DATA NOTE

The genome sequence of the Red-underwing Skipper, *Spialia sertorius* (Hoffmannsegg, 1804) (Lepidoptera: Hesperiidae)

[version 1; peer review: 3 approved]

Kay Lucek[1], Daniel Linke[2], Charlotte J. Wright[3], Joana I. Meier[3], 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

[1]University of Neuchâtel, Neuchâtel, Switzerland
[2]Biology Centre of the Academy of Sciences of the Czech Republic, Senckenberg German Entomological Institute, South Bohemian University in Ceske Budjeovice, Ceske Budjeovice, Czech Republic
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 30 Jul 2025, **10**:390
https://doi.org/10.12688/wellcomeopenres.24648.1
Latest published: 30 Jul 2025, **10**:390
https://doi.org/10.12688/wellcomeopenres.24648.1

Abstract

We present a genome assembly from a female specimen of *Spialia sertorius* (Red-underwing Skipper; Arthropoda; Insecta; Lepidoptera; Hesperiidae). The assembly contains two haplotypes with total lengths of 364.57 megabases and 323.91 megabases. Most of haplotype 1 (99.92%) is scaffolded into 32 chromosomal pseudomolecules, including the W and Z sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.34 kilobases.

Keywords

Spialia sertorius, Red-underwing Skipper, 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. **Elena Pazhenkova**, University of Ljubljana, Ljubljana, Slovenia
2. **Panagiotis Ioannidis**, Foundation for Research & Technology - Hellas, Crete, Greece
3. **Bin Zhang**, Qingdao Agricultural University, Qingdao, China

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: Lucek K: Investigation, Resources, Supervision; **Linke D**: Writing – Original Draft Preparation; **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.

How to cite this article: Lucek K, Linke D, Wright CJ *et al.* **The genome sequence of the Red-underwing Skipper, *Spialia sertorius* (Hoffmannsegg, 1804) (Lepidoptera: Hesperiidae) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, **10**:390 https://doi.org/10.12688/wellcomeopenres.24648.1

First published: 30 Jul 2025, **10**:390 https://doi.org/10.12688/wellcomeopenres.24648.1

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; Hesperioidea; Hesperiidae; Pyrginae; *Spialia*; *Spialia sertorius* (Hoffmannsegg, 1804) (NCBI:txid509483)

Background

Spialia sertorius, the Red-underwing Skipper, is a small butterfly species within the family Hesperiidae. It is widely distributed across Western and Central Europe, extending southward to North Africa and eastward to Poland, Czechia, Slovakia, and Bosnia and Herzegovina (De Jong, 1974; Lelo, 2007; Tolman & Lewington, 2009). However, it is absent from Scandinavia and Great Britain and is considered extinct in the Netherlands (Maes *et al.*, 2019). The species shares a long contact zone with *Spialia orbifer* across central Europe, i.e., Slovenia, Croatia and Austria (Fazekas, 1986; Hesselbarth & Van Oorschot, 1995; Lorkovic, 1973).

This species inhabits dry, calcareous grasslands, meadows, and sunny slopes, often favouring areas with sparse vegetation and favourable microclimates, linked to the specific requirements of its hostplant (*Sanguisorba magnolii*). It is usually only consistently found at elevations below 1 600 m, which corresponds to the upper limit of its hostplant (Hernández-Roldán *et al.*, 2018).

Spialia sertorius is distinguished by its rapid, low flight and characteristic reddish-brown hindwing undersides. However, species within this genus are hard to distinguish based on external morphology (de Jong, 1978; Hernández-Roldán *et al.*, 2016) and unknown cryptic species have recently been discovered in *S. sertorius* (Hernández-Roldán *et al.*, 2016). *S. sertorius* and *S. rosae* seem to be fully indistinguishable based on external morphology and co-occur in Spain at mid-elevations (Hernández-Roldán *et al.*, 2016; Hernández-Roldán *et al.*, 2018). The wingspan ranges from 22 to 26 mm. Adults are bivoltine, with flight periods from April to August, depending on altitude and local conditions (Bräu, 2013).

