Trust's Group of Institutions (SETGOI), West Bengal, India
3. **Oliver Rupp**, Justus-Liebig-University, Giessengiessen, Germany
4. **Jonathan Badger**, National Cancer Institute, Misrata, Libya
National Cancer Institute Center for Cancer Research, Bethesda, 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: Langdon WBV: 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 (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Langdon WBV, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of the Suspected, *Parastichtis suspecta* (Hübner, 1809) (Lepidoptera: Noctuidae) [version 1; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2026, **11**:288 https://doi.org/10.12688/wellcomeopenres.26587.1

First published: 18 May 2026, **11**:288 https://doi.org/10.12688/wellcomeopenres.26587.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Altocrustacea; Allotriocarida; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Eumetabola; Endopterygota; Aparaglossata; Panorpida; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Noctuidae; Xyleninae; *Parastichtis*; *Parastichtis suspecta* (Hübner, 1809) (NCBI: txid988145).

Background

We present a chromosome-level genome sequence for *Parastichtis suspecta* (Hübner, 1817), the Suspected (Lepidoptera: Noctuidae). The common name derives from the Latin *suspecta*, reflecting Hübner's apparent uncertainty about whether this moth could be trusted as a good species — a doubt that proved unfounded (Marren, 2019). The adult moth has a forewing length of 14–16 mm. The forewing is narrow at the base and rather tapered, and variable in colour, ranging from plain greyish brown to reddish or purplish brown, sometimes richly variegated with brownish yellow; variegated and richly marked forms predominate in northern and western Britain (Waring *et al.*, 2017). The species flies in a single generation from July to August, coming to light in small numbers and also to sugar and flowers, particularly ragwort. The larva feeds initially within a spun terminal shoot, later between folded leaves, on birches and probably sallows, and pupates in an underground cocoon (Waring *et al.*, 2017). Habitat associations include fens, carr, woodland, heathland, and moorland with birch scrub. In Great Britain, the species is fairly well distributed across southern and south-east England, East Anglia, and from north Wales and the Midlands north to Ross-shire, with immigrant records from Orkney and Shetland, probably originating from mainland Europe (Waring *et al.*, 2017).

The assembly was produced using the Tree of Life pipeline using a specimen collected from Silchester Common, Hampshire, United Kingdom (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Darwin Tree of Life Project Consortium, 2022).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Parastichtis suspecta* (specimen ID Ox003420, ToLID ilParSusp1; Figure 1), collected from Silchester Common, Hampshire, UK (latitude 51.3554, longitude −1.1139) on 2023-07-07. The specimen was collected and identified by Will Langdon.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode

Figure 1. Photograph of the *parastichtis suspecta* (ilParSusp1) specimen used for genome sequencing.

region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Nucleic acid extraction

Detailed protocols for nucleic acid extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The ilParSusp1 sample was weighed and triaged to determine the appropriate extraction protocol. For HMW DNA extraction, 24 mg of tissue from the head and thorax was used. The tissue was homogenised by powermashing using a PowerMasher II tissue disruptor. High molecular weight (HMW) DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation).

