



DATA NOTE

The genome sequence of the furrow orbweaver, *Larinioides cornutus* (Clerck, 1757)

[version 1; peer review: 3 approved]

Mark G. Telfer¹, Liam M. Crowley^{id}², Craig S. Wilding^{id}³,
 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

¹Independent researcher, Ventnor, Isle of Wight, England, UK²Department of Biology, University of Oxford, Oxford, England, UK³School of Biology and Environmental Science, University College Dublin, Dublin, Leinster, Ireland

V1 First published: 23 May 2025, 10:267
<https://doi.org/10.12688/wellcomeopenres.24288.1>
 Latest published: 23 May 2025, 10:267
<https://doi.org/10.12688/wellcomeopenres.24288.1>

Abstract

We present a genome assembly from a female specimen of *Larinioides cornutus* (furrow orbweaver; Arthropoda; Arachnida; Araneae; Araneidae). The assembly contains two haplotypes with total lengths of 2,304.58 megabases and 2,295.32 megabases. Most of haplotype 1 (98.49%) is scaffolded into 13 chromosomal pseudomolecules, including the X₁ and X₂ sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 14.6 kilobases.

Keywords

Larinioides cornutus, furrow orbweaver, genome sequence, chromosomal, Araneae



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

Open Peer Review

Approval Status

	1	2	3
version 1 23 May 2025	 view	 view	 view

- Peng-Yu Jin** , University of the Chinese Academy of Sciences, Beijing, China
Institute of Zoology, Chinese Academy of Sciences, Beijing, China
- Jesús A Ballesteros** , Kean University, Union, USA
- Gabriela Pozo** , Universidad San Francisco de Quito, Quito, Ecuador

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

Corresponding authors: Liam M. Crowley (liam.crowley@biology.ox.ac.uk), Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Telfer MG: Investigation, Resources; Crowley LM: Investigation, Resources; Wilding CS: Writing – Original Draft Preparation;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (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.

Copyright: © 2025 Telfer MG *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: Telfer MG, Crowley LM, Wilding CS *et al.* **The genome sequence of the furrow orbweaver, *Larinioides cornutus* (Clerck, 1757) [version 1; peer review: 3 approved]** Wellcome Open Research 2025, 10:267 <https://doi.org/10.12688/wellcomeopenres.24288.1>

First published: 23 May 2025, 10:267 <https://doi.org/10.12688/wellcomeopenres.24288.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Chelicerata; Arachnida; Araneae; Araneomorphae; Entelegynae; Orbiculariae; Araneoidea; Araneidae; *Larinioides*; *Larinioides cornutus* (Clerck, 1757) (NCBI:txid336591)

Background

Of the three members of the genus *Larinioides* found in the UK and Ireland (Bee *et al.*, 2020; Lavery, 2019), *Larinioides cornutus* is both the most common and most widespread. It has a holarctic distribution, being found from North America, through Europe, Russia, areas of the Middle East and into Asia (https://wsc.nmbe.ch/species/4277/Larinioides_cornutus). In the UK, it can be found throughout England and Wales but has a more scattered distribution in Scotland. On the island of Ireland, records exist for all but four Irish, and two Northern Irish counties (Van Helsdingen, 1996), although these gaps are likely reflective of gaps in sampling effort or reporting.

Larinioides cornutus reaches sizes of 5–8 mm (♂) and 6–9 mm (♀) (Bee *et al.*, 2020; Locket & Millidge, 1951; Roberts, 1987), and has a pale brown carapace covered with white hairs, and pale, sometimes annulated legs. The abdomen is variable in colour; from cream through to orange-brown and is decorated with an obvious dark folium with light transverse stripes posteriorly, and a pale cardiac region with central dark patch (Bee *et al.*, 2020; Locket & Millidge, 1951; Roberts, 1987).

