



DATA NOTE

The genome sequence of the Garden Tiger, *Arctia caja* (Linnaeus, 1758) (Lepidoptera: Erebidae)

[version 1; peer review: 3 approved]

Yannick Chittaro ¹, Kay Lucek ², Charlotte J. Wright ³, Joana I. Meier³,
Mark L. Blaxter ³,
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, Project Psyche Community

¹Info fauna, Neuchâtel, Switzerland

²University of Neuchâtel, Neuchâtel, Switzerland

³Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 30 Jul 2025, 10:385
<https://doi.org/10.12688/wellcomeopenres.24634.1>
Latest published: 30 Jul 2025, 10:385
<https://doi.org/10.12688/wellcomeopenres.24634.1>

Abstract

We present a genome assembly from a male specimen of *Arctia caja* (Garden Tiger; Arthropoda; Insecta; Lepidoptera; Erebidae). The assembly contains two haplotypes with total lengths of 700.59 megabases and 699.20 megabases. Most of haplotype 1 (99.63%) is scaffolded into 31 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 15.41 kilobases.

Keywords

Arctia caja, Garden Tiger, genome sequence, chromosomal, Lepidoptera



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

Open Peer Review

Approval Status   

	1	2	3
version 1			
30 Jul 2025	view	view	view

1. Will Nash , The University of East Anglia, Norwich, UK
2. Sivasankaran Kuppusamy , Loyola College, Chennai, India
3. Sara Goodwin, Cold Spring Harbor Laboratory, New York, USA

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

Corresponding author: Charlotte J. Wright (cw22@sanger.ac.uk)

Author roles: **Chittaro Y:** Writing – Original Draft Preparation; **Lucek K:** Resources, Supervision, Writing – Review & Editing; **Wright CJ:** Funding Acquisition, Project Administration, Supervision; **Meier JI:** Funding Acquisition, Project Administration, Supervision; **Blaxter ML:** Funding Acquisition, Project Administration, Supervision;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). KL was supported by Swiss National Science Foundation (SNSF), Grant ID 202869.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Chittaro Y *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: Chittaro Y, Lucek K, Wright CJ *et al.* **The genome sequence of the Garden Tiger, *Arctia caja* (Linnaeus, 1758) (Lepidoptera: Erebidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, 10:385 <https://doi.org/10.12688/wellcomeopenres.24634.1>

First published: 30 Jul 2025, 10:385 <https://doi.org/10.12688/wellcomeopenres.24634.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Amphiesmenoptera; Lepidoptera; Glossata; Neolepidoptera; Heteroneura; Ditrysia; Obtectomera; Noctuoidea; Erebidae; Arctiinae; Arctiini; *Arctia*; *Arctia caja* (Linnaeus, 1758) (NCBI:txid289281)

Background

The Garden Tiger moth or Great Tiger moth (*Arctia caja*) is a large (up to 65 mm wingspan), conspicuously coloured moth, with cryptic forewings and hind wings with warning colouration. The upper surface of the forewings is brown, intersected by an irregular network of cream-coloured bands of varying width and extent, while the hindwings are orange (very rarely yellow) and adorned with large, rounded bluish spots. The moth's sequestration of toxic alkaloids (Aplin & Rothschild, 1972; Petzel-Witt *et al.*, 2023; Von Nickisch-Roseneck & Wink, 1993), together with its conspicuous coloration, suggests that it is aposematic. Males have weakly pectinate antennae, while females have filiform antennae.

This species has a Holarctic distribution. Its general range extends from Spain through western and central Europe to eastern Asia and Japan. In the North, it extends as far as Scandinavia, and in the south, across southern Italy and the Peloponnese to Asia Minor and the Himalayas (Ebert, 1997; GBIF Secretariat, 2023). In North America, it colonises the boreal belt of North America and the cordilleras of western North America (Dubatolov & Philip, 2013; GBIF Secretariat, 2023). Within this wide range, several subspecies are known, but also some very similar-looking species (especially *A. martinahoneyi*, *A. opulenta*, *A. brachyptera*) which are genetically only weakly differentiated based on few genetic markers (Rönkä *et al.*, 2016), and whose different taxonomic statuses are not all sufficiently resolved.