The primary larval host plant is *Sanguisorba minor* (Salad Burnet), although feeding on *Potentilla* spp. and *Rubus* spp. has been reported (Tolman & Lewington, 2009; Tshikolovets, 2011). Larvae feed on the plant's flowers and leaves, constructing leaf shelters for protection and pupation occurs on the ground within leaf litter (Bräu, 2013). *S. sertorius* is considered critically endangered in Hungary and endangered in Poland (Maes *et al.*, 2019), it is however classed by IUCN as least concern with stable populations (International Union for Conservation of Nature, 2025). The species is also included in the European Grassland Butterfly Indicator species list, which gives an uncertain population trend for *S. sertorius* (van Swaay *et al.*, 2022). Conservation of *S. sertorius* relies on maintaining open grassland habitats through appropriate grazing or mowing regimes to prevent habitat encroachment, succession and eutrophication.

We present a chromosome-level, haplotype-resolved genome sequence of the Red-underwing Skipper, *Spialia sertorius*, sequenced as part of Project Psyche. The sequence data were derived from a female specimen (Figure 1) collected from Dittingen BL, Switzerland.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult female *Spialia sertorius* (specimen ID SAN28000133, ToLID ilSpiSert1; Figure 1), collected from Dittingen BL, Switzerland (latitude 47.4431, longitude 7.4918; elevation 500 m) on 03/06/2023. The specimen was collected and identified by Yannick Chittaro (Info Fauna, Neuchâtel, Switzerland).

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 ilSpiSert1 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 DNA had a Qubit concentration of 22.77 ng/μL and a yield of 1 070.19 ng, with a fragment size of 15.4 kb. The 260/280 spectrophotometric ratio was 1.94, and the 260/230 ratio was 4.16.

Figure 1. Voucher photograph of the *Spialia sertorius* (ilSpiSert1) specimen used for genome sequencing.

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.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head tissue of the ilSpiSert1 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 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.

Table 1. Specimen and sequencing data for BioProject PRJEB78810.

Platform	PacBio HiFi	Hi-C
ToLID	ilSpiSert1	ilSpiSert1
Specimen ID	SAN28000133	SAN28000133
BioSample (source individual)	SAMEA115110007	SAMEA115110007
BioSample (tissue)	SAMEA115110035	SAMEA115110034
Tissue	thorax	head
Sequencing platform and model	Revio	Illumina NovaSeq X
Run accessions	ERR13485751	ERR13494008
Read count total	3.08 million	889.33 million
Base count total	32.72 Gb	134.29 Gb

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 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 21 breaks, 80 joins, and removal of 36 haplotypic duplications. The curation process is documented 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

Chromosomal painting was performed using lep_busco_painter 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 GoaT 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 (Grüning *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 *Spialia sertorius* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 32.72 Gb (gigabases) from 3.08 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 341.34 Mb, with a heterozygosity of 2.50% and repeat content of 25.59%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 93× coverage. Hi-C sequencing produced 134.29 Gb from 889.33 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 364.57 Mb in 46 scaffolds, with 96 gaps, and a scaffold N50 of 12.35 Mb (Table 2).

Most of the assembly sequence (99.92%) was assigned to 32 chromosomal-level scaffolds, representing 30 autosomes and the W and Z sex chromosomes. The sex chromosomes Z and W were identified by homology with *Thymelicus lineola* (GCA_963932265.1) (Lohse *et al.*, 2024). These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 2; Table 3). Chromosome painting with Merian elements illustrates 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.

Table 2. Genome assembly statistics.

Assembly name	ilSpiSert1.hap1.1	ilSpiSert1.hap2.1
Assembly accession	GCA_964258965.1	GCA_964258975.1
Assembly level	chromosome	scaffold
Span (Mb)	364.57	323.91
Number of chromosomes	32	N/A
Number of contigs	142	133
Contig N50	6.79 Mb	5.9 Mb
Number of scaffolds	46	38
Scaffold N50	12.35 Mb	12.12 Mb
Longest scaffold length (Mb)	24.67	N/A
Sex chromosomes	W and Z	N/A
Organelles	Mitochondrial genome: 15.34 kb	N/A

