Darwin Tree of Life Barcoding collective, , Crowley, LM, Wilding, CS, University of Oxford and Wytham Woods Genome Acquisition Lab, , Darwin Tree of Life Consortium, , Wellcome Sanger Institute Tree of Life Management, Samples and L, , Wellcome Sanger Institute Scientific Operations: Sequencing Oper, , Wellcome Sanger Institute Tree of Life Core Informatics team, and Tree of Life Core Informatics collective,

The genome sequence of an orbweaving spider, Gibbaranea gibbosa (Walckenaer, 1802) [version 1; peer review: 2 approved, 1 approved with reservations]

Article

DATA NOTE

The genome sequence of an orbweaving spider, *Gibbaranea gibbosa* (Walckenaer, 1802)

[version 1; peer review: 2 approved, 1 approved with reservations]

Liam M. Crowley[1], Craig S Wilding[2],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]University of Oxford, Oxford, England, UK
[2]School of Biology and Environmental Science, University College Dublin, Dublin, Leinster, Ireland

v1 **First published:** 24 Feb 2025, **10**:97
https://doi.org/10.12688/wellcomeopenres.23751.1
Latest published: 24 Feb 2025, **10**:97
https://doi.org/10.12688/wellcomeopenres.23751.1

Abstract

We present a genome assembly from a specimen of *Gibbaranea gibbosa* (orbweaving spider; Arthropoda; Arachnida; Araneae; Araneidae). The genome sequence has a total length of 2,816.88 megabases. Most of the assembly (98.61%) is scaffolded into 13 chromosomal pseudomolecules, including the X_1 and X_2 sex chromosomes. The mitochondrial genome has also been assembled and is 14.1 kilobases in length.

Keywords

Gibbaranea gibbosa, orbweaving spider, genome sequence, chromosomal, Araneae

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ? ✓

	1	2	3
version 1 24 Feb 2025	✓ view	? view	✓ view

1. **Peng-Yu Jin**, University of the Chinese Academy of Sciences, Beijing, China
2. **Guilherme Gainett**, Boston Children's Hospital F M Kirby Neurobiology Center, Boston, USA
3. **Jessica E Garb**, University of Massachusetts Lowell,, Lowell, MA, USA

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

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

Author roles: 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.

How to cite this article: 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 Research 2025, **10**:97 https://doi.org/10.12688/wellcomeopenres.23751.1

First published: 24 Feb 2025, **10**:97 https://doi.org/10.12688/wellcomeopenres.23751.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Chelicerata; Arachnida; Araneae; Araneomorphae; Entelegynae; Orbiculariae; Araneoidea; Araneidae; *Gibbaranea*; *Gibbaranea gibbosa* (Walckenaer, 1802) (NCBI:txid1907000)

Background

Gibbaranea gibbosa (Walckenaer, 1802) is a medium-sized orbweaving spider (Araneidae) with adult females ranging from 5–7 mm in length and males from 4–5 mm (Locket & Millidge, 1951; Roberts, 1987). Both sexes are adult from early to midsummer (Harvey *et al.*, 2002). It is eminently recognisable through its distinctive carapace, decorated with two prominent tubercles anterolaterally on the dorsal side of the abdomen (Roberts, 1987) and is unlikely to be confused with other species except in the immature stages when it can superficially resemble *Araneus angulatus* (Bee *et al.*, 2020).

Whilst previously described as *Araneus gibbosa*, Archer (1951) split the genus *Gibbaranea* from *Araneus* Clerck, 1757. Four species of *Gibbaranea* occur in continental Europe – *G. gibbosa*, *G. bituberculata* (Walckenaer, 1802), *G. omoeda* (Thorell, 1870), and *G. ullrichi* (Hahn, 1835) (van Helsdingen, 1996a), though only *G. bituberculata* and *G. gibbosa* have been recorded from the UK and only *G. gibbosa* from Ireland. However, *G. bituberculata* was found in only a single UK locality and with no sightings since 1950 it is potentially now locally extinct (Harvey *et al.*, 2002), making *G. gibbosa* the sole representative of the genus in Britain and Ireland.