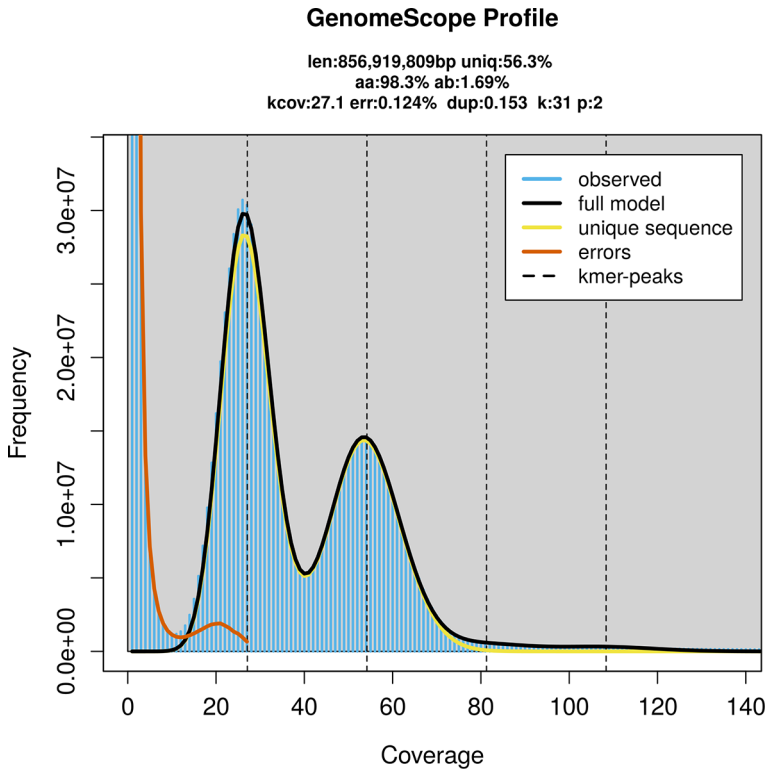

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for *parastichtis suspecta* (BioProject PRJEB80953).

Platform	PacBio HiFi	Hi-C
ToLID	ilParSusp1	ilParSusp1
Specimen ID	Ox003420	Ox003420
BioSample (source individual)	SAMEA114644489	SAMEA114644489
BioSample (tissue)	SAMEA114645099	SAMEA114645099
Tissue	head and thorax	head and thorax
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR13800497	ERR13802639
Read count total	4.47 million	411.18 million
Base count total	48.39 Gb	124.18 Gb

Table 2. **Genome assembly statistics for *parastichtis suspecta*.**

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	ilParSusp1.hap1.1	ilParSusp1.hap2.1
Assembly accession	GCA_964277095.1	GCA_964277165.1
Assembly level	chromosome	scaffold
Span (Mb)	868.84	869.52
Number of chromosomes	30	scaffold-level
Number of contigs	164	192
Contig N50	9.77 Mb	11.15 Mb
Number of scaffolds	59	92
Scaffold N50	29.41 Mb	29.64 Mb
Longest scaffold length (Mb)	73.72	-
Sex chromosomes	Z	-
Organelles	Mitochondrion: 15.38 kb	-

