



## DATA NOTE

# The genome sequence of an orb-weaver spider, *Araneus angulatus* Clerck, 1757 (Araneae: Araneidae)

[version 1; peer review: 2 approved]

Christopher M. Raper<sup>1</sup>, Olga Sivell <sup>1</sup>,  
Natural History Museum 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

<sup>1</sup>Natural History Museum, London, England, UK

**V1** First published: 21 Nov 2025, 10:656  
<https://doi.org/10.12688/wellcomeopenres.25196.1>  
Latest published: 21 Nov 2025, 10:656  
<https://doi.org/10.12688/wellcomeopenres.25196.1>

## Abstract

We present a genome assembly from an individual female *Araneus angulatus* (orb-weaver spider; Arthropoda; Arachnida; Araneae; Araneidae). The assembly contains two haplotypes with total lengths of 2 980.91 megabases and 2 941.08 megabases. Most of haplotype 1 (94.74%) is scaffolded into 28 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 14.53 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

## Keywords

*Araneus angulatus*; orb-weaver spider; genome sequence; chromosomal; Araneae



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

## Open Peer Review

### Approval Status

|                  | 1                                                                                     | 2                                                                                     |
|------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| <b>version 1</b> |  |  |
| 21 Nov 2025      | <a href="#">view</a>                                                                  | <a href="#">view</a>                                                                  |

1. **Adile Akpınar**, Gaziantep University,  
Gaziantep, Turkey

2. **Bastian Fromm** , UiT- The Arctic  
University of Norway, Tromsø, Norway

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

**Corresponding author:** Darwin Tree of Life Consortium ([mark.blaxter@sanger.ac.uk](mailto:mark.blaxter@sanger.ac.uk))

**Author roles:** Raper CM: Investigation, Resources; Sivell O: Writing – Original Draft Preparation, Writing – Review & Editing;

**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 Raper CM *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:** Raper CM, Sivell O, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of an orb-weaver spider, *Araneus angulatus* Clerck, 1757 (Araneae: Araneidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:656 <https://doi.org/10.12688/wellcomeopenres.25196.1>

**First published:** 21 Nov 2025, 10:656 <https://doi.org/10.12688/wellcomeopenres.25196.1>

## Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Chelicerata; Arachnida; Araneae; Araneomorphae; Entelegynae; Orbiculariae; Araneoidea; Araneidae; Araneus; *Araneus angulatus* Clerck, 1757 (NCBI:txid1112382)

## Background

*Araneus angulatus* Clerck 1757 is a species from family Araneidae (orbweb spiders or orb-weavers) and a type species of the genus *Araneus* Clerck, 1757. This is a fairly large spider, with a body length of 12–19 mm in females, and 10–12 mm in males (Roberts, 1996). The spider is greyish-brown in colour, with reddish-brown carapace covered with long white hairs. The pale brown abdomen has a pair of distinct raised “humps”, somewhat similar to *Gibbranea gibbosa*, and a pattern tapering posteriorly, occasionally with white markings medially (Bee *et al.*, 2017; Bee *et al.*, 2020). Males are similar in appearance to females, however with a smaller abdomen and generally darker in colour. The structure of the epigyne and male palpal organ are distinctive and pictured in Roberts (1996).

*Araneus angulatus* occurs in Europe, North Africa, Turkey, Russia (Europe to Far East), Iran, Central Asia and Korea (Nentwig *et al.*, 2025a; World Spider Catalogue, 2025). In Britain it is very local and scarce, mostly restricted to southern coastal counties in England, also recorded from south Wales and Surrey. It occurs in deciduous woodland, occasionally on pine and scrub on maritime heath (Bee *et al.*, 2017; Bee *et al.*, 2020; British Arachnological Society, 2025; Dawson *et al.*,). The webs can be built high up in the trees, suspended from a long frame thread (Bee *et al.*, 2017; Bee *et al.*, 2020).

This species is considered Nationally Scarce in Britain (British Arachnological Society, 2025; Harvey *et al.*, 2017), and is included on the Red List for Sweden (Gärdenfors, 2000). In Britain the males have been recorded from July to August, females from March to August; both sexes mature in June, and mature females can be encountered till September (Bee *et al.*, 2017; Bee *et al.*, 2020; British Arachnological Society, 2025).

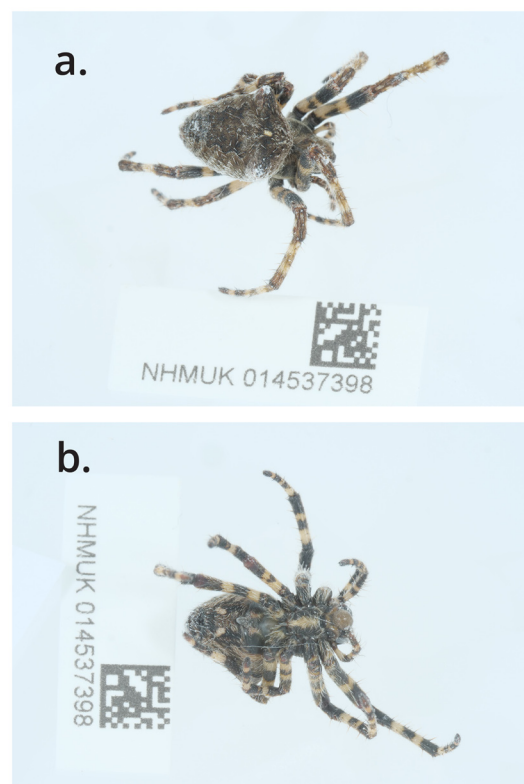
According to the World Spider Catalogue (2025), there are currently 198 genera recognised within Araneidae and the genus *Araneus* includes 550 species. The family Araneidae and the genus *Araneus* are both paraphyletic according to Scharff *et al.* (2020); Scharff & Coddington, (1997), while Dimitrov *et al.* (2017) and Wheeler *et al.* (2017) consider the family to be monophyletic. The phylogenetic relationships and classification of the araneids require more extensive and in-depth research.