This species can be difficult to detect due to its colouration and habitat, being typically drably coloured. In northern Europe, including the UK and Ireland, the opisthoma is typically greenish in colour (Bee *et al.*, 2020; Lissner & Bosmans, 2016), though in southern Europe they are typically less green (Lissner & Bosmans, 2016) with more variable colouration from light brown to greenish grey (Nentwig *et al.*, 2024). There is an obvious folium (the broad leaf-like marking on the dorsal abdomen midline), sometimes with a paler margin (Nentwig *et al.*, 2024). The carapace is brown and legs brown and annulated. Its preferred habitat is often high in trees, though it occurs also on shrubs, hedgerows and scrub (Harvey *et al.*, 2002). Here it spins strong, dense, inconspicuous orb-webs, often in the tree crown.

It is common throughout continental Europe into western Asia occurring as far east as Iran (Zamani *et al.*, 2022a) across Europe to the Iberian Peninsula (Branco *et al.*, 2019), and as far north as Finland (Zamani *et al.*, 2022b). In the UK it is widespread in the south and east but becomes more rare further west and north (https://srs.britishspiders.org.uk/portal.php/p/Summary/s/Gibbaranea+gibbosa). On the island of Ireland there are species records from 11 counties, from Cork in the south to Antrim in the north (van Helsdingen, 1996b) although it likely to be found elsewhere.

Here we present a chromosomally complete genome sequence for *G. gibbosa* based on an individual collected from Wytham woods, Oxfordshire. Spider genomes provide important insights into venomics, evo-devo, silk production and provide additional data to aid in resolution of phylogenetic difficulties (Garb *et al.*, 2018). This additional genome will aid such efforts.

Genome sequence report

Sequencing data

The genome of a specimen of *Gibbaranea gibbosa* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 122.54 Gb from 14.08 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 2,835.31 Mb, with a heterozygosity of 1.11% and repeat content of 37.00%. These values provide an initial assessment of genome complexity and the challenges anticipated during assembly. Based on this estimated genome size, the sequencing data provided approximately 42.0x coverage of the genome. Chromosome conformation Hi-C sequencing produced 240.61 Gb from 1,593.46 million reads.

Table 1 summarises the specimen and sequencing information, including the BioProject, study name, BioSample numbers, and sequencing data for each technology.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The assembly was improved by manual curation, which corrected 77 misjoins or missing joins and removed 16 haplotypic duplications. These interventions reduced the total assembly length by 0.72%, decreased the scaffold count by 9.88%, and increased the scaffold N50 by 5.26%. The final assembly has a total length of 2,816.88 Mb in 227 scaffolds, with 208 gaps, and a scaffold N50 of 226.11 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.

Figure 1. Photograph of the *Gibbaranea gibbosa* (qqGibGibb2) specimen used for genome sequencing.

Table 1. **Specimen and sequencing data for *Gibbaranea gibbosa*.**

Project information			
Study title	Gibbaranea gibbosa		
Umbrella BioProject	PRJEB73645		
Species	*Gibbaranea gibbosa*		
BioSpecimen	SAMEA10166988		
NCBI taxonomy ID	1907000		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	qqGibGibb2	SAMEA10201191	cephalothorax
Hi-C sequencing	qqGibGibb1	SAMEA10200839	Whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR12737269	1.59e+09	240.61
PacBio Revio	ERR12736872	9.41e+06	72.19
PacBio Sequel IIe	ERR12736873	2.28e+06	22.74
PacBio Sequel IIe	ERR12736871	2.38e+06	27.61

Most of the assembly sequence (98.51%) was assigned to 13 chromosomal-level scaffolds, representing 11 autosomes and the X_1 and X_2 sex chromosome. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, chromosomes X_1 and X_2 were assigned based on Hi-C signal.

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 in GenBank.

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 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.