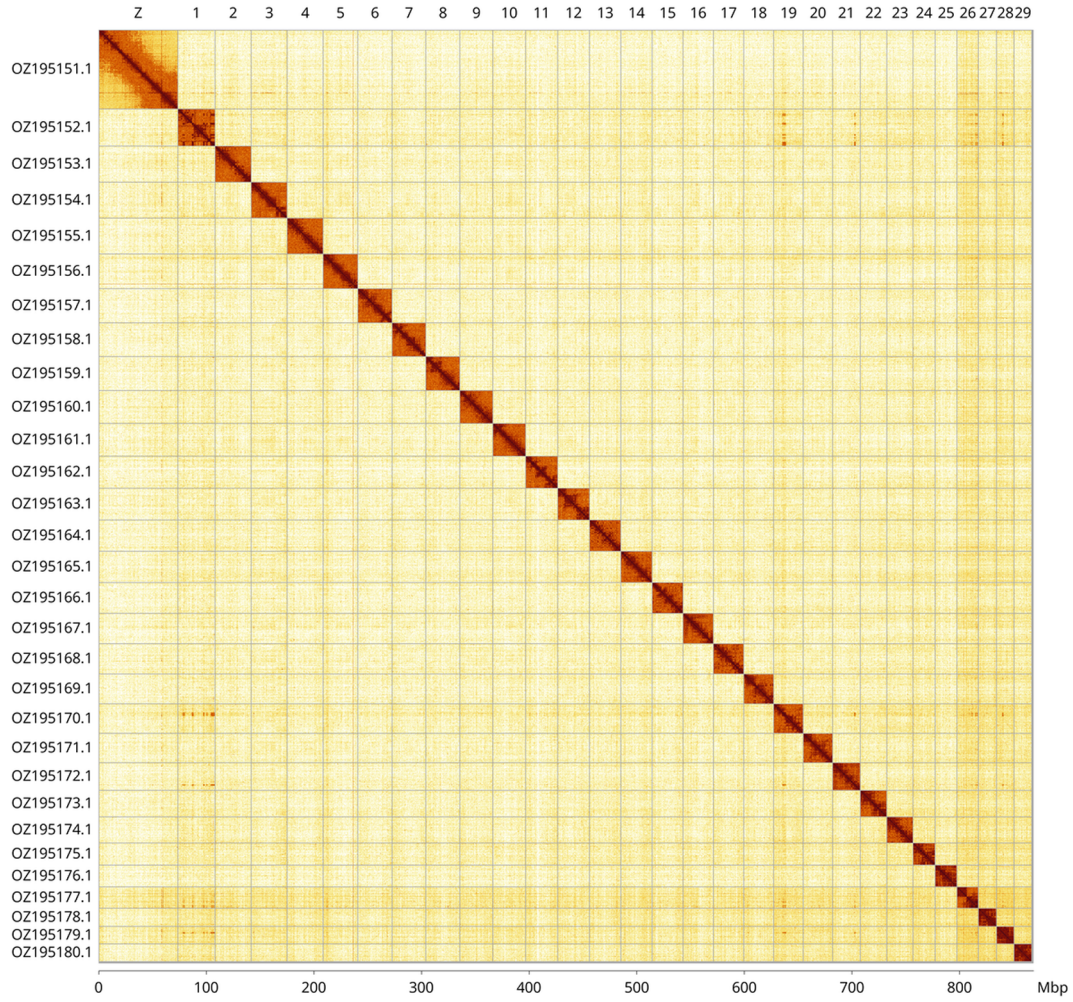

Figure 3. **Hi-C contact map of the *parastichtis suspecta* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 49.4 ng/μL and a yield of 2 321.80 ng, with a fragment size of 15.7 kb. The Genomic Quality Number (GQN) was 6.9.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *parastichtis suspecta* ilParSusp1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ195152.1	1	34.49	37.50
OZ195153.1	2	33.59	37.50
OZ195154.1	3	33.52	37.50
OZ195155.1	4	33.27	37.50
OZ195156.1	5	32.40	38
OZ195157.1	6	31.64	37.50
OZ195158.1	7	31.60	37.50
OZ195159.1	8	31.55	37.50
OZ195160.1	9	30.71	37.50
OZ195161.1	10	30.26	37.50
OZ195162.1	11	29.95	37.50
OZ195163.1	12	29.41	37.50
OZ195164.1	13	29.27	37.50
OZ195165.1	14	29.08	37.50
OZ195166.1	15	28.71	37.50
OZ195167.1	16	28.19	37.50
OZ195168.1	17	28.17	37.50
OZ195169.1	18	27.75	37.50
OZ195170.1	19	27.64	37.50
OZ195171.1	20	27.08	37.50
OZ195172.1	21	25.75	37.50
OZ195173.1	22	24.75	37.50
OZ195174.1	23	24.40	37.50
OZ195175.1	24	20.52	37.50
OZ195176.1	25	20.23	37.50
OZ195177.1	26	19.91	38
OZ195178.1	27	16.85	38
OZ195179.1	28	16.38	38.50
OZ195180.1	29	16.30	37.50
OZ195151.1	Z	73.72	37.50

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head and thorax tissue from the ilParSusp1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/µL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

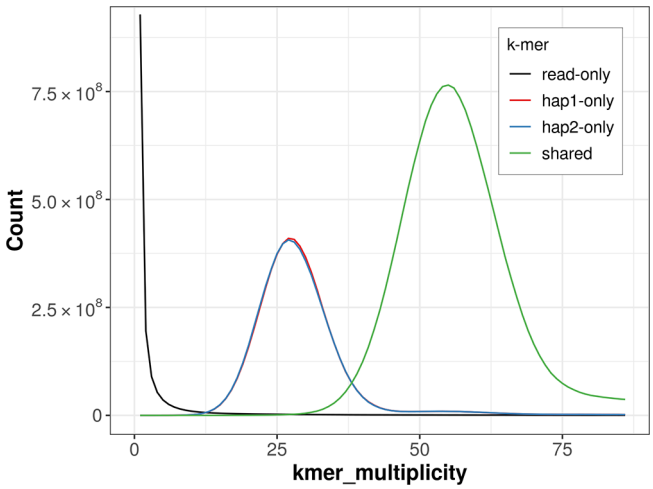

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MerquryFK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