The recent phylogenetic analysis of the Araneidae (Scharff *et al.*, 2020) did not include *A. angulatus* due to the lack of suitable samples. The sequences published more recently on BOLD and GenBank suggest the *A. angulatus* may be more than one species (Ratnasingham *et al.*, 2024; Scharff *et al.*, 2020). However, it is also possible that some of the specimens of other species, e.g. *A. pallidus*, might have been misidentified as *A. angulatus*. *Araneus angulatus pallidus* (Franganillo, 1909),

*A. a. crucineptus* (Franganillo, 1909), *A. a. fuscus* (Franganillo, 1909), *A. a. iberoi* (Franganillo, 1909), *A. a. castaneus* (Franganillo, 1909), previously considered subspecies of *A. angulatus*, have since been synonymised with *Araneus pallidus* (Olivier, 1789), a variable species with several colour forms, occurring in France, Portugal, Spain, Algeria and Morocco (Breitling *et al.*, 2016; Nentwig *et al.*, 2025b; World Spider Catalogue, 2025; Wunderlich, 2023). *Araneus angulatus afolius* (Franganillo, 1909), *A. a. levifolius* (Franganillo, 1909), *A. a. niger* (Franganillo, 1918) and *A. a. nitidifolius* (Franganillo, 1909) were synonymised with *A. angulatus* (Breitling *et al.*, 2016). Characters distinguishing *A. angulatus* and *A. pallidus* as well as different colour forms are discussed by Breitling *et al.* (2016), the pedipalpus of *A. pallidus* is figured in Grasshoff (1968) and Wunderlich (2023) and a key to *Araneus* Clerck, 1757 is available online.

The identity of *A. angulatus* is important for our understanding of the phylogeny and taxonomy of Araneidae and this publication aims to contribute to the body of knowledge by providing a high-quality genome of *A. angulatus* from Britain. The mitochondrial genome of this species was sequenced from a specimen collected from China (Wang *et al.*, 2019).

The high-quality genome of *Araneus angulatus* was sequenced from a single male (NHMUK014537398; SAMEA11024980) from Kings Wood, St Austell, England (Figure 1). The genome



**Figure 1.** Photographs of the *Araneus angulatus* (qqAraAngu1) specimen used for genome sequencing.

of *A. angulatus* presented here was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. It will aid research into phylogeny of spiders and taxonomy, biology and ecology of the species.

## Methods

### Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Araneus angulatus* (specimen ID NHMUK014537398, ToLID qqAraAngul; [Figure 1](#)), collected from Kings Wood, St Austell, England, United Kingdom (latitude 50.3045, longitude -4.8001) on 2021-06-28. The specimen was collected and identified by Chris Raper and Chris Spilling, and processed at the Natural History Museum before being shipped to the Wellcome Sanger Institute. Sample metadata were collected in line with the Darwin Tree of Life project standards described by [Lawniczak et al. \(2022\)](#).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by [Twyford et al. \(2024\)](#). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the [protocol](#)). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification ([Crowley et al., 2023](#)). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI ([Twyford et al., 2024](#)). The standard operating procedures for Darwin Tree of Life barcoding are available on [protocols.io](#).

### Nucleic acid extraction

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 qqAraAngul sample was weighed and [triaged](#) to determine the appropriate extraction protocol. Tissue from the abdomen 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 70.8 ng/μL and a yield of 3 405.48 ng, with a fragment size of 12.5 kb.

### 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 using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

## Hi-C

### Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen cephalothorax tissue of the qqAraAngul 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 were created. 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 6000.

## 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 organelle genomes were assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

## 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 63 breaks and 167 joins, which reduced the scaffold count by 3.3%. The curation process is described 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

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate  $k$ -mer completeness and assembly quality for both haplotypes using the  $k$ -mer 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 version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The

BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

## Genome sequence report

### Sequence data

PacBio sequencing of the *Araneus angulatus* specimen generated 129.23 Gb (gigabases) from 12.35 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 2 902.58 Mb, with a heterozygosity of 0.68% and repeat content of 39.48% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 51× coverage. Hi-C sequencing produced 239.79 Gb from 1 588.02 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 2 980.91 Mb in 1 908 scaffolds, with 499 gaps, and a scaffold N50 of 102.38 Mb (Table 2).

Most of the haplotype 1 assembly sequence (94.74%) was assigned to 28 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Sex chromosomes could not be identified, due to the lack of a related comparator. Scaffolds on Chromosome 18 in the regions 9.7–16.1 Mb and 76.5–85.3 Mb have uncertain order and orientation.