The species colonises a wide variety of environments, from open forests, clearings, edges and hems, roadsides, extensive dry or wet grasslands, and sometimes even gardens, from the plains to over 2,300 m (LSPN, 2005). This species prefers seasonal and temperate environments. Eggs are laid in plates under the leaves of the caterpillar's host plants. The caterpillar is extremely polyphagous, feeding on a wide range of low growing plants (such as *Rumex*, *Rubus* and *Urtica*) and shrubs (*Lonicera*, *Salix*, etc.) (Ebert, 1997). The caterpillar hibernates in the ground vegetation while it is still small, i.e. less than 10 mm (LSPN, 2005). Pupation takes place in a web woven on the ground. The species is univoltine, with most moths observed from mid-June to mid-August. The imagoes fly at night. The species is still widespread and sometimes common, but has declined in several regions, particularly in the UK (Conrad *et al.*, 2002). Its decline has been accompanied by phenotypic changes and reduced genetic diversity (Anderson *et al.*, 2008).

The reference genome of *Arctia caja* presented here provides the opportunity to resolve the remaining taxonomic uncertainties in the *Arctia caja* group (Joshi *et al.*, 2024; Rönkä *et al.*, 2016). The sequence data was derived from a male specimen (Figure 1) collected from Unterschächen, Switzerland.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult male *Arctia caja* (specimen ID SAN28000109, ToLID ilArcCaja1; Figure 1), collected from Unterschächen, Switzerland (latitude 46.8641, longitude 8.7826; elevation 1 053 m) on 11/08/2023.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The ilArcCaja1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the thorax was homogenised by powermashing using a PowerMasher II tissue disruptor.

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). 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



Figure 1. Voucher photograph of the *Arctia caja* (ilArcCaja1) specimen used for genome sequencing.

DNA had a Qubit concentration of 59.35 ng/μL and a yield of 2 789.45 ng, with a fragment size of 13.6 kb. The 260/280 spectrophotometric ratio was 1.96, and the 260/230 ratio was 2.75.

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.

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 25M 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.

Specimen details, sequencing platforms, and data yields are summarised in [Table 1](#).

Table 1. Specimen and sequencing data for BioProject PRJEB78788.

Platform	PacBio HiFi	Hi-C
ToLiD	ilArcCaja1	ilArcCaja1
Specimen ID	SAN28000109	SAN28000109
BioSample (source individual)	SAMEA115109945	SAMEA115109945
BioSample (tissue)	SAMEA115109988	SAMEA115109987
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485740	ERR13493997
Read count total	2.46 million	898.07 million
Base count total	24.88 Gb	135.61 Gb

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head tissue of the ilArcCaja1 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 Diagenode 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 and 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.

Specimen details, sequencing platforms, and data yields are summarised in [Table 1](#).

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](#); [Cheng et al., 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 MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cointons 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 18 breaks and 58 joins. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

Chromosomal painting was performed using lep_buscoPainter using Merian elements, which represent the 32 ancestral linkage groups in Lepidoptera (Wright *et al.*, 2024). Painting was based on gene locations from the lepidoptera_odb10 BUSCO analysis and chromosome lengths from the genome index produced using SAMtools faidx (Danecek *et al.*, 2021). Each complete BUSCO (including both single-copy and duplicated BUSCOs) was assigned to a Merian element using a reference database, and coloured positions were plotted along chromosomes drawn to scale.

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases (*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 BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO 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 package management via Conda and Bioconda (Grünig *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Arctia caja* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 24.88 Gb (gigabases) from 2.46 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 691.06 Mb, with a heterozygosity of 1.70% and repeat content of 29.18%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 35× coverage. Hi-C sequencing produced 135.61 Gb from 898.07 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

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 700.59 Mb in 108 scaffolds, with 150 gaps, and a scaffold N50 of 24.69 Mb (Table 2).

Assembly statistics

Most of the assembly sequence (99.63%) was assigned to 31 chromosomal-level scaffolds, representing 30 autosomes and the Z sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; Table 3). Chromosome painting with Merian elements illustrates

Table 2. Genome assembly statistics.