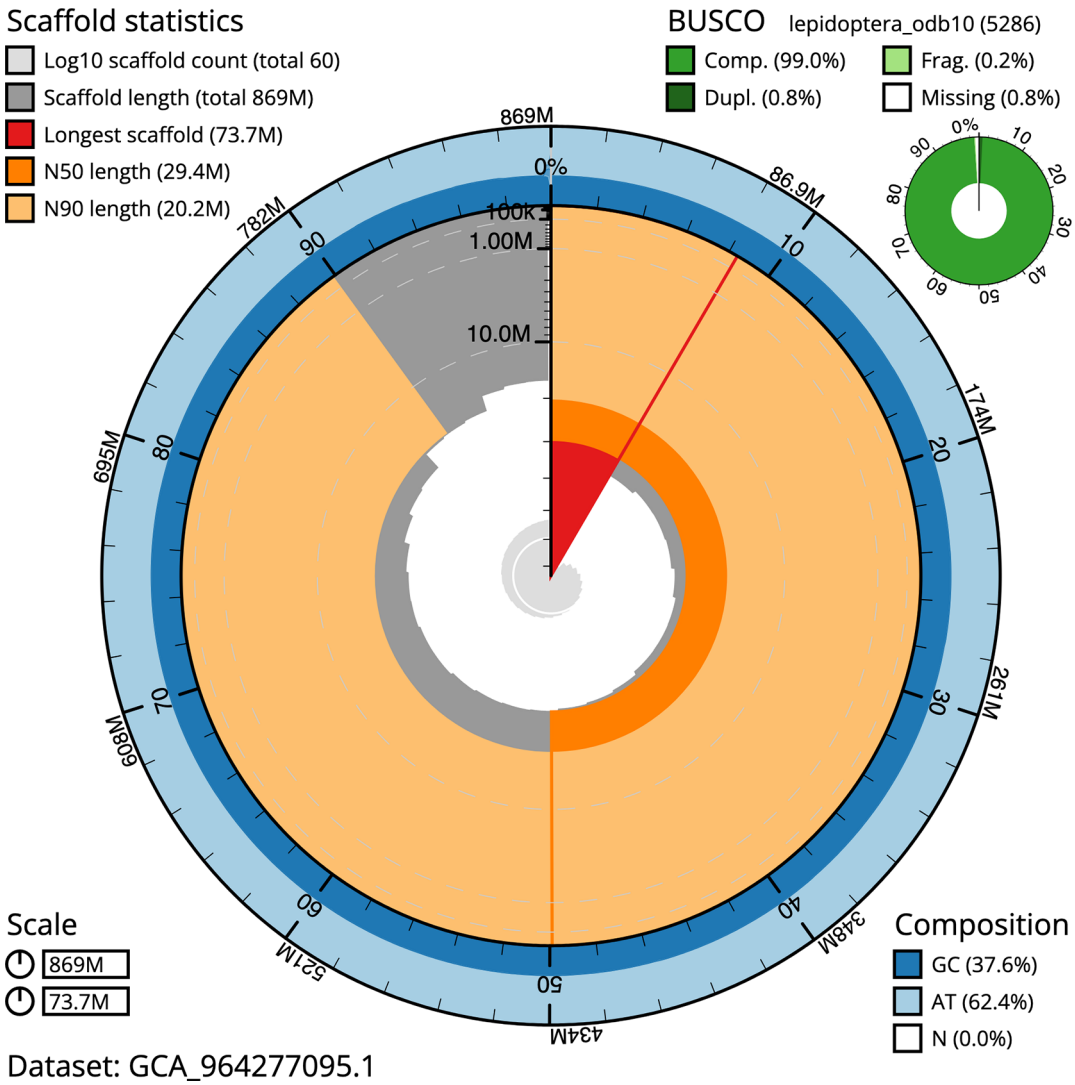

Figure 5. Assembly metrics for ilParSusp1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included five breaks and 12 joins. This reduced the scaffold count by 3.2%. The curation process is described at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The MerquryFK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer database (k = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021)