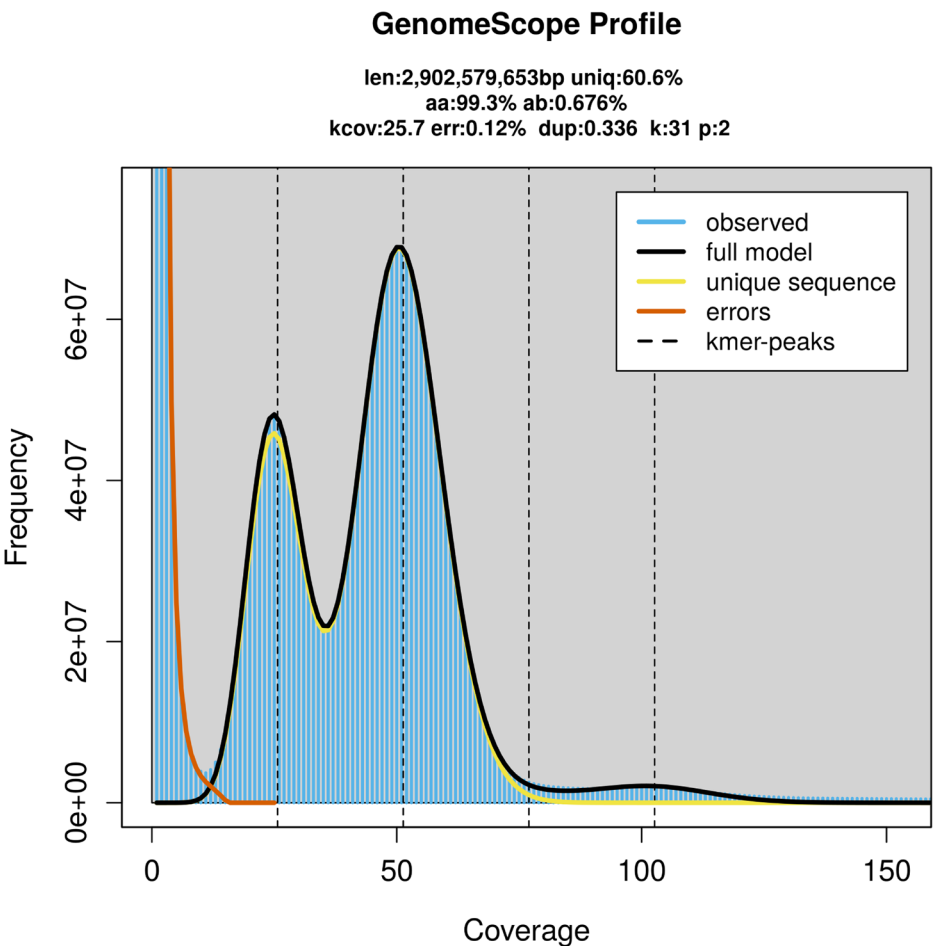
The mitochondrial genome was also assembled (length 14.53 kb, OZ287229.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 59.2, and for haplotype 2, 60.2. When the two haplotypes are combined, the assembly achieves an estimated QV of 59.7. The  $k$ -mer completeness is 84.41% for haplotype 1, 84.43% for haplotype 2, and 99.13% for the combined haplotypes (Figure 4).

BUSCO analysis using the arachnida\_odb10 reference set ( $n = 2 934$ ) identified 98.5% of the expected gene set (single = 92.3%, duplicated = 6.2%) 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) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric,





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

**Table 1.** Specimen and sequencing data for BioProject PRJEB73665.

| Platform                      | PacBio HiFi                                                 | Hi-C                  |
|-------------------------------|-------------------------------------------------------------|-----------------------|
| ToLID                         | qqAraAngu1                                                  | qqAraAngu1            |
| Specimen ID                   | NHMUK014537398                                              | NHMUK014537398        |
| BioSample (source individual) | SAMEA11024980                                               | SAMEA11024980         |
| BioSample (tissue)            | SAMEA11025175                                               | SAMEA11025178         |
| Tissue                        | abdomen                                                     | cephalothorax         |
| Instrument                    | Sequel IIe                                                  | Illumina NovaSeq 6000 |
| Run accessions                | ERR12736889;<br>ERR12736890;<br>ERR12736891;<br>ERR12736892 | ERR12737284           |
| Read count total              | 12.35 million                                               | 1 588.02 million      |
| Base count total              | 129.23 Gb                                                   | 239.79 Gb             |

Table 2. Genome assembly statistics.

|                              |                         |                   |
|------------------------------|-------------------------|-------------------|
| Assembly name                | qqAraAngu1.hap1.1       | qqAraAngu1.hap2.1 |
| Assembly accession           | GCA_965645275.1         | GCA_965645295.1   |
| Assembly level               | chromosome              | scaffold          |
| Span (Mb)                    | 2 980.91                | 2 941.08          |
| Number of chromosomes        | 28                      | scaffold-level    |
| Number of contigs            | 2 407                   | 1 259             |
| Contig N50                   | 10.18 Mb                | 11.16 Mb          |
| Number of scaffolds          | 1 908                   | 779               |
| Scaffold N50                 | 102.38 Mb               | 97.17 Mb          |
| Longest scaffold length (Mb) | 160.13                  |                   |
| Organelles                   | Mitochondrion: 14.53 kb |                   |