Assembly name	ilArcCaja1.hap1.1	ilArcCaja1.hap2.1
Assembly accession	GCA_964264495.1	GCA_964264445.1
Assembly level	chromosome	scaffold
Span (Mb)	700.59	699.20
Number of chromosomes	31	N/A
Number of contigs	258	199
Contig N50	6.9 Mb	6.59 Mb
Number of scaffolds	108	55
Scaffold N50	24.69 Mb	24.63 Mb
Longest scaffold length (Mb)	33.39	N/A
Sex chromosomes	Z	N/A
Organelles	Mitochondrial genome: 15.41 kb	N/A

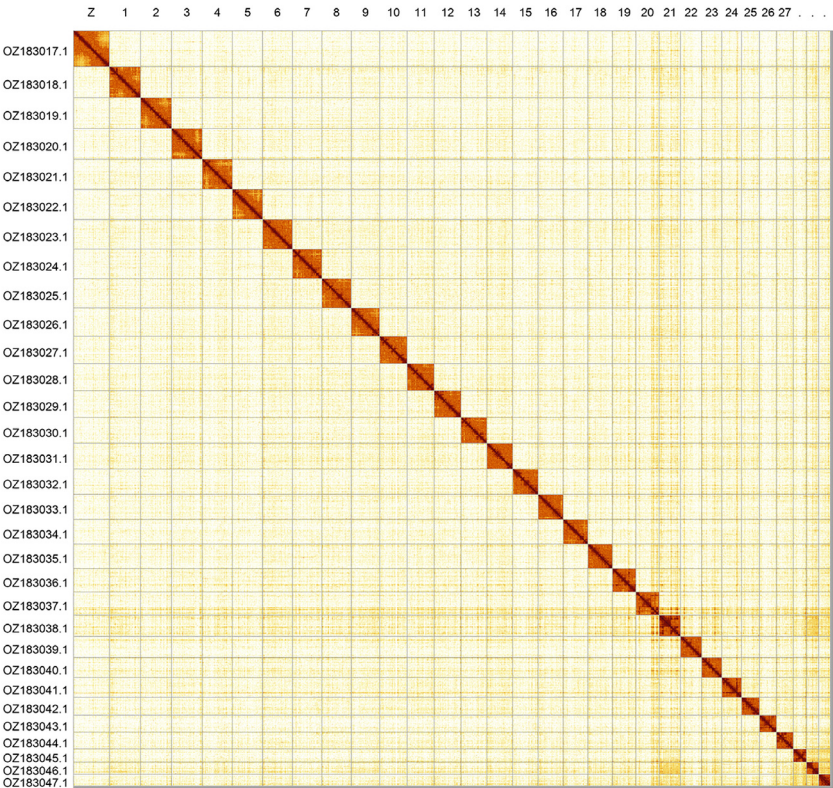


Figure 2. Hi-C contact map of the *Arctia caja* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Arctia caja* ilArcCaja1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ183018.1	1	28.72	36	M17;M20
OZ183019.1	2	28.58	36	M2
OZ183020.1	3	28.20	36	M9
OZ183021.1	4	27.94	36.50	M3
OZ183022.1	5	27.80	36	M1
OZ183023.1	6	27.39	36.50	M5
OZ183024.1	7	27.14	36	M8
OZ183025.1	8	27.13	36	M12
OZ183026.1	9	26.22	36	M7
OZ183027.1	10	25.28	36	M18
OZ183028.1	11	24.92	36	M16
OZ183029.1	12	24.69	36	M6
OZ183030.1	13	23.95	36	M21
OZ183031.1	14	23.90	36	M22
OZ183032.1	15	23.30	36	M15

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ183033.1	16	23.14	36	M10
OZ183034.1	17	22.83	36	M4
OZ183035.1	18	22.48	36.50	M11
OZ183036.1	19	21.62	36.50	M14
OZ183037.1	20	21.53	36	M23
OZ183038.1	21	19.77	38	M30
OZ183039.1	22	19.39	36	M13
OZ183040.1	23	18.64	36.50	M19
OZ183041.1	24	18.12	36.50	M24
OZ183042.1	25	16.54	36.50	M26
OZ183043.1	26	15.70	36.50	M28
OZ183044.1	27	15.42	37	M27
OZ183045.1	28	12.17	37.50	M25
OZ183046.1	29	11.34	38	M31
OZ183047.1	30	10.75	37.50	M29
OZ183017.1	Z	33.39	36	MZ
OZ183048.1	MT	0.02	19.50	N/A

the distribution of orthologues along chromosomes and highlights patterns of chromosomal evolution relative to Lepidopteran ancestral linkage groups (Figure 3).