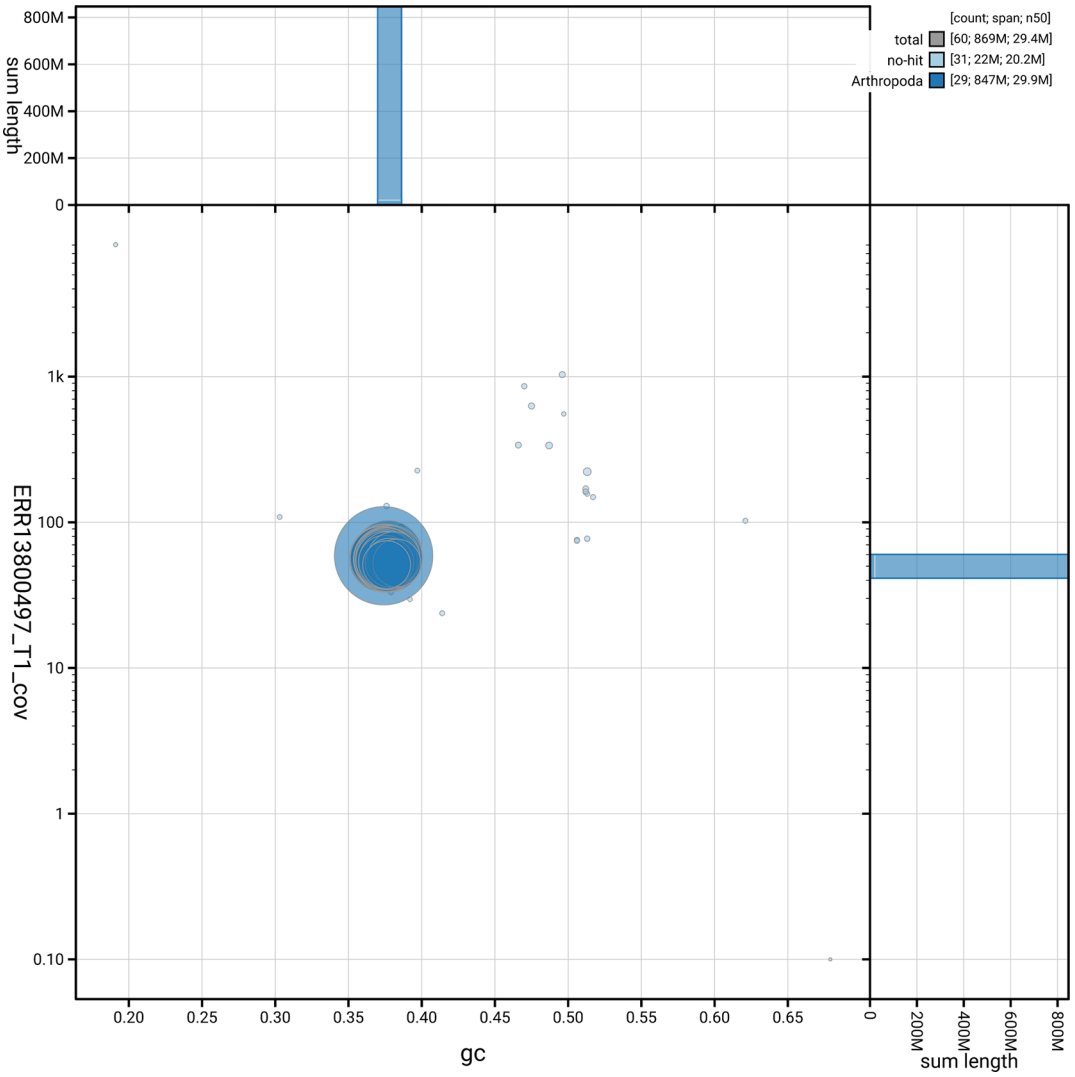

Figure 6. BlobToolKit blob plot for ilParSusp1.hap1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Parastichtis suspecta* specimen generated 48.39 Gb (gigabases) from 4.47 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 856.92 Mb, with a heterozygosity of 1.69% and repeat content of 43.76% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 54× coverage. Hi-C sequencing produced 124.18 Gb from 411.18 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 868.84 Mb in 59 scaffolds, with 105 gaps, and a scaffold N50 of 29.41 Mb (Table 2).