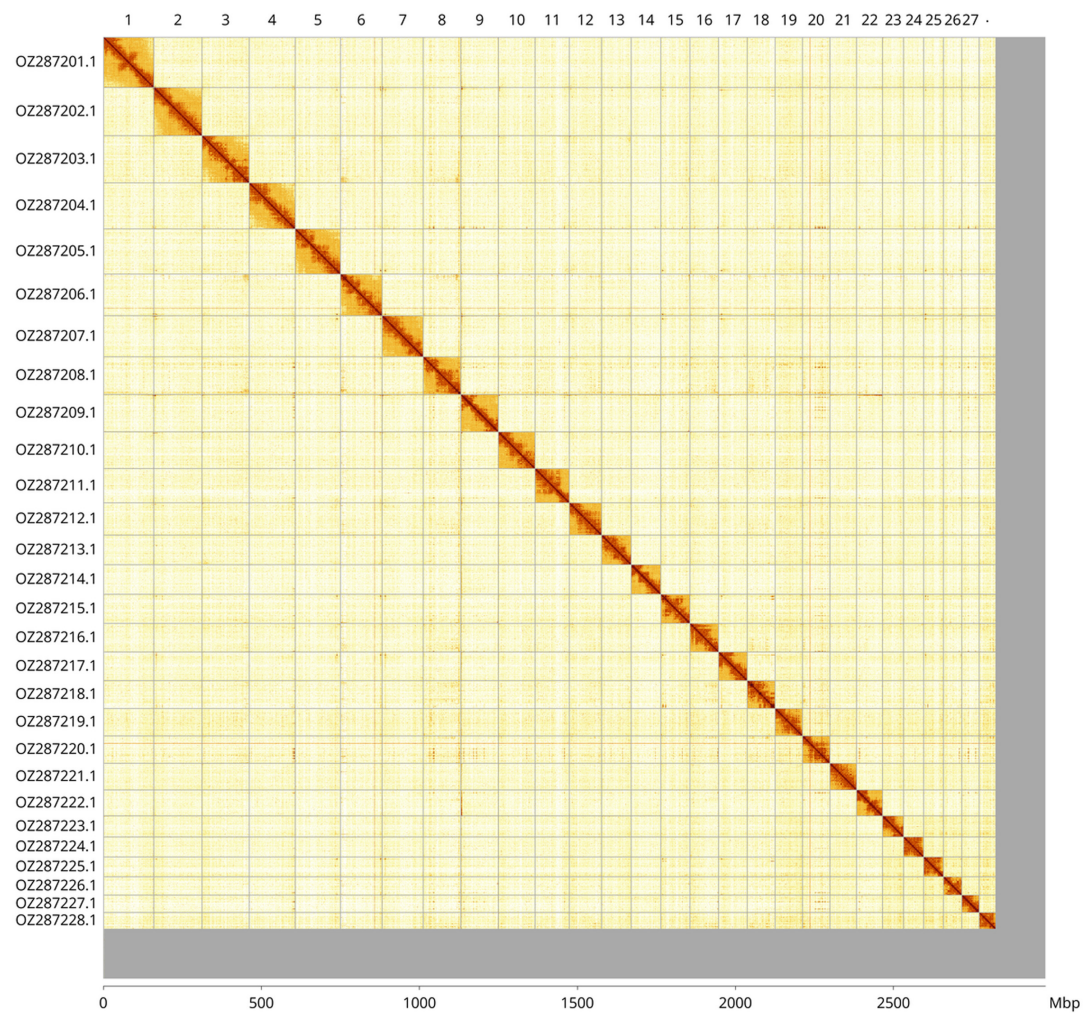
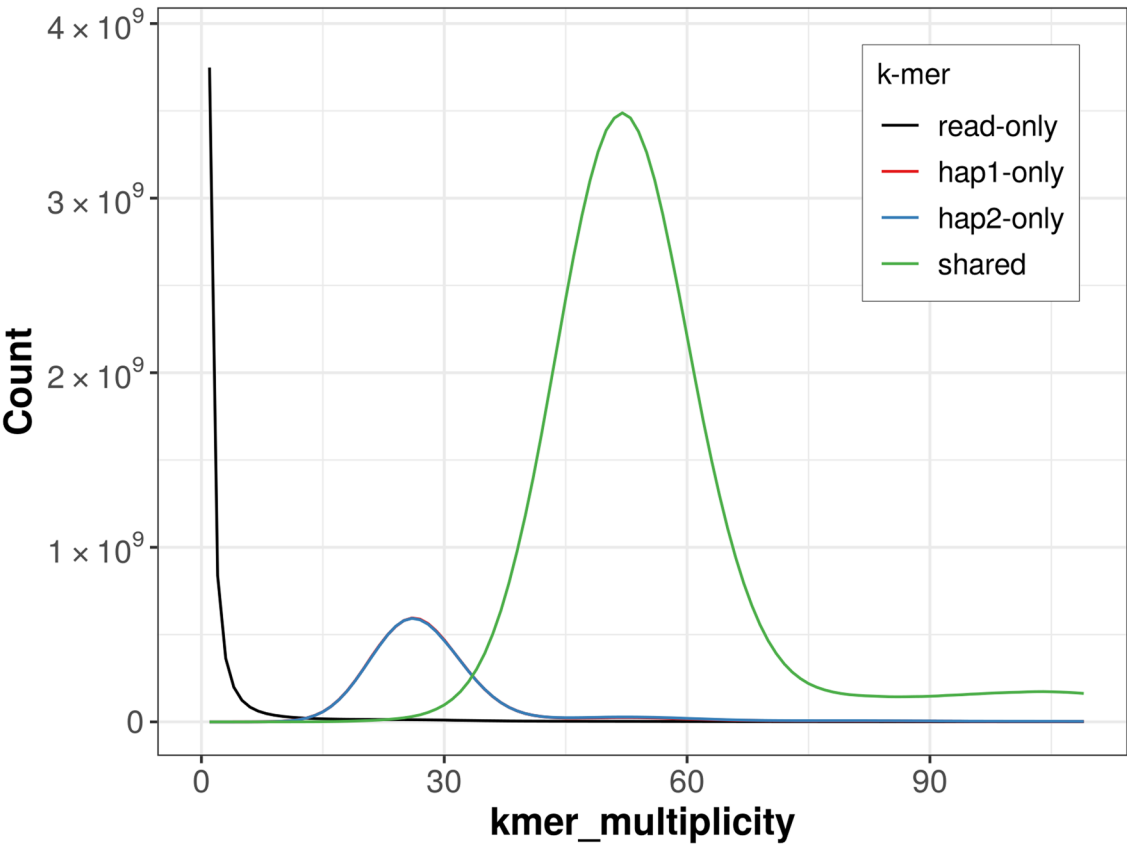