The mitochondrial genome was also assembled. 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 66.7, and for haplotype 2, 66.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 66.6. The *k*-mer completeness

is 70.62% for haplotype 1, 70.53% for haplotype 2, and 99.62% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set ($n = 5\,286$) (Kriventseva *et al.*, 2019) identified 99.0% of the expected gene set (single = 98.2%, duplicated = 0.8%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length 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) the Earth BioGenome Project Report on

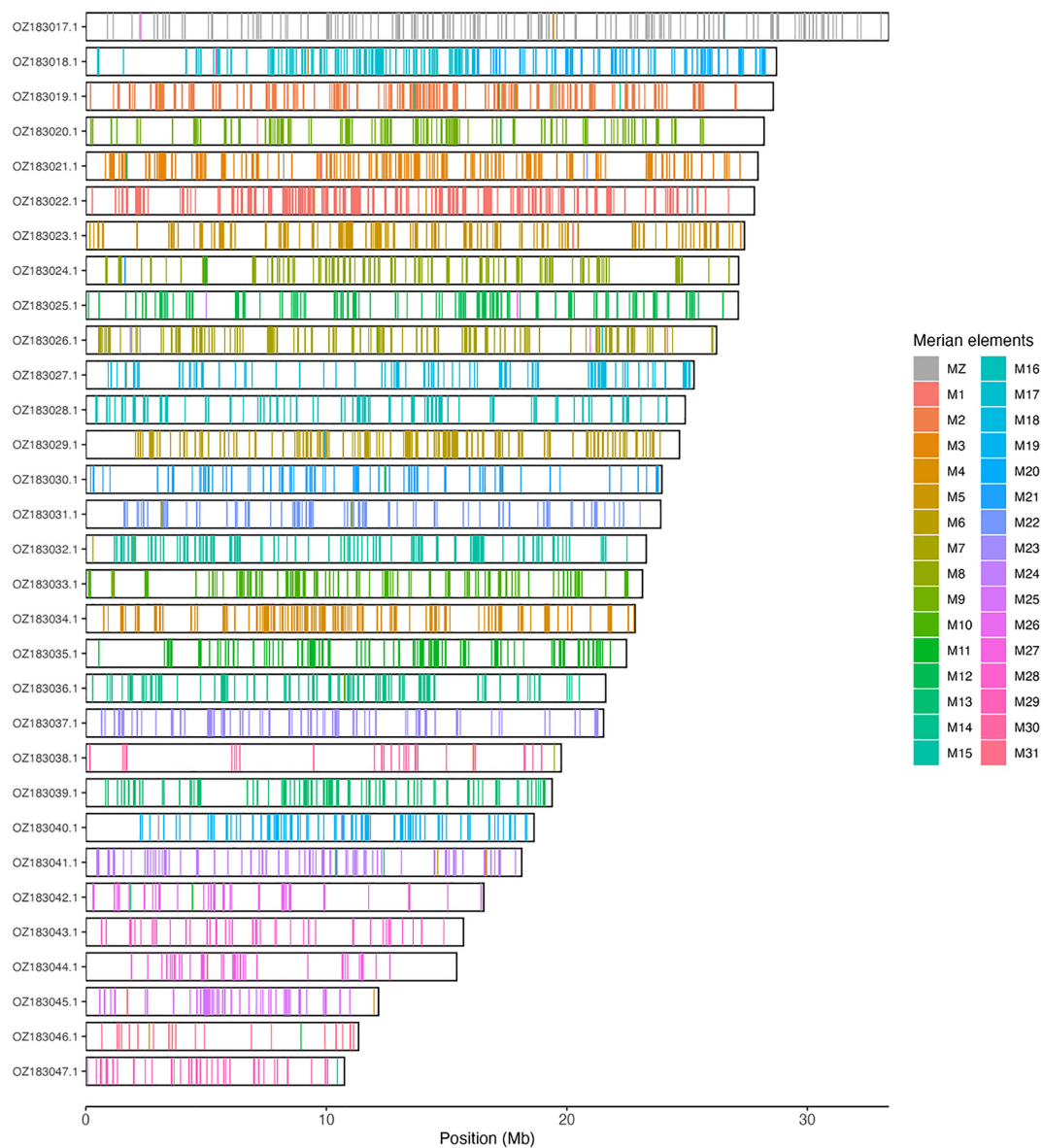


Figure 3. Merian elements painted across chromosomes in the iArcCaja1.hap1.1 assembly of *Arctia caja*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

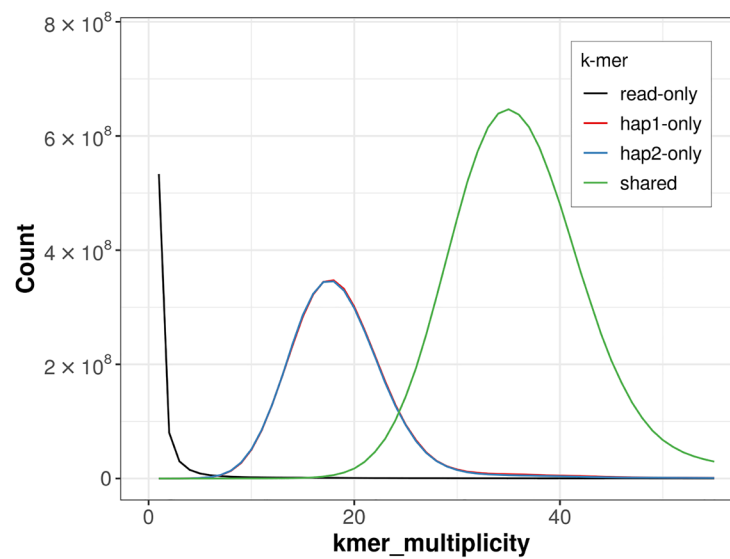


Figure 4. Evaluation of *k*-mer completeness using MerquyFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

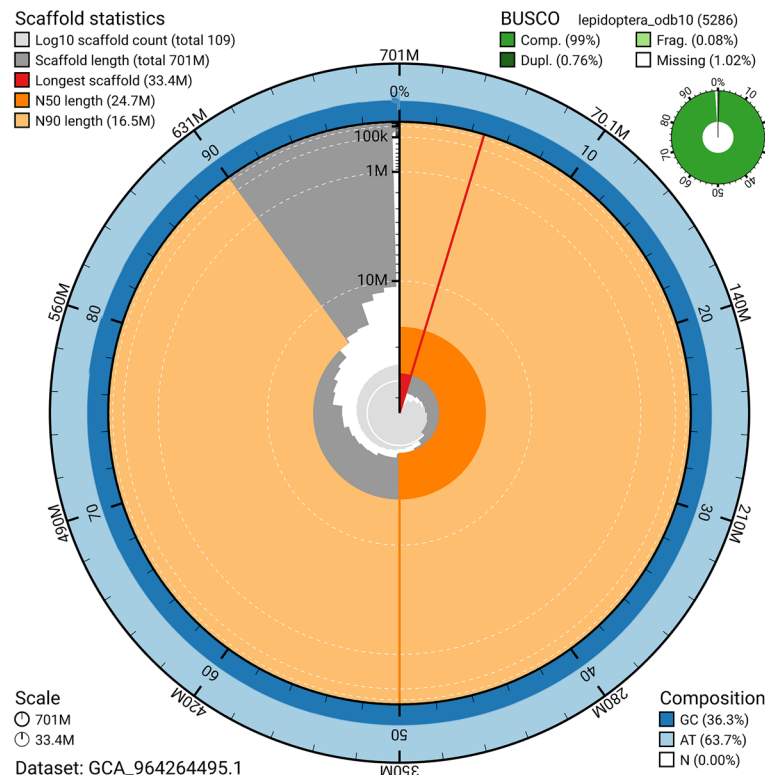


Figure 5. Assembly metrics for ilArcCaja1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