This species is typically found close to water on tall grasses and reeds, as well as bridges and posts, where it spins large, sometimes prominent webs; however, it may be found also in damp, rough grassland and meadows (Bee *et al.*, 2020; Harvey *et al.*, 2002). The web may appear empty during the day since the spider hides in a silken retreat constructed high up in vegetation adjacent to the web and camouflaged with plant and animal matter, emerging only at night.

Whilst a transcriptome for *L. cornutus* has been produced previously (Wilson *et al.*, 2020), no genome resource exists. To this end, we present a chromosomally complete genome sequence for *L. cornutus* based on an individual (Figure 1) collected from Cothill Fen, Oxfordshire. This will aid in future population genetic studies of this widespread species, as well as contributing to broader studies of the biology and phylogeny of the order Araneae (Garb *et al.*, 2018).

Genome sequence report

Sequencing data

The genome of a specimen of *Larinioides cornutus* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 67.91 Gb from 8.03 million reads, which were used to assemble the genome. GenomeScope analysis estimated the haploid genome size at 2,258.74 Mb, with a heterozygosity of 2.43% and repeat content of 27.74%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided



Figure 1. Photograph of the *Larinioides cornutus* (qqLarCorn3) specimen used for genome sequencing.

approximately 29x coverage. Hi-C sequencing produced 480.07 Gb from 3,179.24 million reads and was 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 assembly was improved by manual curation, which corrected 96 misjoins or missing joins. These interventions decreased the scaffold count by 6.52%. The final assembly has a total length of 2,304.58 Mb in 501 scaffolds, with 732 gaps, and a scaffold N50 of 175.46 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (98.49%) was assigned to 13 chromosomal-level scaffolds, representing 11 autosomes and the X₁ and X₂ sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, the X₁ and X₂ chromosomes were assigned based on homology to the genome of *Gibbaranea gibbosa* (GCA_964059485.1) (Crowley *et al.*, 2025).

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 (Accession number OZ204543.1).

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer

Table 1. Specimen and sequencing data for *Larinioides cornutus*.

Project information			
Study title	Larinioides cornutus		
Umbrella BioProject	PRJEB82167		
Species	<i>Larinioides cornutus</i>		
BioSpecimen	SAMEA114644584		
NCBI taxonomy ID	336591		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	qqLarCorn3	SAMEA114645200	head and thorax
Hi-C sequencing	qqLarCorn3	SAMEA114645200	head and thorax
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR13947480	1.58e+09	239.33
Hi-C Illumina NovaSeq X	ERR13947481	1.59e+09	240.74
PacBio Revio	ERR13946461	8.03e+06	67.91

Table 2. Genome assembly data for *Larinioides cornutus*.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	qqLarCorn3.hap1.1	qqLarCorn3.hap2.1
Assembly accession	GCA_964340895.1	GCA_964340905.1
Assembly level	chromosome	scaffold
Span (Mb)	2,304.58	2,295.32
Number of contigs	1,233	1,568
Number of scaffolds	501	871
Longest scaffold (Mb)	195.78	-
Assembly metrics (benchmark)	Haplotype 1	Haplotype 2
Contig N50 length (≥ 1 Mb)	5.57 Mb	5.23 Mb
Scaffold N50 length (= chromosome N50)	175.46 Mb	171.8 Mb
Consensus quality (QV) (≥ 40)	61.9	61.7
<i>k</i> -mer completeness	64.94%	64.69%
Combined <i>k</i> -mer completeness (≥ 95%)	99.59%	
BUSCO* (S > 90%; D < 5%)	C:98.2%[S:92.5%,D:5.6%], F:0.8%,M:1.1%,n:2,934	C:98.1%[S:92.4%,D:5.7%], F:0.7%,M:1.2%,n:2,934
Percentage of assembly mapped to chromosomes (≥ 90%)	98.49%	-
Sex chromosomes (localised homologous pairs)	X ₁ and X ₂	-
Organelles (one complete allele)	Mitochondrial genome: 14.6 kb	-