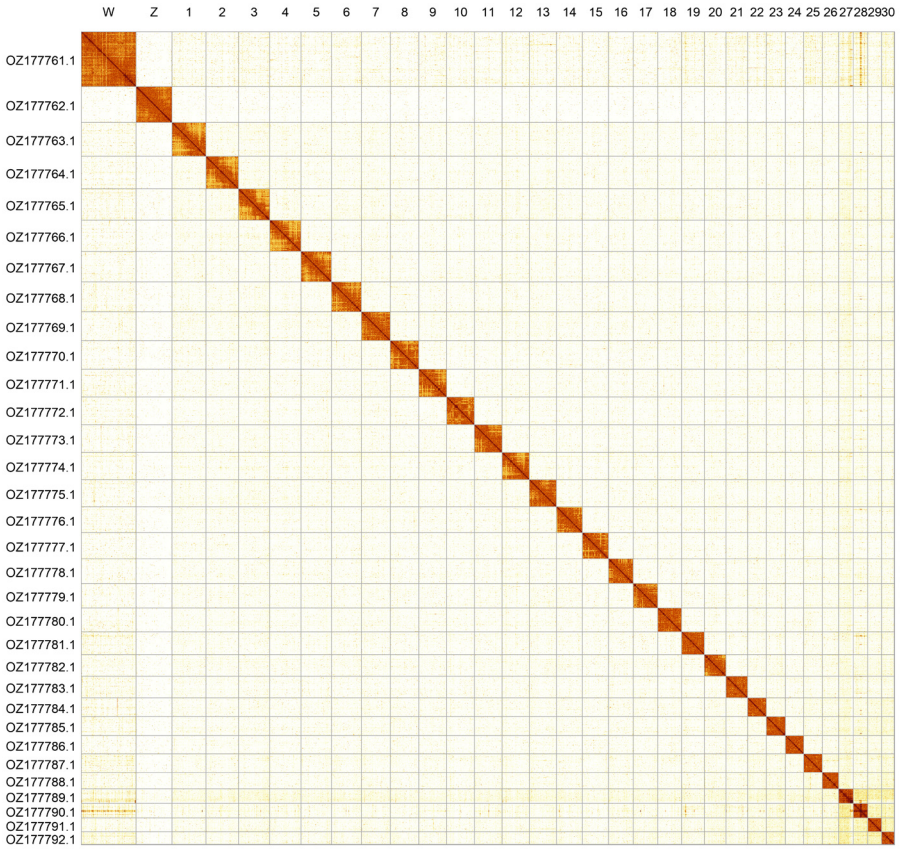

Figure 2. Hi-C contact map of the *Spialia sertorius* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Spialia sertorius* ilSpiSert1.

INSDC accession	Molecule	Length (Mb)	GC%	Assigned Merian elements
OZ177763.1	1	15.09	35.50	M14;M2
OZ177764.1	2	14.62	36	M1
OZ177765.1	3	14.04	36	M17;M20
OZ177766.1	4	14.01	36	M3
OZ177767.1	5	13.63	35.50	M9
OZ177768.1	6	13.41	36	M8
OZ177769.1	7	12.93	35.50	M5
OZ177770.1	8	12.73	35.50	M12
OZ177771.1	9	12.52	35.50	M16
OZ177772.1	10	12.42	35.50	M18
OZ177773.1	11	12.35	36	M7
OZ177774.1	12	12.24	36	M4
OZ177775.1	13	12.10	36	M6
OZ177776.1	14	11.71	36	M15
OZ177777.1	15	11.66	35.50	M21
OZ177778.1	16	11.01	36.50	M11
OZ177779.1	17	11.01	36	M10
OZ177780.1	18	10.70	36.50	M22
OZ177781.1	19	10.25	36	M13
OZ177782.1	20	9.68	36.50	M14
OZ177783.1	21	9.61	36	M23
OZ177784.1	22	8.48	36	M24
OZ177785.1	23	8.44	37	M19
OZ177786.1	24	8.28	36.50	M28
OZ177787.1	25	8.23	36	M26
OZ177788.1	26	7.40	36.50	M27
OZ177789.1	27	6.58	40.50	M31
OZ177790.1	28	6.50	40.50	M30
OZ177791.1	29	6.11	37	M25
OZ177792.1	30	5.77	38	M29
OZ177761.1	W	24.67	38	N/A
OZ177762.1	Z	16.14	35.50	MZ
OZ177793.1	MT	0.02	19	N/A