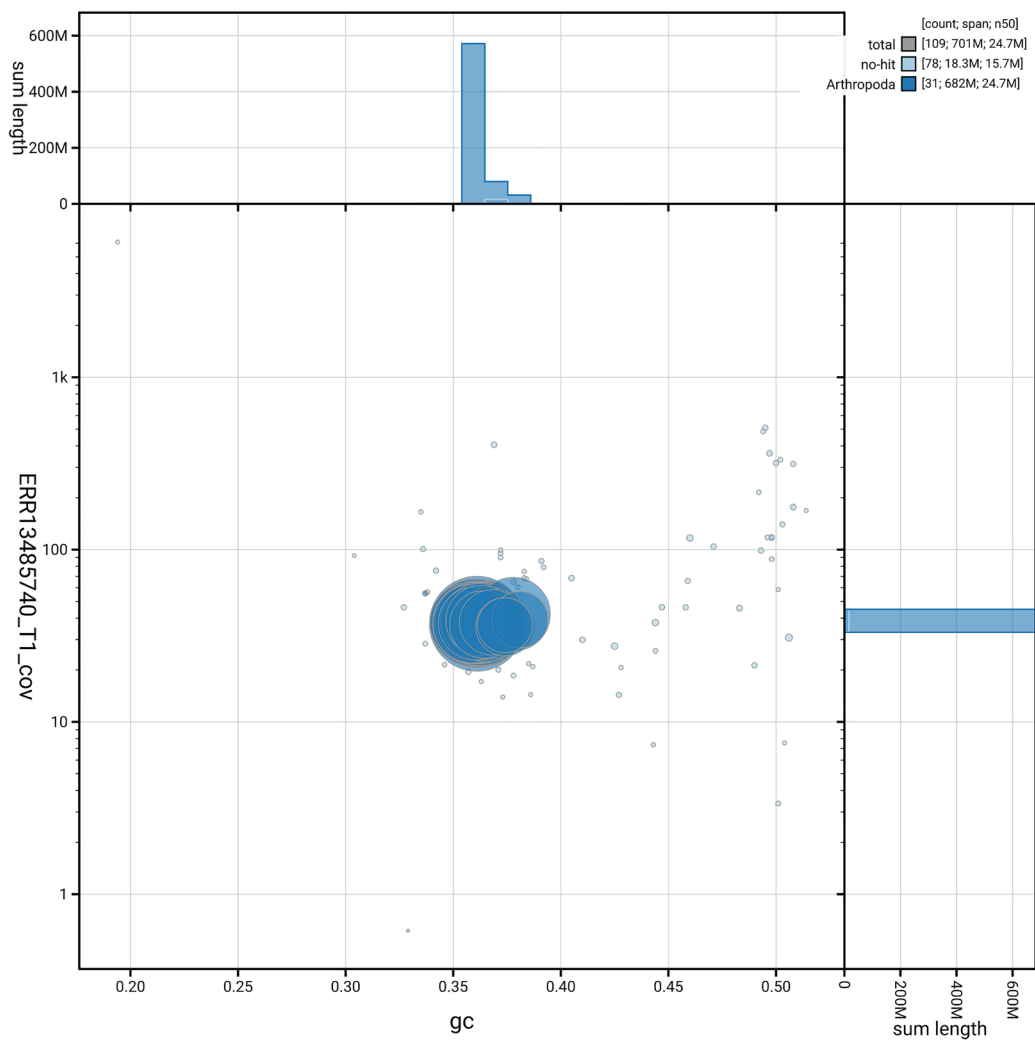


Figure 6. BlobToolKit GC-coverage plot for ilArcCaja1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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](#).

Table 4. Earth Biogenome Project summary metrics for the *Arctia caja* assembly.