*BUSCO scores based on the arachnida_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

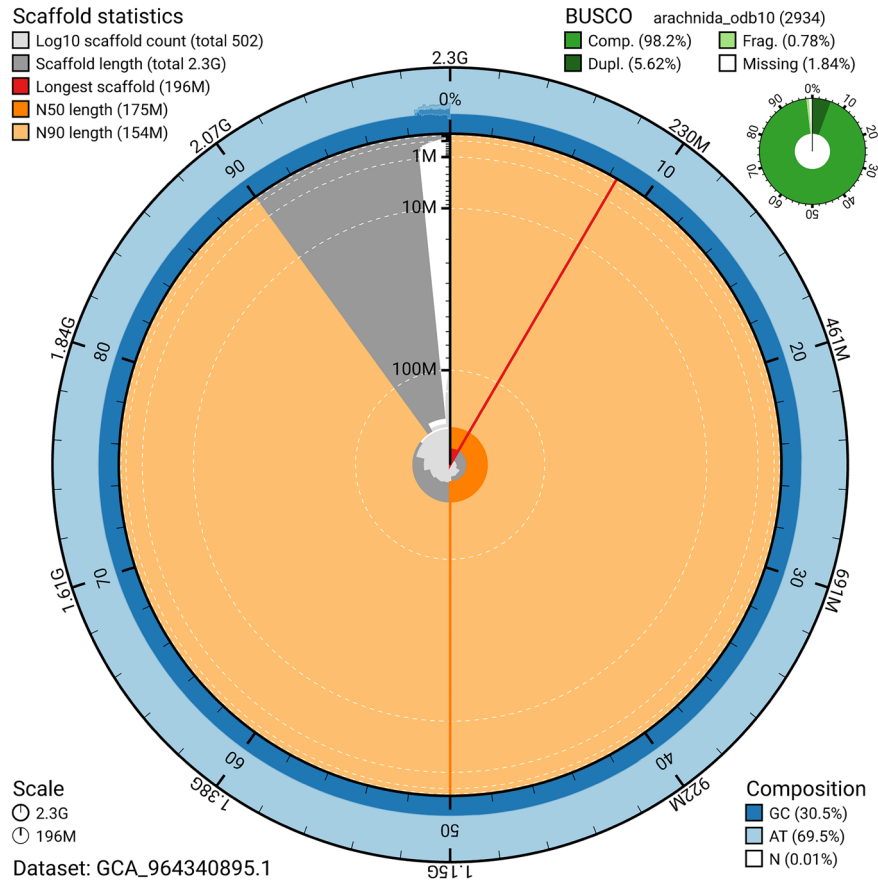


Figure 2. Genome assembly of *Larinioides cornutus*, qqLarCorn3.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the 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 arachnida_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964340895.1/dataset/GCA_964340895.1/snail.

completeness indicates the proportion of expected k -mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

For haplotype 1, the estimated QV is 61.9, and for haplotype 2, 61.7. When the two haplotypes are combined, the assembly achieves an estimated QV of 61.8. The k -mer recovery for haplotype 1 is 64.94%, and for haplotype 2 64.69%, while the combined haplotypes have a k -mer recovery of 99.59%. BUSCO v.5.5.0 analysis using the arachnida_odb10 reference set ($n = 2,934$) identified 98.2% of the expected gene set (single = 92.5%, duplicated = 5.6%) for haplotype 1.

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The haplotype 1 assembly achieves the EBP reference standard of 6.C.Q61.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Larinioides cornutus* (specimen ID Ox003518, ToLID qqLarCorn3), collected from Cothill Fen, Oxfordshire, United Kingdom (latitude 51.695, longitude -1.335) on 2023-05-12, using a sweep net. The specimen was collected by Mark Telfer and Liam Crowley (University of Oxford) and identified by Liam Crowley, and then preserved on dry ice.