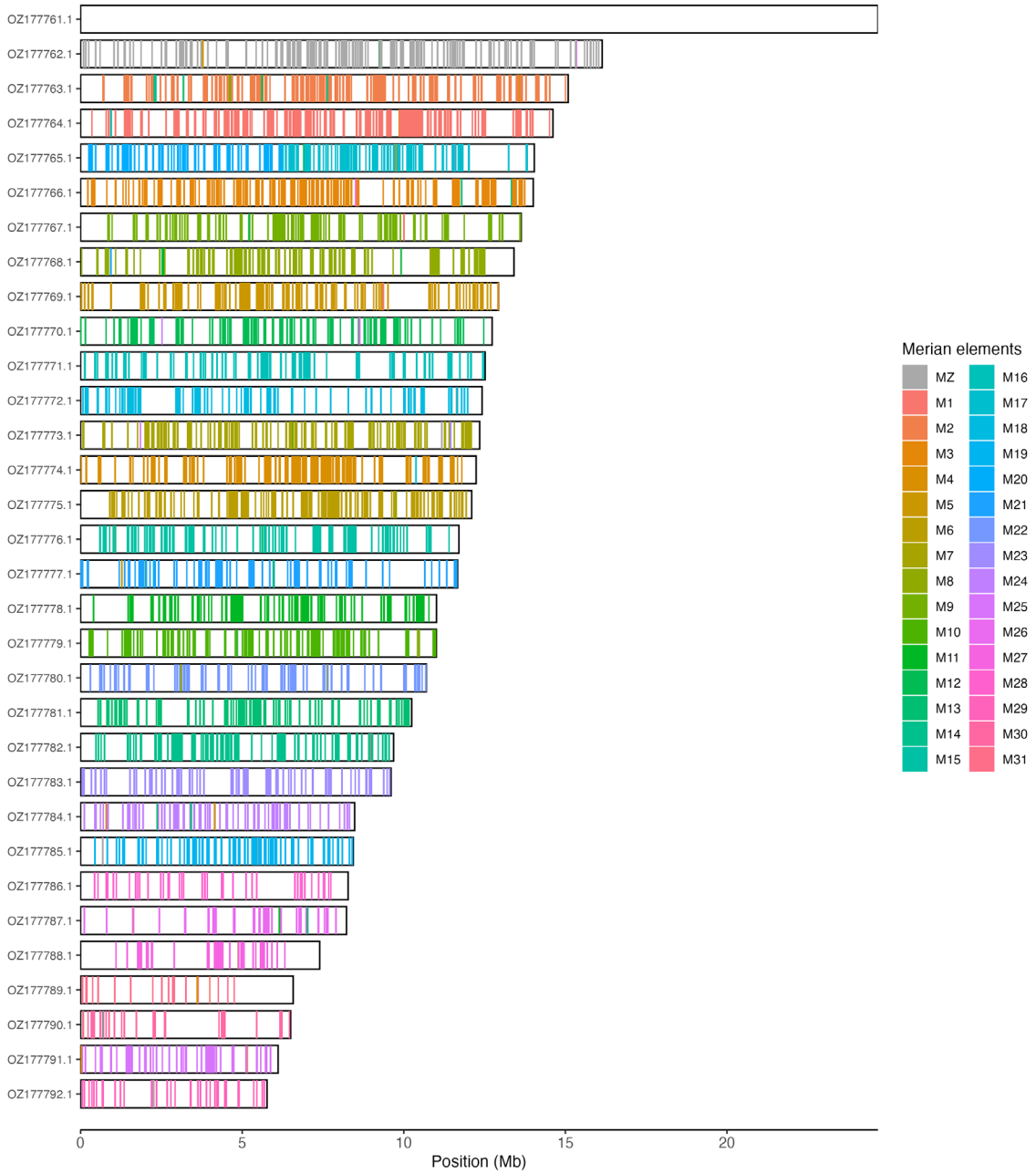

Figure 3. Merian elements painted across chromosomes in the ilSpiSert1.hap1.1 assembly of *Spialia sertorius*. Chromosomes are drawn to scale, with the positions of orthologues shown as coloured bars. Each orthologue is coloured by the Merian element that it belongs to. All orthologues which could be assigned to Merian elements are shown.