The primary haplotype has a QV of 66.8, and the combined primary and alternate assemblies achieve an estimated QV of 65.1. The *k*-mer completeness for the primary haplotype is 77.86%, and for the alternate haplotype it is 75.90%. The combined primary and alternate assemblies achieve a *k*-mer completeness of 99.47%. BUSCO analysis using the arachnida_odb10 reference set (n = 2,934) indicated a completeness score of 98.3% (single = 92.4%, duplicated = 5.9%).

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The achieves the EBP reference standard of 7.C.Q66.

Methods

Sample acquisition and DNA barcoding

An adult female *Gibbaranea gibbosa* (specimen ID Ox001412, ToLID qqGibGibb2) was collected from Wytham Woods, Oxfordshire, United Kingdom (latitude 51.77, longitude -1.33) on 2021-06-03 by potting. The specimen used for Hi-C sequencing (specimen ID Ox001199, ToLID qqGibGibb1) was collected from the same location on 2021-04-13 by potting. Both specimens were collected and identified by Liam Crowley (University of Oxford) and 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

Table 2. Genome assembly data for *Gibbaranea gibbosa*.

Genome assembly		
Assembly name	qqGibGibb2.1	
Assembly accession	GCA_964059485.1	
Alternate haplotype accession	*GCA_964059525.1*	
Assembly level for primary assembly	chromosome	
Span (Mb)	2,816.88	
Number of contigs	435	
Number of scaffolds	227	
Longest scaffold (Mb)	435.79	
Assembly metrics	**Measure**	***Benchmark***
Contig N50 length	50.82 Mb	*≥ 1 Mb*
Scaffold N50 length	226.11 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 66.8; alternate: 63.9; combined 65.1	*≥ 40*
k-mer completeness	Primary: 77.86%; alternate: 75.90%; combined: 99.47%	*≥ 95%*
BUSCO*	C:98.3%[S:92.4%,D:5.9%], F:0.6%,M:1.1%,n:2,934	*S > 90%, D < 5%*
Percentage of assembly mapped to chromosomes	98.51%	*≥ 90%*
Sex chromosomes	X_1 and X_2	*localised homologous pairs*
Organelles	Mitochondrial genome: 14.1 kb	*complete single alleles*

* 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.

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. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The qqGibGibb2 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). 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

Tissue from the whole organism of the qqGibGibb1 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, 20–50 mg of frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration. After crosslinking, the tissue was homogenised using the Diagnocine