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) (Pereira *et al.*, 2022). 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

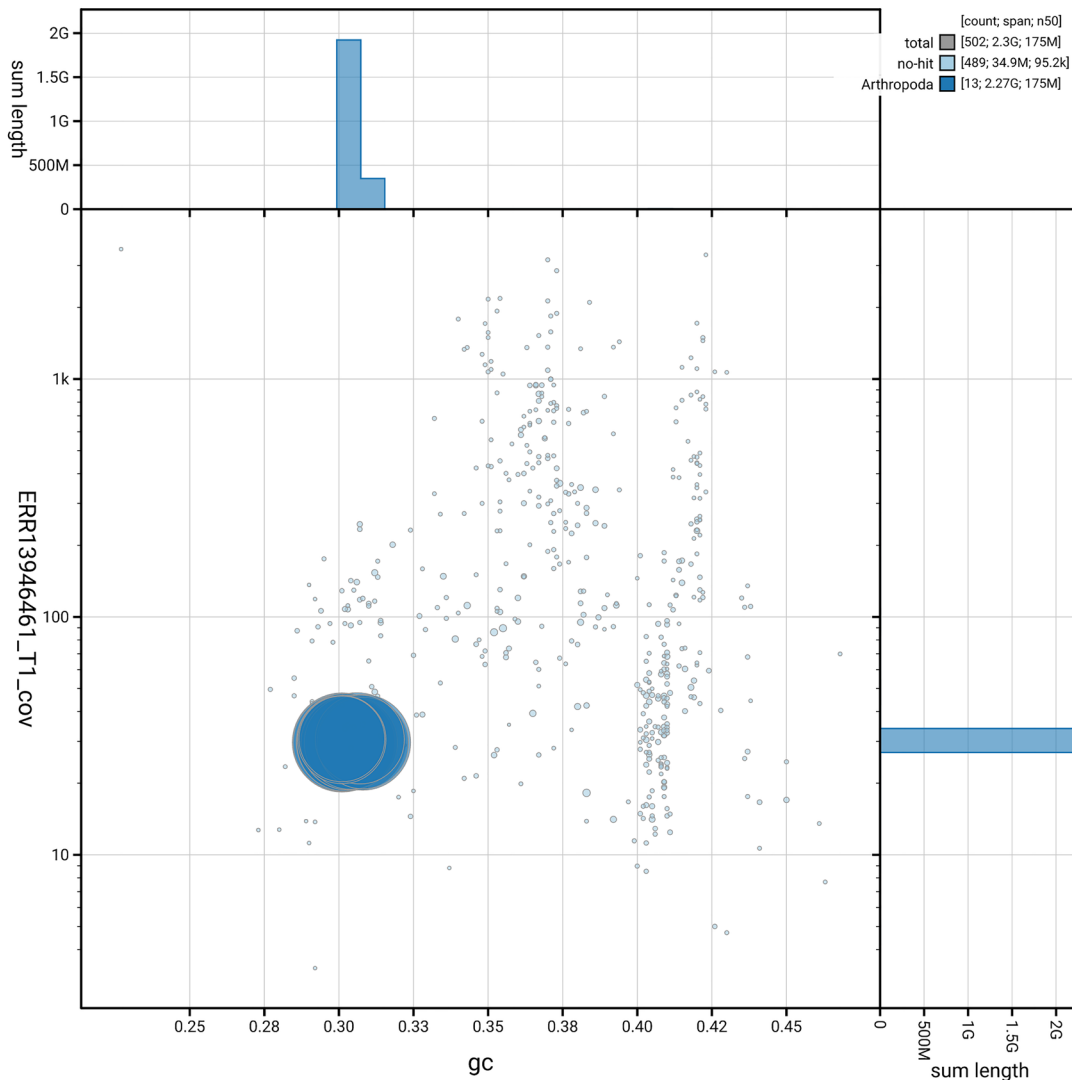


Figure 3. Genome assembly of *Larinioides cornutus*, qqLarCorn3.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum 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 at https://blobtoolkit.genomehubs.org/view/GCA_964340895.1/dataset/GCA_964340895.1/blob.

identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode 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 have been deposited on protocols.io (Beasley *et al.*, 2023).