Assembly quality metrics

For haplotype 1, the estimated QV is 66.3, and for haplotype 2, 65.9. When the two haplotypes are combined, the assembly achieves an estimated QV of 66.1. The *k*-mer completeness is 67.42% for haplotype 1, 62.42% for haplotype 2, and 99.80% for the combined haplotypes (Figure 4). BUSCO analysis using the lepidoptera_odb10 reference set (n = 5 286) (Kriventseva *et al.*, 2019) identified 98.5% of the expected gene set (single = 98.3%, duplicated = 0.2%) 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 Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q66**, meeting the recommended reference standard.

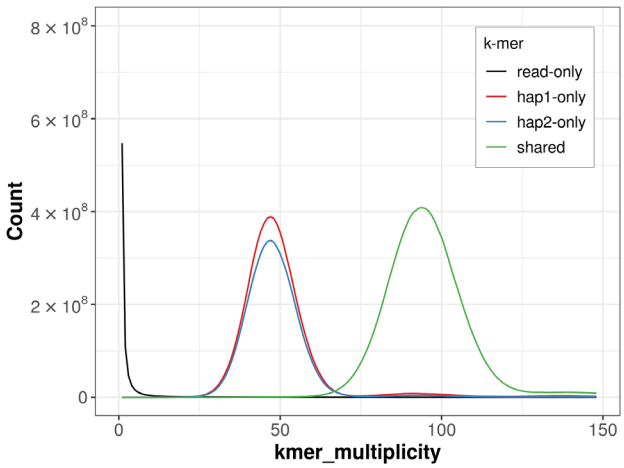

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve (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.

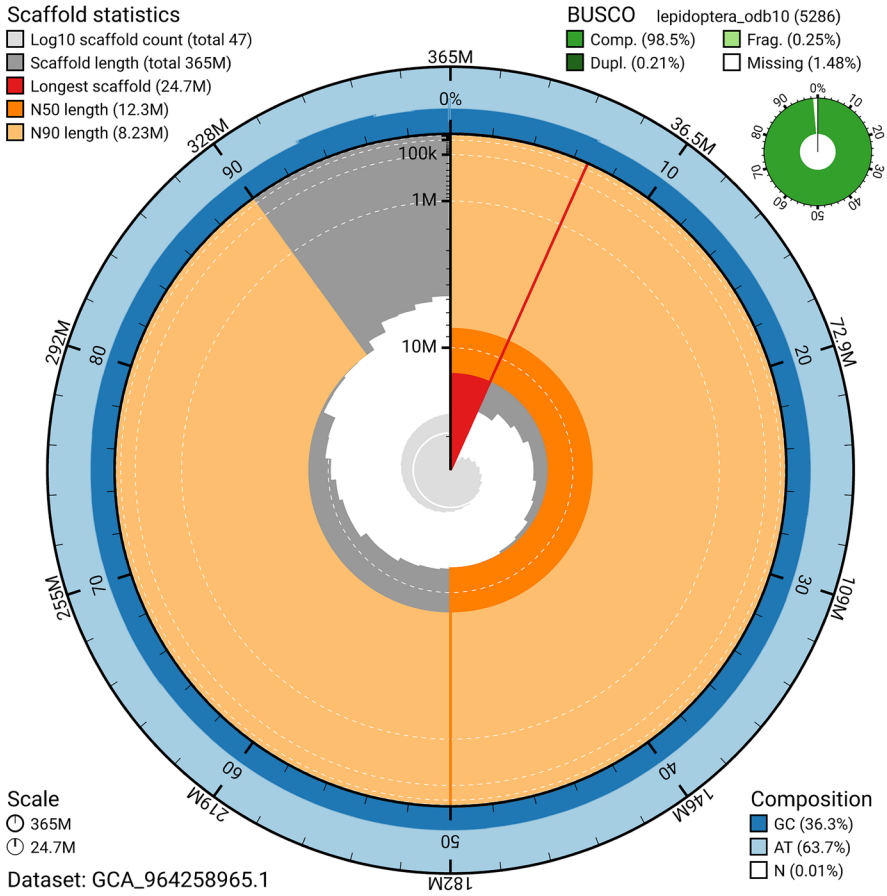

Figure 5. Assembly metrics for ilSpiSert1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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.