Figure 3. Hi-C contact map of the *Araneus angulatus* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

**Table 3.** Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Araneus angulatus* qqAraAngu1.

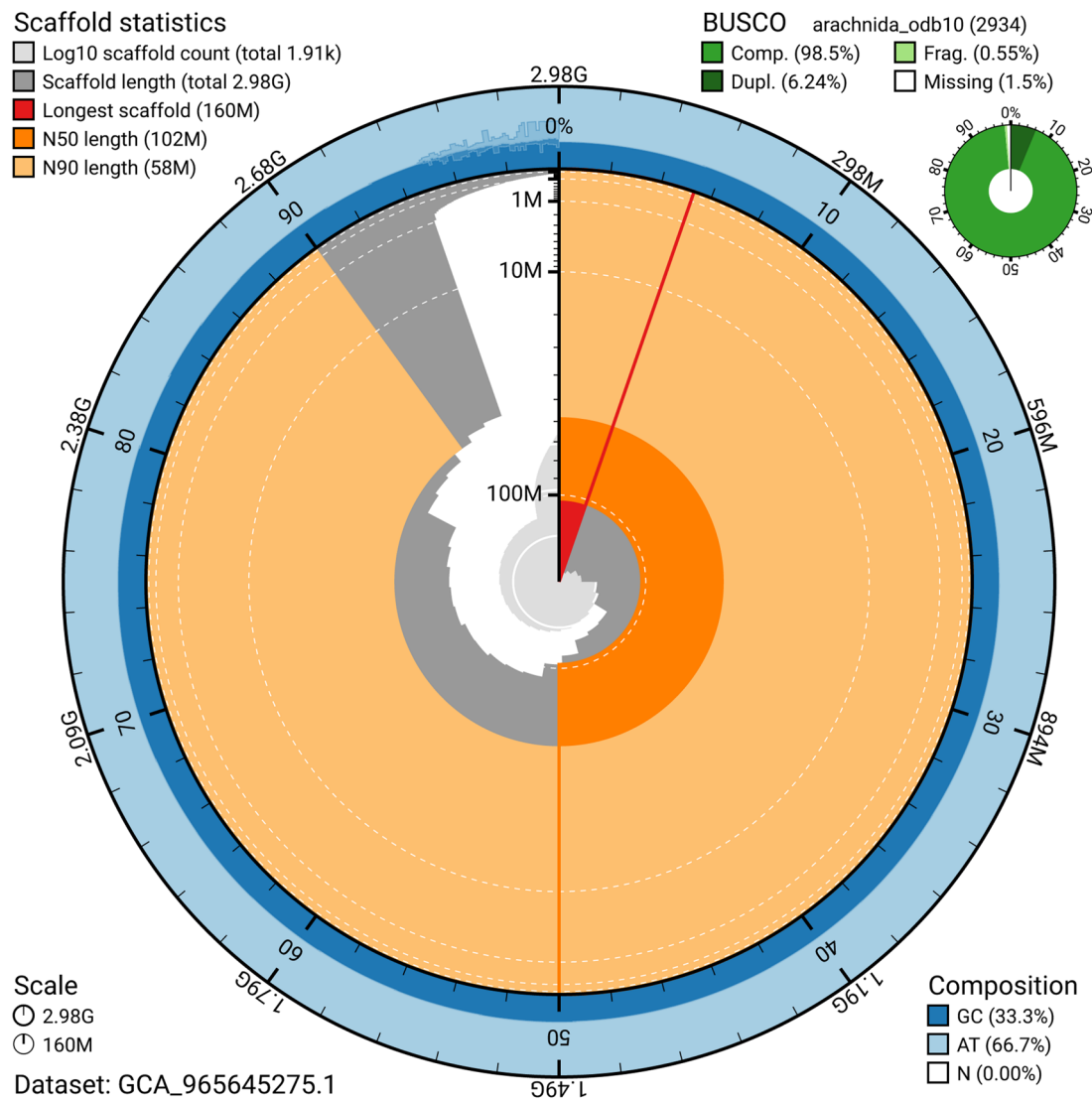
| INSDC accession | Molecule | Length (Mb) | GC%   |
|-----------------|----------|-------------|-------|
| OZ287201.1      | 1        | 160.13      | 33    |
| OZ287202.1      | 2        | 152.26      | 33.50 |
| OZ287203.1      | 3        | 149.62      | 33.50 |
| OZ287204.1      | 4        | 145.17      | 33    |
| OZ287205.1      | 5        | 143.20      | 32.50 |
| OZ287206.1      | 6        | 132.08      | 33    |
| OZ287207.1      | 7        | 129.73      | 33    |
| OZ287208.1      | 8        | 119.32      | 33    |
| OZ287209.1      | 9        | 118.51      | 33    |
| OZ287210.1      | 10       | 116.54      | 33    |
| OZ287211.1      | 11       | 107.85      | 33    |
| OZ287212.1      | 12       | 102.38      | 33.50 |
| OZ287213.1      | 13       | 94.09       | 33    |