Metadata collection for samples adhered to the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life

Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification (Howard *et al.*, 2025). Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The qqLarCorn3 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the cephalothorax was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023).

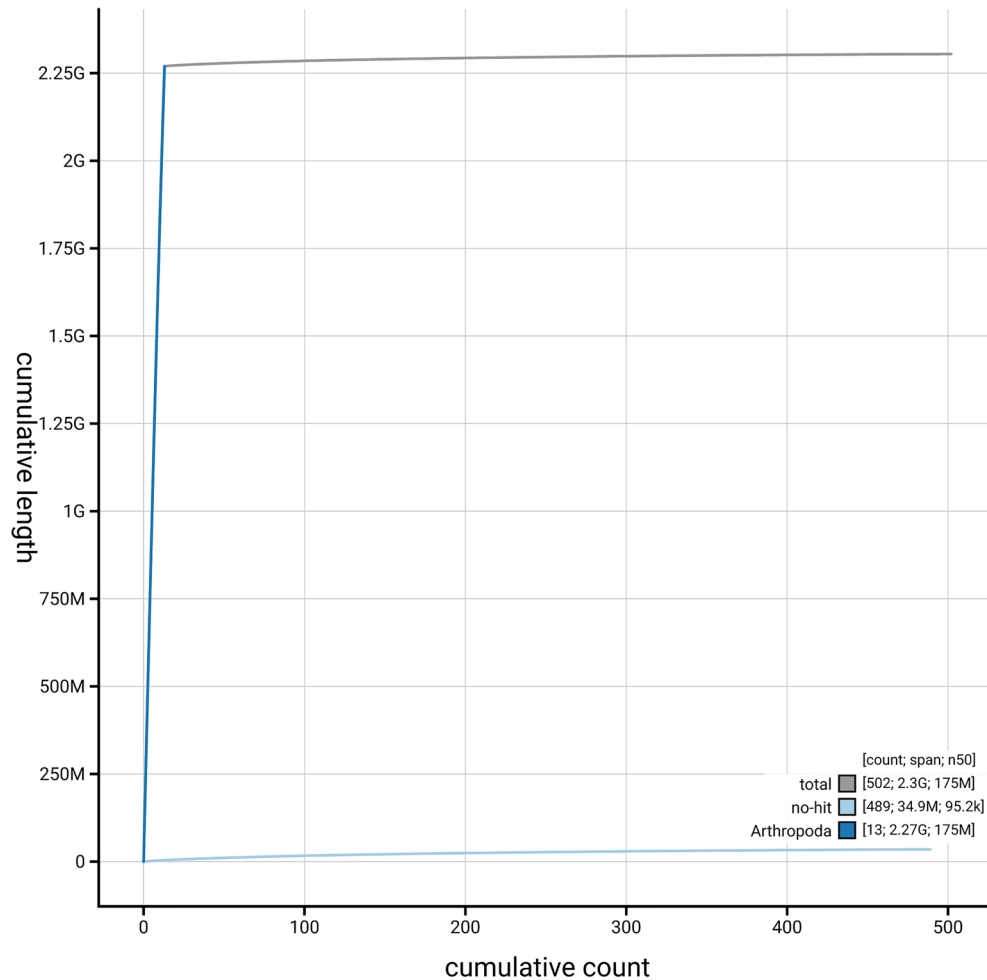


Figure 4. Genome assembly of *Larinioides cornutus*, qqLarCorn3.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964340895.1/dataset/GCA_964340895.1/cumulative.

Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). 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.