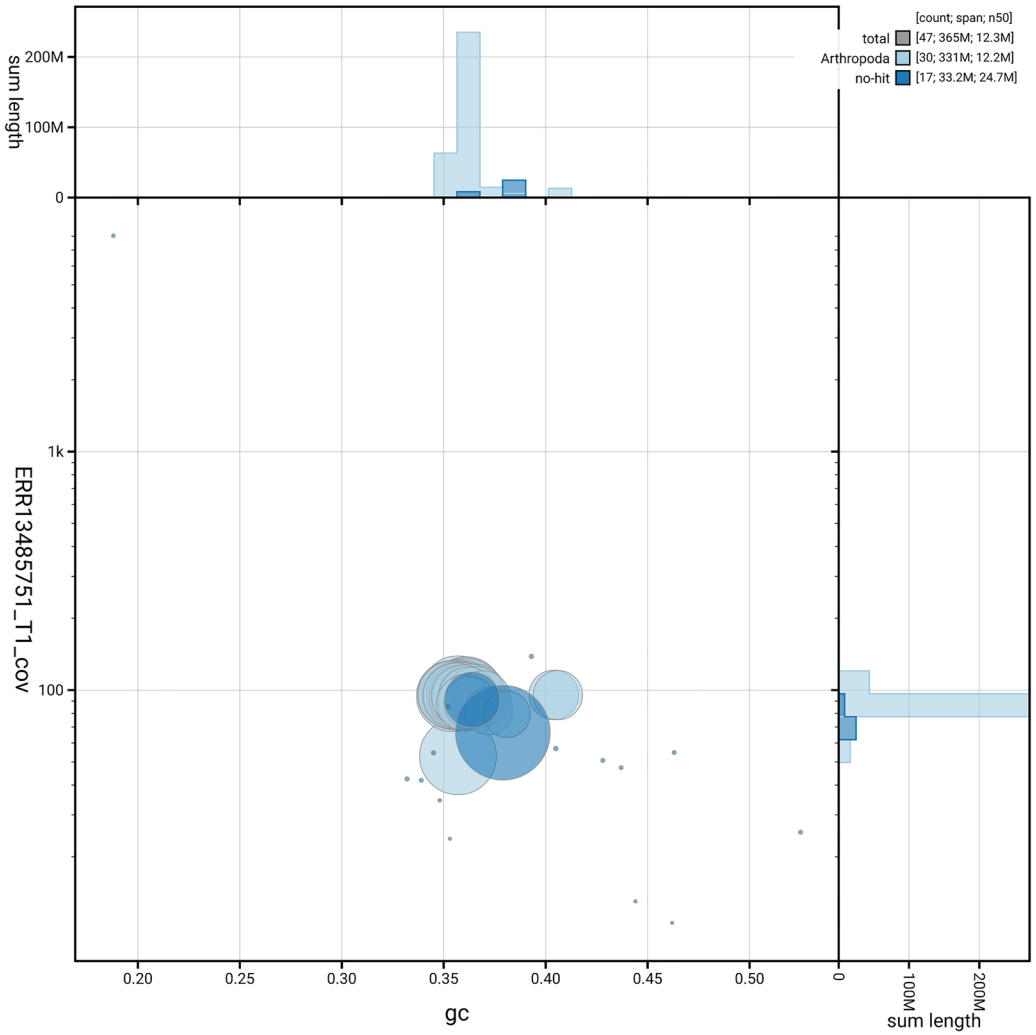

Figure 6. BlobToolKit GC-coverage plot for ilSpiSert1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 *Spialia sertorius* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q66	6.C.Q40
Contig N50 length	6.79 Mb	≥ 1 Mb
Scaffold N50 length	12.35 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 66.3; haplotype 2: 65.9; combined: 66.1	≥ 40
k-mer completeness	Haplotype 1: 67.42%; Haplotype 2: 62.42%; combined: 99.80%	≥ 95%
BUSCO	C:98.5%[S:98.3%,D:0.2%], F:0.2%,M:1.2%,n:5 286	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.92%	≥ 90%

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: Spialia sertorius (red underwing skipper). Accession number PRJEB78810. The genome sequence is released openly for reuse. The *Spialia sertorius* 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.