| INSDC accession | Molecule | Length (Mb) | GC%   |
|-----------------|----------|-------------|-------|
| OZ287214.1      | 14       | 93.29       | 32.50 |
| OZ287215.1      | 15       | 92.47       | 33    |
| OZ287216.1      | 16       | 90.60       | 33.50 |
| OZ287217.1      | 17       | 90.35       | 33.50 |
| OZ287218.1      | 18       | 87.93       | 34    |
| OZ287219.1      | 19       | 87.13       | 34    |
| OZ287220.1      | 20       | 86.29       | 34    |
| OZ287221.1      | 21       | 84.42       | 33.50 |
| OZ287222.1      | 22       | 82.67       | 32.50 |
| OZ287223.1      | 23       | 65.91       | 34    |
| OZ287224.1      | 24       | 64.04       | 33    |
| OZ287225.1      | 25       | 62.51       | 33.50 |
| OZ287226.1      | 26       | 57.99       | 33.50 |
| OZ287227.1      | 27       | 54.69       | 33    |
| OZ287228.1      | 28       | 53          | 34    |



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





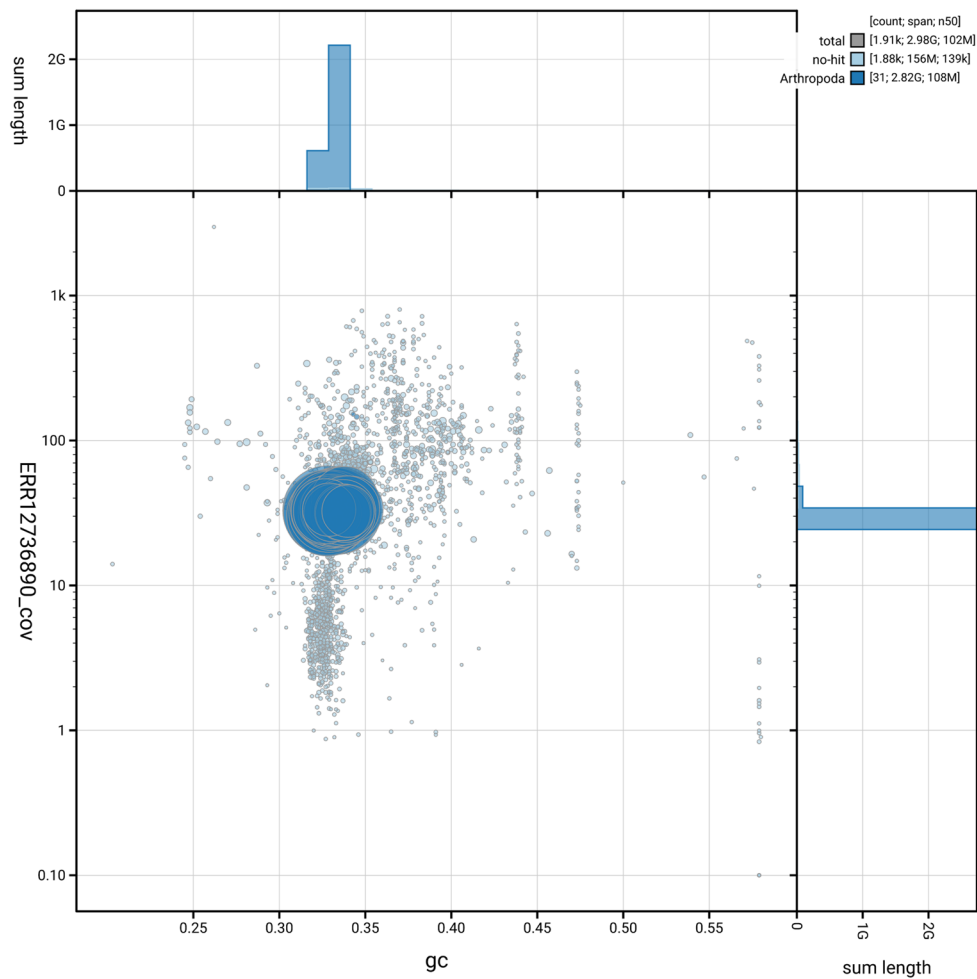
**Figure 5. Assembly metrics for qqAraAngu1.hap1.1.** The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost 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](#).

calculated for the haplotype 1, is **7.C.Q59**, 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 Darwin Tree of Life Partner. The submission

of materials by a Darwin Tree of Life Partner is subject to the **‘Darwin Tree of Life Project Sampling Code of Practice’**, which can be found in full on the [Darwin Tree of Life website](#). By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards



**Figure 6. BlobToolKit GC-coverage plot for qqAraAngu1.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 *Araneus angulatus* assembly.**

| Measure                                        | Value                                                      | Benchmark        |
|------------------------------------------------|------------------------------------------------------------|------------------|
| EBP summary (haplotype 1)                      | 7.C.Q59                                                    | 6.C.Q40          |
| Contig N50 length                              | 10.18 Mb                                                   | ≥ 1 Mb           |
| Scaffold N50 length                            | 102.38 Mb                                                  | = chromosome N50 |
| Consensus quality (QV)                         | Haplotype 1: 59.2; haplotype 2: 60.2; combined: 59.7       | ≥ 40             |
| <i>k</i> -mer completeness                     | Haplotype 1: 84.41%; Haplotype 2: 84.43%; combined: 99.13% | ≥ 95%            |
| BUSCO                                          | C:98.5% [S:92.3%; D:6.2%]; F:0.5%; M:1.0%; n:2 934         | S > 90%; D < 5%  |
| Percentage of assembly assigned to chromosomes | 94.74%                                                     | ≥ 90%            |

set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Araneus angulatus*. Accession number [PRJEB73665](#). The genome sequence is released openly for reuse. The *Araneus angulatus* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and

the Sanger Institute Tree of Life Programme (PRJEB43745). 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](#).