Measure (Benchmark)	Value
EBP summary for haplotype 1	6.C.Q66
Contig N50 length (≥ 1 Mb)	6.90 Mb
Scaffold N50 length (= chromosome N50)	24.69 Mb
Consensus quality, QV (≥ 40)	Haplotype 1: 66.7; haplotype 2: 66.5; combined: 66.6
k-mer completeness (≥ 95%)	Haplotype 1: 70.62%; Haplotype 2: 70.53%; combined: 99.62%
BUSCO (S > 90%; D < 5%)	C:99.0%[S:98.2%,D:0.8%],F:0.1%,M:0.9%,n:5 286
Percentage of assembly assigned to chromosomes (≥ 90%)	99.63%

Assembly Standards [September 2024](#). The EBP metric, calculated for the haplotype 1, is **6.C.Q66**, meeting the recommended reference standard.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. 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 undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Arctia caja* (garden tiger moth). Accession number [PRJEB78788](#). The genome sequence is released openly for reuse. The *Arctia caja* genome sequencing initiative is part of the 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 [Ensembl](#) at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#) and [Table 2](#).

Pipelines used for genome assembly at the WSI Tree of Life are available at <https://pipelines.tol.sanger.ac.uk/pipelines>. [Table 5](#) lists software versions used in this study.

Contributors are listed at the following links:

- [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](#)
- [Project Psyche Community](#).

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
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
Cooler	0.8.11	https://github.com/open2c/cooler
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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
lep_buscoPainter	1.0.0	https://github.com/charlottewright/lep_buscoPainter
MerquryFK	1.1.1	https://github.com/thegenemyers/MERQURY.FK

Software	Version	Source
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
PretextViewSnapshot	N/A	https://github.com/sanger-tol/PretextViewSnapshot
PretextViewView	0.2.5	https://github.com/sanger-tol/PretextViewView
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
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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](#)

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410. [PubMed Abstract](#) | [Publisher Full Text](#)

Anderson SJ, Conrad KF, Gillman MP, *et al.*: **Phenotypic changes and reduced genetic diversity have accompanied the rapid decline of the garden tiger moth (*Arctia caja*) in the UK.** *Ecol Entomol.* 2008; **33**(5): 638–45. [Publisher Full Text](#)

Aplin RT, Rothschild M: **Poisonous alkaloids in the body tissues of the garden tiger moth (*Arctia caja* L.) and the cinnabar moth (*Tyria (=Callimorpha) jacobaeae* L.) (Lepidoptera).** In: A. De Vries and E. Kochva (eds), *Toxins of animal and plant origin*, Vol 2. New York: Gordon; Breach Sci Pub, 1972; 579–95. [Reference Source](#)

Bateman A, Martin MJ, 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 HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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](#)

Conrad KF, Woiwod IP, Perry JN: **Long-term decline in abundance and distribution of the garden tiger moth (*Arctia caja*) in Great Britain.** *Biol Conserv.* 2002; **106**(3): 329–37. [Publisher 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](#)

Dubatolov VV, Philip KW: **Review of the northern Holarctic *Arctia caja* complex (Lepidoptera: Noctuidae: Arctiinae).** *Can Entomol.* 2013; **145**(2): 147–54. [Publisher Full Text](#)

Ebert G: **Die Schmetterlinge Baden-Württembergs. Band 5: Nachtfalter III.** Verlag Eugen Ulmer, 1997. [Reference Source](#)

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.** *Nat Biotechnol.* 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](#)

GBIF Secretariat: ***Arctia caja* (Linnaeus, 1758) in GBIF Backbone Taxonomy.** 2023. [Publisher Full Text](#)

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476. [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): g1aa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Joshi M, Espeland M, Huemer P, *et al.*: **Species delimitation under allopatry: genomic insights within and across continents in Lepidoptera.** *Insect Syst Divers.* 2024; **8**(1): 1–16. [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](#)

Kriventseva EV, Kuznetsov D, Tegenfeldt F, *et al.*: **OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.** *Nucleic Acids Res.* 2019; **47**(D1): D807–D811.

[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](#)

LSPN: **Les Papillons Et Leurs Biotopes. Espèces. Dangers Qui Les Menacent. Protection. Volume 3.** L.S.P.N. Pro Natura, Bâle, 2005.

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](#)

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

[Reference Source](#)

Petzel-Witt S, Wunder C, Pogoda W, *et al.*: **Missed chances? Sequestration and non-sequestration of alkaloids by moths (Lepidoptera).** *Toxicon.* 2023; **227**: 107098.