Author information

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_busco_painter	1.0.0	https://github.com/charlottewright/lep_busco_painter
MerquryFK	1.1.2	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
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	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
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

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

Bräu M, ed: **Tagfalter in bayern.** Stuttgart (Hohenheim): Ulmer, 2013.
Reference Source

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

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

De Jong R: **Systematics and evolution of the Palaearctic *Spialia* species (Lepidoptera, Hesperiidae).** *Tijdschr Entomol.* 1974; **117**(6): 225–271.
Reference Source

de Jong R: **Monograph of the genus *Spialia* Swinhoe (Lepidoptera, Hesperiidae).** *Tijdschr Entomol.* 1978; **121**: 23–146.
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

Fazekas I: **Die *Spialia*-Arten des Karpatenbeckens und ihre Verbreitung (Lepidoptera: Hesperiidae).** *Nachr Entomol Ver Apollo.* 1986; **7**(2/3): 49–55.
Reference Source

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

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

Hernández-Roldán JL, Dapporto L, Dincă V, *et al.*: **Integrative analyses unveil speciation linked to host plant shift in *Spialia* butterflies.** *Mol Ecol.* 2016; **25**(17): 4267–84.
PubMed Abstract | Publisher Full Text

Hernández-Roldán JL, Vicente JC, Vila R, *et al.*: **Natural history and immature stage morphology of *Spialia* Swinhoe, 1912 in the Iberian Peninsula (Lepidoptera, Hesperiidae).** *Nota Lepidopterol.* 2018; **41**(1): 1–22.
Publisher Full Text

Hesselbarth G, Van Oorschot H: **Die Tagfalter Der Türkei: Unter Berücksichtigung Der Angrenzenden Länder, Vol. 1.** S. Wagener, 1995.
Reference Source

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

International Union for Conservation of Nature: **The IUCN red list of threatened species.** Version 2025-1, 2025.
Reference Source

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

Lelo S: **Contribution to knowledge of the fauna of butterflies in Bosnia and Herzegovina.** *Acta entomologica serbica.* 2007; **12**(2): 73–92.
Reference Source

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

Lohse K, Vila R, Hayward A, *et al.*: **The genome sequence of the Essex Skipper butterfly, *Thymelicus lineola* (Ochsenheimer, 1808) [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 452.
PubMed Abstract | Publisher Full Text | Free Full Text

Lorkovic Z: **150 jahre bis zur entdeckung der präimaginalstadien von Spialia orbifer hbn.(Lep., Hesperiidae).** *Acta Entomologica Yugoslavica.* 1973; **9**: 67–70.

Maes D, Verovnik R, Wiemers M, *et al.*: **Integrating national Red Lists for prioritising conservation actions for European butterflies.** *J Insect Conserv.* 2019; **23**(2): 301–30.
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): 4647–4654.
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, [Accessed 2 April 2024].
Reference Source

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

Tolman T, Lewington R: **Collins butterfly guide.** HarperCollins Publishers, 2009.
Reference Source

Tshikolovets VV: **Butterflies of Europe & the mediterranean area.** 2011.
Reference Source

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

Van Swaay CA, Dennis EB, Schmucki R, *et al.*: **European Grassland Butterfly Indicator 1990-2020: technical report.** 2022.
Reference Source

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

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 24 October 2025

https://doi.org/10.21956/wellcomeopenres.27162.r135178

✔ **Bin Zhang**
Qingdao Agricultural University, Qingdao, China

This data note presents a high-quality, chromosome-level genome assembly of *Spialia sertorius* (Red-underwing Skipper), a species of ecological and conservation interest in the family Hesperiidae. The assembly meets the Earth BioGenome Project (EBP) reference standards (6.C.Q40) and provides detailed sequencing, assembly, and curation workflows. By integrating PacBio HiFi long reads and Hi-C data, the authors successfully generated a haplotype-resolved assembly, with 99.92% of Haplotype 1 assigned to chromosomal pseudomolecules (including W/Z sex chromosomes). This work fills a critical gap in genomic resources for Hesperiidae and supports future research on butterfly phylogenomics, population genetics, and conservation. However, several minor issues related to sample representativeness, genomic detail, and functional context need to be addressed to maximize the manuscript's utility.