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

Author information

Contributors are listed at the following links:

- Members of the [Natural History Museum Genome Acquisition Lab](#)
- Members of the [Darwin Tree of Life Barcoding collective](#)
- Members of the [Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team](#)
- Members of [Wellcome Sanger Institute Scientific Operations – Sequencing Operations](#)
- Members of the [Wellcome Sanger Institute Tree of Life Core Informatics team](#)
- Members of the [Tree of Life Core Informatics collective](#)
- Members of the [Darwin Tree of Life Consortium](#)

Table 5. Software versions and sources.

| Software       | Version     | Source                                                                                                                       |
|----------------|-------------|------------------------------------------------------------------------------------------------------------------------------|
| BEDTools       | 2.30.0      | <a href="https://github.com/arq5x/bedtools2">https://github.com/arq5x/bedtools2</a>                                          |
| BLAST          | 2.14.0      | <a href="ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/">ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/<br/></a> |
| BlobToolKit    | 4.4.6       | <a href="https://github.com/blobtoolkit/blobtoolkit">https://github.com/blobtoolkit/blobtoolkit</a>                          |
| BUSCO          | 5.8.3       | <a href="https://gitlab.com/ezlab/busco">https://gitlab.com/ezlab/busco</a>                                                  |
| bwa-mem2       | 2.2.1       | <a href="https://github.com/bwa-mem2/bwa-mem2">https://github.com/bwa-mem2/bwa-mem2</a>                                      |
| Cooler         | 0.8.11      | <a href="https://github.com/open2c/cooler">https://github.com/open2c/cooler</a>                                              |
| DIAMOND        | 2.1.8       | <a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>                                      |
| fasta_windows  | 0.2.4       | <a href="https://github.com/tolkit/fasta_windows">https://github.com/tolkit/fasta_windows</a>                                |
| FastK          | 1.1         | <a href="https://github.com/thegenemyers/FASTK">https://github.com/thegenemyers/FASTK</a>                                    |
| GenomeScope2.0 | 2.0.1       | <a href="https://github.com/tbenavi1/genomescope2.0">https://github.com/tbenavi1/genomescope2.0</a>                          |
| Gfastats       | 1.3.6       | <a href="https://github.com/vgl-hub/gfastats">https://github.com/vgl-hub/gfastats</a>                                        |
| Hifiasm        | 0.19.8-r603 | <a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a>                                      |
| HiGlass        | 1.13.4      | <a href="https://github.com/higlass/higlass">https://github.com/higlass/higlass</a>                                          |
| MercuryFK      | 1.1.2       | <a href="https://github.com/thegenemyers/MERURY.FK">https://github.com/thegenemyers/MERURY.FK</a>                            |
| Minimap2       | 2.28-r1209  | <a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>                                                |

| Software                   | Version             | Source                                                                                                            |
|----------------------------|---------------------|-------------------------------------------------------------------------------------------------------------------|
| MitoHiFi                   | 3                   | <a href="https://github.com/marcelauliano/MitoHiFi">https://github.com/marcelauliano/MitoHiFi</a>                 |
| MultiQC                    | 1.14; 1.17 and 1.18 | <a href="https://github.com/MultiQC/MultiQC">https://github.com/MultiQC/MultiQC</a>                               |
| Nextflow                   | 24.10.4             | <a href="https://github.com/nextflow-io/nextflow">https://github.com/nextflow-io/nextflow</a>                     |
| PretextViewSnapshot        | N/A                 | <a href="https://github.com/sanger-tol/PretextViewSnapshot">https://github.com/sanger-tol/PretextViewSnapshot</a> |
| PretextViewView            | 1.0.3               | <a href="https://github.com/sanger-tol/PretextViewView">https://github.com/sanger-tol/PretextViewView</a>         |
| samtools                   | 1.21                | <a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>                           |
| sanger-tol/ascc            | 0.1.0               | <a href="https://github.com/sanger-tol/ascc">https://github.com/sanger-tol/ascc</a>                               |
| sanger-tol/blobtoolkit     | v0.8.0              | <a href="https://github.com/sanger-tol/blobtoolkit">https://github.com/sanger-tol/blobtoolkit</a>                 |
| sanger-tol/curationpretext | 1.4.2               | <a href="https://github.com/sanger-tol/curationpretext">https://github.com/sanger-tol/curationpretext</a>         |
| Seqtk                      | 1.3                 | <a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>                                           |
| Singularity                | 3.9.0               | <a href="https://github.com/sylabs/singularity">https://github.com/sylabs/singularity</a>                         |
| TreeVal                    | 1.4.0               | <a href="https://github.com/sanger-tol/treeval">https://github.com/sanger-tol/treeval</a>                         |
| YaHS                       | 1.2a.2              | <a href="https://github.com/c-zhou/yahs">https://github.com/c-zhou/yahs</a>                                       |

References

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](#)