Most of the haplotype 1 assembly sequence (99.8%) was assigned to 30 chromosomal-level scaffolds, representing 29 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The Z sex chromosome was identified by homology to the genome of *Pyrrhia umbra.*

The mitochondrial genome was also assembled (length 15.38 kb, OZ195181.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 69.7, and for haplotype 2, 68.7. When the two haplotypes are combined, the assembly achieves an estimated QV of 69.2. The *k*-mer completeness is 70.91% for haplotype 1, 70.90% for haplotype 2, and 99.35% for the combined haplotypes (Figure 4).

BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) identified 99.0% of the expected gene set (single = 98.2%, duplicated = 0.8%) in haplotype 1. For haplotype 2, BUSCO v.5.5.0 analysis identified 99.0% of the expected gene set (single = 98.1%, duplicated = 0.9%). The snail plot in Figure 5 summarises the scaffold length

Table 4. Earth Biogenome Project summary metrics for the *Parastichtis suspecta* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q69	6.C.Q40
Contig N50 length	9.77 Mb	≥ 1 Mb
Scaffold N50 length	29.41 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 69.7; haplotype 2: 68.7; combined: 69.2	≥ 40
k-mer completeness	Haplotype 1: 70.91%; Haplotype 2: 70.90%; combined: 99.35%	≥ 95%
BUSCO	C:99.0% [S:98.2%, D:0.8%], F:0.2%, M:0.8%, n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.80%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues

Table 5. Software versions and sources used for *Parastichtis suspecta*.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards January 2026. The EBP metric, calculated for the haplotype 1, is **6.C.Q69**, meeting the recommended reference standard.

Data availability

European Nucleotide Archive: Parastichtis suspecta (suspected). Accession number PRJEB80953. The genome sequence is released openly for reuse. The *Parastichtis suspecta* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), Sanger Institute Tree of Life Programme (PRJEB43745) and Project Psyche (PRJEB71705). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Tables 1 and 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Author information

Contributors are listed at the following links:

- Members of the University of Oxford and Wytham Woods Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

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. 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.

References

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *Journal of Molecular Biology.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nature Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 Genes|Genomes|Genetics.* 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

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial Arthropod fauna.** *Wellcome Open Res.* 2023; **8**: 123.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
PubMed Abstract | Publisher Full Text | Free Full Text

Darwin Tree of Life Project Consortium: **Sequence locally, think globally: The Darwin Tree of Life Project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: Summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nature Biotechnology.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs.**

Bioinformatics. 2022; **38**(17): 4214–4216.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
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**

Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Li H: **Minimap2: Pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Marren P: *Emperors, Admirals and Chimney Sweepers.* Bridport: Little Toller Books; 2019.

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Publisher Full Text

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project.** *Wellcome Open Res.* 2024; **9**: 339.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: A Python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–24.
Publisher Full Text

Waring P, Townsend M, Lewington R: *Field Guide to the Moths of Great Britain and Ireland.* London: Bloomsbury Wildlife; 2017.

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 17 June 2026

https://doi.org/10.21956/wellcomeopenres.29280.r156712

✔ **Jonathan Badger**
[1] National Cancer Institute, Misrata, Libya
[2] Laboratory of Integrative Cancer Immunology (LICI), National Cancer Institute Center for Cancer Research, Bethesda, Maryland, USA

The authors describe a chromosome-level genome assembly of the noctuid moth *Parastichtis suspecta*, which inhabits Britain. The more complete haplotype was scaffolded into 30 pseudomolecules including the Z sex chromosome (which many insect genome projects fail to obtain, so this is good to see). The manuscript is correctly formatted for a data note and the standard metrics such as N50 and BUSCO are reported and look reasonable. I have checked to see if the data is in the ENA as reported and it is.