Hi-C sample preparation and crosslinking

Hi-C data were generated from the qqLarCorn3 sample using the Arima-HiC v2 kit (Arima Genomics) with 20–50 mg of frozen tissue (stored at -80°C). As per manufacturer's instructions, tissue was fixed, and the DNA crosslinked using a TC buffer with a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagenocine

Power Masher-II. The crosslinked DNA was digested using a restriction enzyme master mix, then biotinylated and ligated. A clean up was performed with SPRIselect beads prior to library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit, and sample biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core.

PacBio HiFi

Samples with an average fragment size greater than 8 kb and total mass exceeding 400 ng were eligible for the low-input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California,

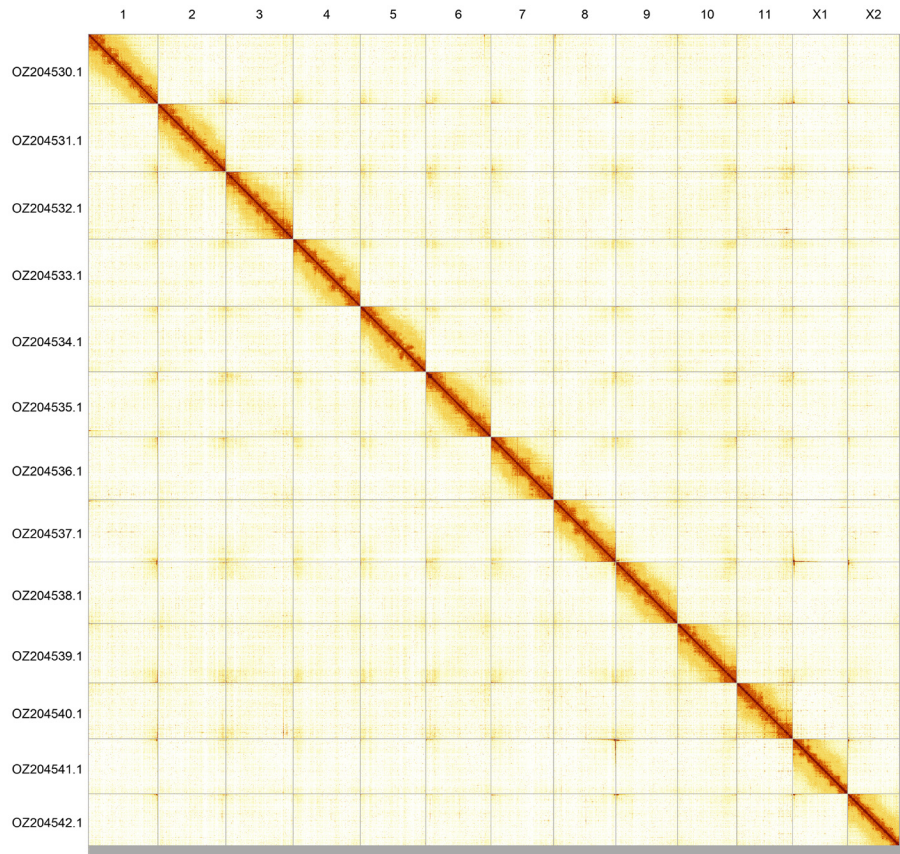


Figure 5. Genome assembly of *Larinioides cornutus*. Hi-C contact map of the qqLarCorn3.hap1.1 assembly, generated using PretextSnapshot. Chromosomes are shown in order of size and labelled with chromosome numbers (top) and chromosome accession numbers (left).

Table 3. Chromosomal pseudomolecules in the genome assembly of *Larinioides cornutus*, qqLarCorn3.