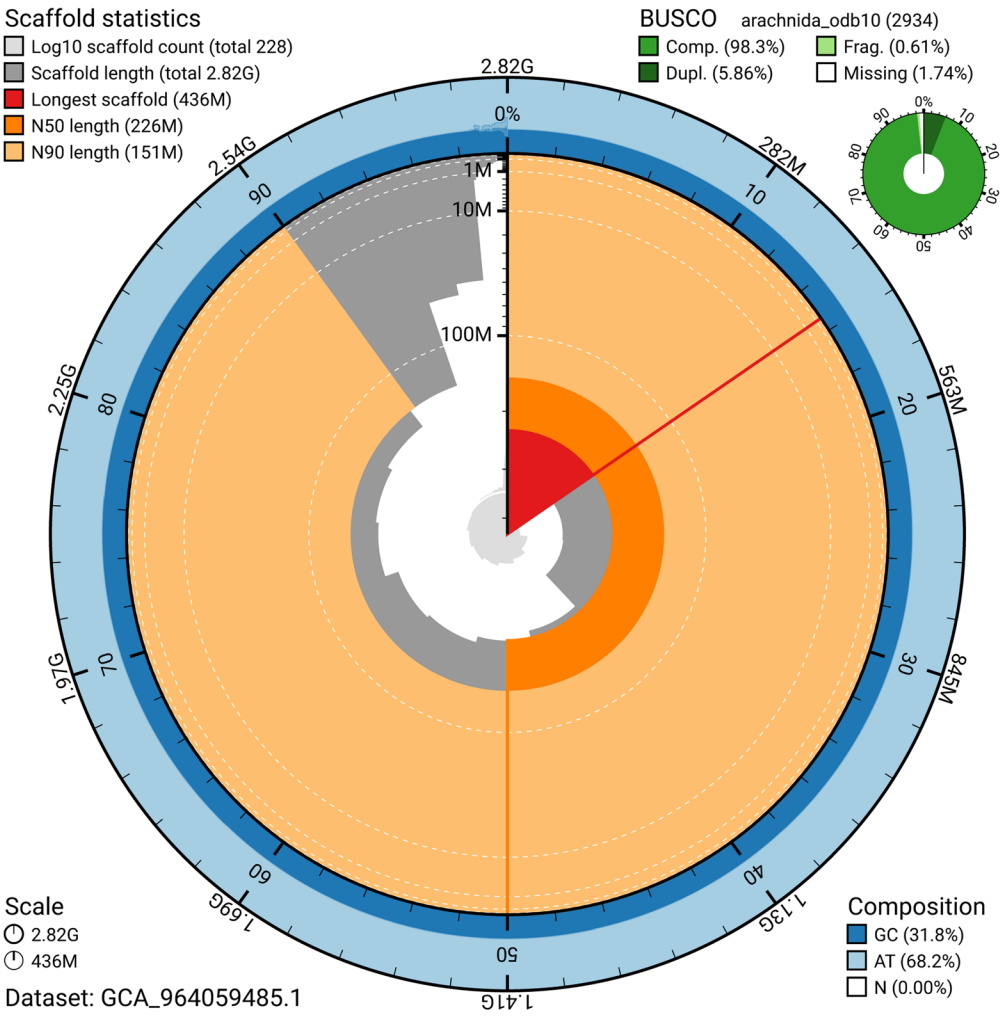

Figure 2. Genome assembly of *Gibbaranea gibbosa*, qqGibGibb2.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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_964059485.1/dataset/GCA_964059485.1/snail.

Power Masher-II and BioMasher-II tubes and pestles. Following the Arima-HiC v2 kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation. Additionally, the biotinylation percentage was estimated using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and Arima-HiC v2 QC beads.