[PubMed Abstract](#) | [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.*: **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](#)

Rönkä K, Mappes J, Kaila L, *et al.*: **Putting *Parasemia* in its phylogenetic place: a molecular analysis of the subtribe *Arctiina* (Lepidoptera).** *Syst Entomol.* 2016; **41**(4): 844–53.

[Publisher Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.

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

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.

[Publisher Full Text](#)

Von Nickisch-Rosenegk E, Wink M: **Sequestration of pyrrolizidine alkaloids in several *Arctiid* moths (Lepidoptera: Arctiidae).** *J Chem Ecol.* 1993; **19**(9): 1889–1903.

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

Wright CJ, Stevens L, Mackintosh A, *et al.*: **Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera.** *Nat Ecol Evol.* 2024; **8**(4): 777–790.

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

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 22 November 2025

<https://doi.org/10.21956/wellcomeopenres.27148.r134206>

© 2025 Goodwin S. 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.



Sara Goodwin

Cold Spring Harbor Laboratory, New York, USA

Review of "The genome sequence of the Garden Tiger, *Arctia caja* (Linnaeus, 1758) (Lepidoptera: Erebidae)
"

The Darwin Tree of Life Project is an exciting genomic endeavor. This latest manuscript reads as a typical article that is part of this project and is concise. This genome note reports a high-quality, chromosome-scale assembly for the Garden Tiger moth *Arctia caja*. Both PacBio HiFi Arima Hi-C were used to generate a phased assembly. This assembly achieved 99.0% complete genes for hap1 (BUSCO). Raw data and assemblies have been deposited at ENA under PRJEB78788

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: Plant and animal genomics and sequencing technology

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 24 October 2025

<https://doi.org/10.21956/wellcomeopenres.27148.r134207>

© 2025 Kuppusamy S. 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.



Sivasankaran Kuppusamy 

Loyola College, Chennai, Tamil Nadu, India

The authors have performed a chromosome-level genome assembly of the Garden Tiger, *Arctia caja* (Linnaeus, 1758). They generated 160.49 Gb of genomic data using PacBio and Illumina sequencing platforms. The assembly resulted in two haplotypes with total lengths of 700.59 Mb (chromosome level) and 699.20 Mb, respectively. High-molecular-weight DNA was isolated using appropriate methods, and suitable software tools were employed for genome assembly and annotation.

Comments on the manuscript

In the Background section, in the first sentence of the third line, the term “hind wings” appears as two separate words. It would be more appropriate to write it as “hindwings” (a single word) for consistency with standard entomological usage.

The authors have produced 160.49 Gb of robust genome data. However, the numbers of protein-coding genes, non-coding genes, and gene transcripts have not been identified through genome annotation and are not listed in the table. Could the authors clarify the reason for this omission?

The authors have used the full form of the genus name *Arctia* throughout the manuscript. The genus name should be written in full only at its first mention and may thereafter be abbreviated (e.g., *A. caja*).

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

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Phylogenetic analysis of Noctuoidea moths using mitogenome analysis

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 30 September 2025

<https://doi.org/10.21956/wellcomeopenres.27148.r134205>

© 2025 Nash W. 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.



Will Nash 

The University of East Anglia, Norwich, Norwich, UK

A chromosome level assembly for the Lepidopteran *Artica caja* is presented. This assembly was presumably sequenced as part of Project Psyche? This is not explicitly stated in the Background and could be clarified?

The assembly uses cutting edge methodologies. These are well documented and reproducible. PacBio sequencing is produced from the Revio platform, and this is scaffolded with Hi-C data. RNA-seq data is generated for annotation. Bioinformatic pipelines are well documented, and versions are provided.

The assembly is 700.59 Mb over 108 scaffolds with an N50 of 24.69Mb. 99.63% of the assembly is assigned into 31 chromosome level scaffolds. The assembly is high quality: QV 66.6, 99.62 kmer completeness, and representing 99.0% of the lepidoptera_odb10 BUSCO set.

Assemblies and raw data is available through ENA.

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: Population Genomics, 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.