INSDC accession	Name	Length (Mb)	GC%
OZ204530.1	1	195.78	30
OZ204531.1	2	189.23	30
OZ204532.1	3	188.7	30.5
OZ204533.1	4	188.04	30
OZ204534.1	5	183.06	30
OZ204535.1	6	182.09	31
OZ204536.1	7	175.46	30.5
OZ204537.1	8	174.12	30.5
OZ204538.1	9	172.09	30.5
OZ204539.1	10	165.83	31
OZ204540.1	11	156.76	30.5
OZ204543.1	MT	0.01	22.5
OZ204541.1	X ₁	153.77	30
OZ204542.1	X ₂	144.78	30

USA), depending on genome size and required sequencing depth. Libraries were prepared using the SMRTbell Prep Kit 3.0 according to 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 μ L 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.

Hi-C

For Hi-C library preparation, the biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size-selected to 400–600 bp using SPRISelect beads. DNA was then enriched using Arima-HiC v2 Enrichment beads. The NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) was used for end repair, A-tailing, and adapter ligation, following a modified protocol in which library preparation is carried out while the DNA remains bound to the enrichment beads. PCR amplification was performed using KAPA HiFi HotStart mix and custom dual-indexed adapters (Integrated DNA Technologies) in a 96-well plate format. Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, samples were amplified for 10–16 PCR cycles. Post-PCR clean-up was carried out using SPRISelect beads. The libraries were quantified using the Accuclear Ultra High Sensitivity dsDNA Standards Assay kit (Biotium) and normalised to 10 ng/μL before sequencing. Hi-C sequencing was performed on the Illumina NovaSeq X instrument using 150 bp paired-end reads.

Genome assembly, curation and evaluation

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), resulting in a pair of haplotype-resolved assemblies. The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided 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. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. Sex

chromosomes were identified by synteny analysis. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextView was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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 (Di Tommaso *et al.*, 2017) 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).

Table 4 contains a list of relevant software tool versions and sources.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the ‘**Darwin Tree of Life Project Sampling Code of Practice**’, which can be found in full on the Darwin Tree of Life website [here](https://www.darwintreeoflife.org/). 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.

Table 4. Software tools: versions and sources.

Software tool	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	666652151335353eef2fcd58880bcef5bc2928e1	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	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	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
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

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.

Data availability

European Nucleotide Archive: *Larinioides cornutus*. Accession number PRJEB82167; <https://identifiers.org/ena.embl/PRJEB82167>. The genome sequence is released openly for reuse. The *Larinioides cornutus* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665). 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 [Table 1](#) and [Table 2](#).

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12157525>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.4783558>.

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](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor³ for LI PacBio.** *protocols.io.* 2023. [Publisher Full Text](#)
- Beasley J, Uhl R, Forrest LL, *et al.*: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023; [Accessed 25 June 2024]. [Publisher Full Text](#)
- Bee L, Oxford G, Smith H: **Britain's spiders (Second Edition).** Old Basing, Hampshire, UK: WILDGuides, 2020. [Publisher 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](#)
- Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crowley LM, Wilding CS, University of Oxford and Wytham Woods Genome Acquisition Lab, *et al.*: **The genome sequence of an orbweaving spider, *Gibbaranea gibbosa* (Walckenaer, 1802) [version 1; peer review: 2 approved, 1 approved with reservations].** *Wellcome Open Res.* 2025; **10**: 97. [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): g1ab008. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a. [Publisher Full Text](#)
- Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b. [Publisher Full Text](#)
- Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319. [PubMed Abstract](#) | [Publisher Full Text](#)
- Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74. [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.** *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](#)
- Garb JE, Sharma PP, Ayoub NA: **Recent progress and prospects for advancing arachnid genomics.** *Curr Opin Insect Sci.* 2018; **25**: 51–57. [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](#)
- Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps.** 2022. [Reference Source](#)
- Harvey PR, Nellist DR, Telfer MG, *et al.*: **Provisional Atlas of British Spiders (Arachnida, Aranae), volumes 1 & 2.** Huntingdon: Biological Records Centre, 2002. [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): g1aa153. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023. [Publisher Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)

Lavery A: **A revised checklist of the spiders of Great Britain and Ireland.** *Arachnology.* 2019; **18**(3): 196–212.