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: I am a computational genomicist. While I primarily work in the biomedical field, I have also worked on several insect genomes as well.

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 June 2026

https://doi.org/10.21956/wellcomeopenres.29280.r156714

Oliver Rupp
Justus-Liebig-University, Giessengiessen, Germany

This article presents the assembly of Parastichtis suspecta.
The data creation and data analyses follow state-of-the-art methods for eukaryotic genome assemblies established at the Darwin Tree of Life Consortium, including PacBio Hifi and Hi-C sequencing. The assembly with hifiasm in Hi-C mode was followed by Hi-C scaffolding and manual curation.
The assembly quality assessment using k-mer and orthology based methods like Merqury.FK or BUSCO showed a high-quality sequence for both haplotypes.
The methods are clearly presented, and the quality of the data is shown.
Overall, no major issues are to 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: genome assembly

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 04 June 2026

https://doi.org/10.21956/wellcomeopenres.29280.r156713

?

Anwesha Das
Sanaka Educational Trust's Group of Institutions (SETGOI), West Bengal, India

This manuscript entitled "The genome sequence of the Suspected, *Parastichtis suspecta* (Hübner, 1809) (Lepidoptera: Noctuidae)" reports a **chromosome-level genome assembly** for *Parastichtis suspecta* (Lepidoptera: Noctuidae), generated as part of the **Darwin Tree of Life (DToL) Project**.
The authors sequenced an individual male specimen collected in Hampshire, UK, using **PacBio HiFi long reads** (≈54× coverage) and **Hi-C data** for scaffolding. Key results include:
Assembly size: ~869 Mb for each haplotype
Haplotype 1: 99.8% assigned to **30 chromosomal pseudomolecules**, including the Z chromosome
Scaffold N50: ~29.4 Mb; contig N50 ~9.77 Mb
BUSCO completeness: ~99% (lepidoptera_odb10)
Estimated QV: ~69
Combined haplotype k-mer completeness: ~99.35%. The manuscript provides detailed methodological descriptions, sequencing protocols, assembly pipeline steps, and quality metrics. Data are publicly available via ENA. This work contributes to ongoing efforts to generate **reference genomes for biodiversity research**.
This is a **high-quality genome resource paper** consistent with DToL "data note" standards. The assembly appears technically robust and meets **Earth BioGenome Project (EBP)** benchmarks (6.C.Q69). However, as a scientific contribution, the manuscript is **limited in biological interpretation and contextualisation**. While this is acceptable for a data note, several aspects require clarification or expansion to ensure **scientific transparency, reproducibility, and utility**.

Major Strengths

High assembly quality
Chromosome-level scaffolding with strong contiguity and completeness
Use of modern sequencing approaches
PacBio HiFi + Hi-C phasing
Thorough methodological documentation
Clear protocols and software versions
Data availability
Open access raw data and assembly

Major Concerns
Lack of Biological Insight or Interpretation
Issue:
The manuscript provides almost no biological interpretation of the genome.
Why this matters:
Even for data notes, minimal discussion of genome features improves scientific value and reuse.
Required actions:
Add a short section addressing:

Genome size comparison with related Noctuidae species
Repeat content interpretation (43.76% reported but not discussed)
Any notable features (e.g., chromosome count consistency with Lepidoptera norms)
Provide at least 1–2 comparative references to contextualise results

Incomplete Reporting of Assembly Quality Metrics
Issue:
Some key metrics are insufficiently explained or missing:

No explicit contamination statistics reported despite ASCC use
No detailed BUSCO lineage justification
No structural accuracy validation beyond Hi-C visual inspection

Required actions:

Add:
Quantitative contamination results (e.g., % removed, taxonomic origin)
Clarification of why lepidoptera_odb10 was selected
Statement on structural accuracy validation (e.g., misjoins detected, correction rationale)

Limited Explanation of Haplotype Treatment
Issue:
Two haplotypes were generated, but the biological relevance and handling are unclear.
Required actions:

Clarify:

Whether haplotypes represent a phased diploid genome or alternative assemblies
Why is haplotype 2 only scaffold-level
Recommended version for downstream analysis

Provide clear guidance for users of the dataset

K-mer Completeness Interpretation is Confusing
Issue:

Individual haplotypes show ~70% completeness vs 99.35% combined
This is not explained

Required actions:

Explain:

Why haplotype-specific completeness is low
Whether this reflects heterozygosity (1.69%) or partitioning effects

Add a brief explanatory paragraph to prevent misinterpretation

Missing Annotation Information
Issue:
Genome annotation is deferred but not described.
Required actions:

Clearly state:
Expected timeline or status of annotation
Whether RNA-seq data exists
Where annotation will be hosted (Ensembl mentioned but not elaborated)

Minor Comments

Taxonomic Inconsistency

The species authority differs: “(Hübner, 1809)” vs “(Hübner, 1817)”
Fix: Ensure consistency throughout

Typographical / Formatting Issues

Spacing issues (e.g., “fromOrkney”, “northWales”)
Figure captions sometimes difficult to parse
Fix: Proofread carefully

Figures

Figures are referenced but not fully interpreted
Fix: Add brief interpretation (e.g., Hi-C confirms chromosomal structure)

Sampling Context

Only one individual was analysed
Fix: Add statement acknowledging limitations (e.g., population variation not captured)

Methods Clarity

Some protocols referenced externally only
Fix: Summarise key steps inline (without requiring external lookup)

Aspects Cannot Be Fully Assessed

Biological accuracy of taxonomy and species identification beyond the provided description
Raw data integrity (FASTQ/BAM not evaluated)
Performance of assembly under alternative pipelines
Functional annotation quality (not included)

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

Partly

Are the protocols appropriate and is the work technically sound?
Partly

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Medicinal chemistry

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 02 June 2026

https://doi.org/10.21956/wellcomeopenres.29280.r155911

Gulab D Khedkar
Dr Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

Review report on the manuscript entitled, “The genome sequence of the Suspected, Parastichtis suspecta (Hübner, 1809) (Lepidoptera: Noctuidae)”
This manuscript presents a high-quality chromosome-level genome assembly of *Parastichtis suspecta*, an ecologically important noctuid moth species and provides a valuable genomic resource for Lepidopteran biology, comparative genomics, biodiversity assessment, and evolutionary studies. The manuscript is well organized, technically rigorous, and clearly written. The authors successfully generated a highly contiguous and well-curated assembly using PacBio HiFi long-read sequencing combined with Hi-C scaffolding. The assembly statistics are excellent, with 99.8% of the assembly assigned to 30 chromosomal pseudomolecules, including the Z chromosome. The reported scaffold N50 (29.41 Mb), high consensus QV values (~69), and BUSCO completeness of 99.0% clearly demonstrate the robustness and reliability of the assembly. The assembly also meets the Earth BioGenome Project reference standards, highlighting the quality of the genomic resource.
A major strength of the manuscript is the comprehensive methodological transparency. The workflows for specimen collection, DNA barcoding verification, nucleic acid extraction, genome assembly, curation, and quality assessment are described in sufficient detail to ensure reproducibility. The inclusion of openly accessible accession numbers, software versions, and

analysis pipelines further enhances the utility of the work for the broader scientific community.
The background section provides useful ecological and taxonomic context for *P. suspecta*, including its distribution, habitat associations, and biological characteristics. The authors appropriately emphasize the importance of generating genomic resources for biodiversity documentation and future conservation-oriented studies.
The manuscript is scientifically sound and technically complete. The discussion could briefly elaborate on how this genome may contribute to future comparative analyses within Noctuidae and broader Lepidopteran phylogenomics. It is a minor and do not detract from the overall scientific quality of the work.
Overall, the manuscript presents an important and high-quality genomic resource generated using state-of-the-art methodologies. The study will be valuable to researchers working in Lepidoptera genomics, evolutionary biology, biodiversity science, and conservation genomics. I recommend the manuscript for acceptance after minor revision.

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: Genomics, biodiversity research

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.