Bee L, Oxford G, Smith H: **Britain's spiders: a field guide.** WILDGuides, Princeton University Press, 2017.  
[Reference Source](#)

Bee L, Oxford G, Smith H: **Britain's spiders: a field guide - fully revised and updated second edition.** WILDGuides, Princeton University Press, 2020.  
[Reference Source](#)

Breitling R, Bauer T, Schäfer M, *et al.*: **Phantom spiders 2: more notes on dubious spider species from Europe.** *Arachnologische Mitteilungen.* 2016; **52**: 50–77.  
[Publisher Full Text](#)

British Arachnological Society: **Summary for *Araneus angulatus* (Araneae) in the Spider and Harvestman Recording Scheme.** 2025.  
[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, 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](#)

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](#)

Dawson IK, Harvey PR, Merrett P, *et al.*: **A review of the scarce and threatened spiders (Arachnida: Araneae) of Great Britain.**

Dimitrov D, Benavides LR, Arnedo MA, *et al.*: **Rounding up the usual suspects: a standard target-gene approach for resolving the interfamilial phylogenetic relationships of ecribellate orb-weaving spiders with a new family-rank classification (Araneae, Araneoidea).** *Cladistics.* 2017; **33**(3): 221–250.  
[PubMed Abstract](#) | [Publisher 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](#)

Gärdenfors U: **Rödlistade Arter i Sverige 2000 - the 2000 Red List of Swedish Species.** Uppsala: ArtDatabanken, SLU, 2000.  
[Reference Source](#)

Grasshoff M: **Morphologische Kriterien als Ausdruck von artgrenzen bei radnetzspinnen der subfamilie Araneinae (Arachnida: Araneae: Araneidae).** *Abhandlungen der Senckenbergischen Naturforschenden Gesellschaft.* 1968; **516**: 1–100.  
[Reference Source](#)

Harvey P, Davidson M, Dawson I, *et al.*: **A review of the scarce and threatened spiders (Araneae) of Great Britain: Species Status no. 22.** 2017; 101.  
[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](#)

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](#)

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](#)

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](#)

Nentwig W, Blick T, Bosmans R, *et al.*: **Spiders of Europe, version 9.2025:**

*Araneus angulatus* (Clerck, 1757). 2025a.

[Publisher Full Text](#)

Nentwig W, Blick T, Bosmans R, *et al.*: **Spiders of Europe, version 9.2025:**

*Araneus pallidus* (Olivier, 1789). 2025b.

[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](#)

Ratnasingham S, Wei C, Chan D, *et al.*: **BOLD v4: a centralized bioinformatics platform for DNA-based biodiversity data**. In: *Methods Mol Biol*. New York, NY: Springer US, 2024; **2744**: 403–441.

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

Roberts MJ: **Collins Field Guide. Spiders of Britain and Northern Europe**. HarperCollins Publishers, 1996.

Scharff N, Coddington JA: **A phylogenetic analysis of the orb-weaving spider family Araneidae (Arachnida, Araneae)**. *Zoological Journal of the Linnean Society*. 1997; **120**(4): 355–434.

[Publisher Full Text](#)

Scharff N, Coddington JA, Blackledge TA, *et al.*: **Phylogeny of the orb-weaving spider family Araneidae (Araneae: Araneoidea)**. *Cladistics*. 2020; **36**(1): 1–21.

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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: a comprehensive update on curation, resources and tools**. *Database (Oxford)*. 2020; **2020**: baaa062.

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

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

Wang ZL, Wang ZY, Huang J, *et al.*: **The complete mitochondrial genome of an orb-weaver spider *Araneus angulatus* (Araneae: Araneidae)**. *Mitochondrial DNA B Resour*. 2019; **4**(2): 3870–3871.

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

Wheeler WC, Coddington JA, Crowley LM, *et al.*: **The spider Tree of Life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling**. *Cladistics*. 2017; **33**(6): 574–616.

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

World Spider Catalogue: **World Spider Catalogue: *Araneus angulatus***. 2025.

[Reference Source](#)

Wunderlich J: **Contribution to the spider (Araneida) fauna of the Algarve, Portugal, with notes on the phylogeny of the superfamily Araneoidea and the description of the new araneoid family Fonteferridae**. *Beiträge zur Araneologie*. 2023; **16**: 4–112, 231–33.

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 07 January 2026

<https://doi.org/10.21956/wellcomeopenres.27772.r143716>

© 2026 Fromm B. 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.



**Bastian Fromm** 

UiT- The Arctic University of Norway, Tromsø, Norway

Not sure if this is a fully automatically generated article, as I have received another review request for a spider genome for Wellcome Open Research that is written exactly the same way (and, hence, will get the same review).

The report is state-of-the-art report of genome assembly techniques and statistics of a spider (*Araneus angulatus*).

I am personally surprised by the lack of a genome annotation and would like to ask the authors for a statement why it was not included.

I like the relatively extensive introduction to the species.

**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:** non-coding genes

**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 December 2025

<https://doi.org/10.21956/wellcomeopenres.27772.r142355>

© 2025 Akpınar A. 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.



**Adile Akpınar**

Gaziantep University, Gaziantep, Turkey

Morphological data for the species *Araneus angulatus* may be insufficient for species identification and may lead to synonymy. This article presents a comprehensive genomic study that details all molecular data for *Araneus angulatus*. The results of this article serve as a reference for future studies. There are many closely related species among spiders, which can lead to misidentification of species and synonymy.

**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:** Araneae taxonomy

**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.**

---