1. While the mitochondrial genome length (15.34 kb) is reported, key features are missing: (1) gene order (e.g., arrangement of protein-coding genes, rRNA, tRNA); (2) GC content of mitochondrial regions (e.g., control region vs. coding regions); (3) comparisons with mitochondrial genomes of congeneric (e.g., *S. orbifer*) or confamilial species. This limits its application in evolutionary studies (e.g., mitochondrial phylogenetics of Hesperiidae).
2. The manuscript states the genome "will be annotated using available RNA-Seq data" but provides no preliminary results (e.g., number of predicted protein-coding genes, functional categories of BUSCO genes) or timelines for annotation release. Without even basic functional insights, it is impossible to link the assembly to *S. sertorius*' unique characters (e.g., genes related to hostplant specialization on *Sanguisorba minor* or rapid flight behavior).
3. While the authors reference *Thymelicus lineola* (GCA_963932265.1) for W/Z sex chromosome identification, no detailed comparison with *S. sertorius* is provided. Key evolutionary insights (e.g., synteny conservation, chromosome fusion/fission events) are not discussed, missed opportunities to contextualize the assembly within Lepidoptera genome evolution.

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 and Bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 24 October 2025

https://doi.org/10.21956/wellcomeopenres.27162.r135182

Panagiotis Ioannidis
Foundation for Research & Technology - Hellas, Crete, Greece

This manuscript describes the sequencing and assembly of the lepidopteran insect *Spialia sertorius* . All the methodology used for assembling is the standard one used by other Psyche/DToL genome projects and as such the resulting genome assembly is suitable for downstream analyses. This is demonstrated by the relatively low contig/scaffold to chromosome ratio (46/32=1.43), as well as the good Merqury QV and BUSCO scores.

I should note here that the lack of a predicted gene set precludes the use of this assembly from a large number of researchers, because gene prediction is not trivial thing to do. Therefore, it would certainly be desirable to also have a predicted gene set (together with its BUSCO score!).

I find it very interesting that chromosomes 27 and 28 have a considerably different GC% from the remaining chromosomes. Was this expected? Maybe the authors should briefly comment on it. And I think that such findings should be brought up in the text. I understand that this is a genome report, but in the same way that the authors (very appropriately) write details about the ecology and the life cycle of the insect, in that same sense they should also bring up "peculiar" findings from the analysis of their genome assembly. One sentence should suffice!

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

Are the protocols appropriate and is the work technically sound?

Yes

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: insect genomics; bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 29 August 2025

https://doi.org/10.21956/wellcomeopenres.27162.r128398

Elena Pazhenkova
University of Ljubljana, Ljubljana, Slovenia

The article presents a comprehensive genome assembly of the Red-underwing Skipper butterfly, *Spialia sertorius* (Lepidoptera, Hesperiidae). The authors have successfully constructed a high-quality assembly from a female specimen by utilizing Pacific Biosciences HiFi long-read sequencing and Hi-C data for scaffolding and phasing. The resulting assembly comprises two haplotypes, with total lengths of 364.57 megabases and 323.91 megabases. Notably, 99.92% of haplotype 1 is scaffolded into 32 chromosomal pseudomolecules, including both the W and Z sex chromosomes, while haplotype 2 was assembled to scaffold level. The mitochondrial genome, spanning 15.34 kilobases, has also been assembled.

An estimated haploid genome size was 341.34 megabases, with a heterozygosity of 2.50% and a repeat content of 25.59%, providing a strong foundation for assembly expectations. High sequencing coverage for the nuclear genome, supplemented by extensive Hi-C data that facilitated chromosome-level assembly of haplotype, which was manually curated. The chromosomal-level assembly represents a valuable genomic resource. It lays the groundwork for future research into the population genomics and evolutionary history of *S. sertorius*, especially within the context of its broad distribution and areas of sympatry with related species. The combination of extensive sequencing coverage, high contiguity, and careful curation makes this work a significant contribution to the field of butterfly genomics and biodiversity research.

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.

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.