[Publisher Full Text](#)

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.

[Publisher Full Text](#)

Li H: **Minimap2: pairwise alignment for nucleotide sequences.**

Bioinformatics. 2018; **34**(18): 3094–3100.

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

Lockett GH, Millidge AF: **British spiders.** The Ray Society, 1951.

[Reference Source](#)

Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.

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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].

[Reference Source](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.

[Publisher Full Text](#)

Pereira L, Sivell O, Sivess L, *et al.*: **DTOL taxon-specific standard operating procedure for the terrestrial and freshwater arthropods working group.** 2022.

[Publisher Full Text](#)

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.

[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.*: **Mercury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.

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

Roberts MJ: **The spiders of Great Britain and Ireland. Volume 1. Atyidae – Theridiosomatidae.** Harley Books, 1987.

[Reference Source](#)

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

[Publisher Full Text](#)

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *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](#)

Van Helsdingen PJ: **The county distribution of Irish spiders, incorporating a revised catalogue of the species.** *Ir Nat J.* 1996; **25**: 1–92.

[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](#)

Wilson RJ, Ahmed TH, Rahman MM, *et al.*: **Identification and activity of monoamine oxidase in the orb-weaving spider *Larinioides cornutus*.** *Gen Comp Endocrinol.* 2020; **299**: 113580.

[PubMed Abstract](#) | [Publisher 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 05 November 2025

<https://doi.org/10.21956/wellcomeopenres.26786.r126103>

© 2025 Pozo G. 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.



Gabriela Pozo 

Universidad San Francisco de Quito, Quito, Ecuador

This manuscript presents a haplotype-resolved genome assembly for a spider species, generated using long-read sequencing and phased with Hi-C data. The authors report a genome size of approximately 2,258.74 Mb and provide assemblies for two haplotypes. The study is well-structured, with clearly described methods, informative results, and public accessibility of the genomic data. This genome represents an important contribution to arachnid genomics and provides a valuable resource for future studies of spider biology, evolution, and comparative genomics.

I have two suggestions for the authors:

1. You mention that this study will contribute to broader studies of the biology and phylogeny of the order Araneae. I suggest briefly stating here what the current state of sequenced genomes in this order is, if there is representativity or not.
2. In table 2, I suggest stating how many BUSCO groups the database you used has (arachnida_odb10), like you do for the snailplot.

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

Partly

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics, genomics

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

Reviewer Report 04 November 2025

<https://doi.org/10.21956/wellcomeopenres.26786.r126100>

© 2025 Ballesteros J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jesús A Ballesteros 

Kean University, Union, New Jersey, USA

This manuscript presents the chromosome genome assembly of the furrow spider "Lariniodes cornutus". This orbweaving spider, common across the holartic region, represents a valuable genomic resource for the broad scientific community. The dataset presented in this contribution can be readily useful across various biological disciplines: population genetics, phylogenetics, silk, venom, etc. The detail and clarity in the method descriptions are also very valuable for future projects targeting non model organism. In consequence I believe the manuscript is indexable in its current format.

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: arachnids, phylogenetics, 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 23 June 2025

<https://doi.org/10.21956/wellcomeopenres.26786.r124283>

© 2025 Jin P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Peng-Yu Jin 

¹ University of the Chinese Academy of Sciences, Beijing, Beijing, China

² Institute of Zoology, Chinese Academy of Sciences, Beijing, Beijing, China

The manuscript presents a high-quality chromosomal-level genome assembly of *Larinioides cornutus*, an ecologically widespread orb-weaving spider. They generated a robust dataset using PacBio HiFi and Hi-C, achieving a highly contiguous assembly. Quality assessments indicate excellent completeness and accuracy. The study follows current community standards for genome assembly, including EBP reference standard, and provides sufficient methodological transparency to ensure reproducibility. Overall, this is a strong contribution and merits publication.

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: Spider evolution; Single-cell RNA-seq; Spider genomics

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