Library preparation and sequencing

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

PacBio HiFi

At a minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California,

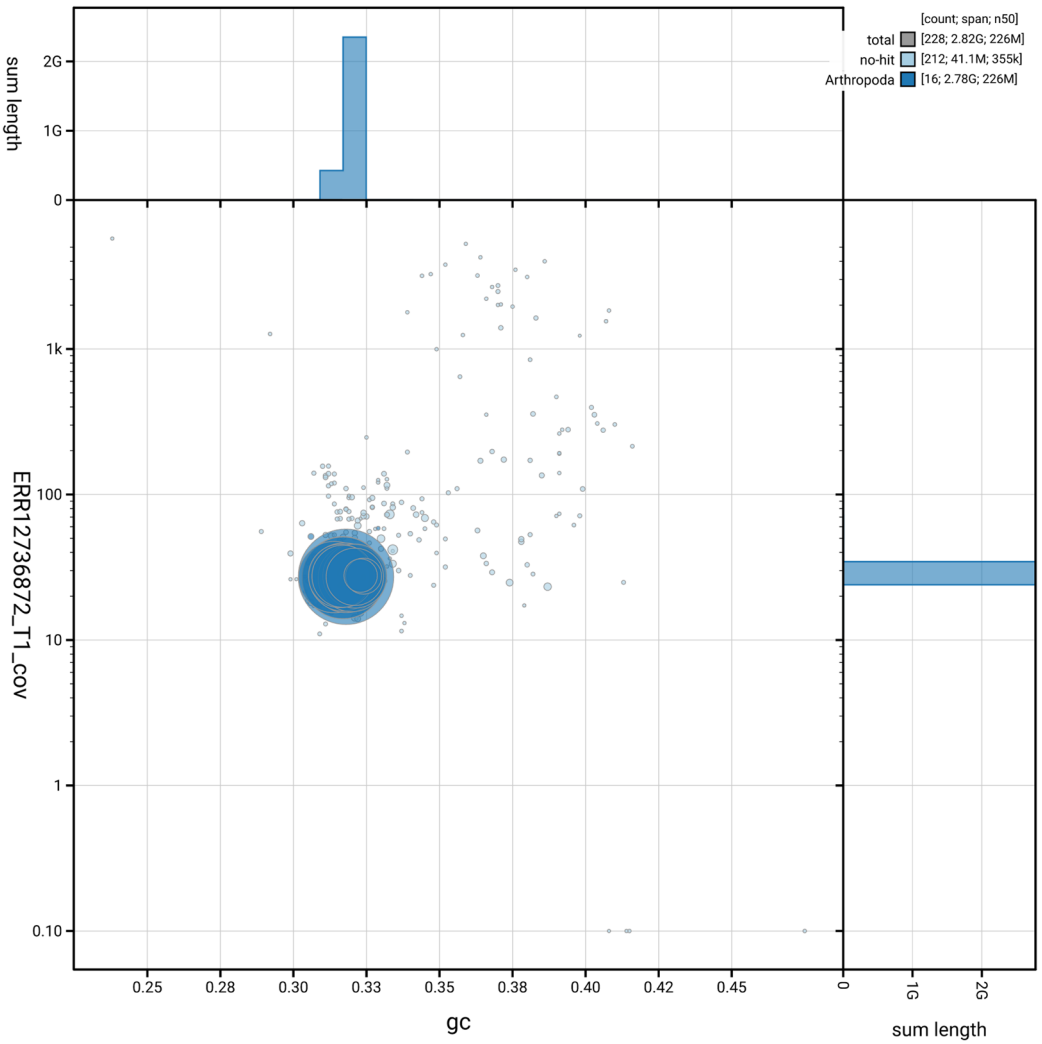

Figure 3. Genome assembly of *Gibbaranea gibbosa*, qqGibGibb2.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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_964059485.1/blob.

USA) as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and gDNA 55kb BAC analysis kit.

Samples were 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, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using

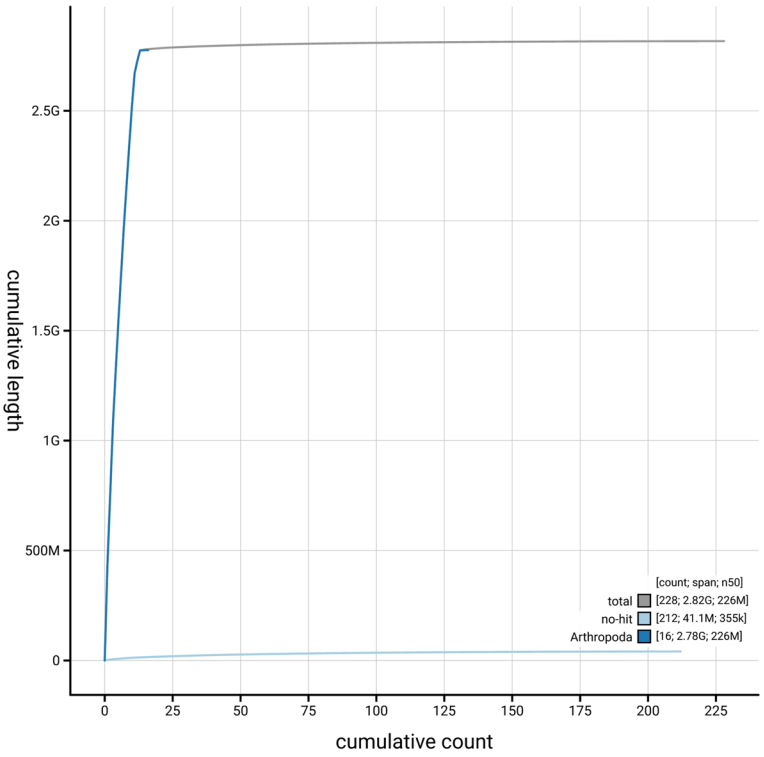

Figure 4. Genome assembly of *Gibbaranea gibbosa,* qqGibGibb2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964059485.1/dataset/GCA_964059485.1/cumulative.

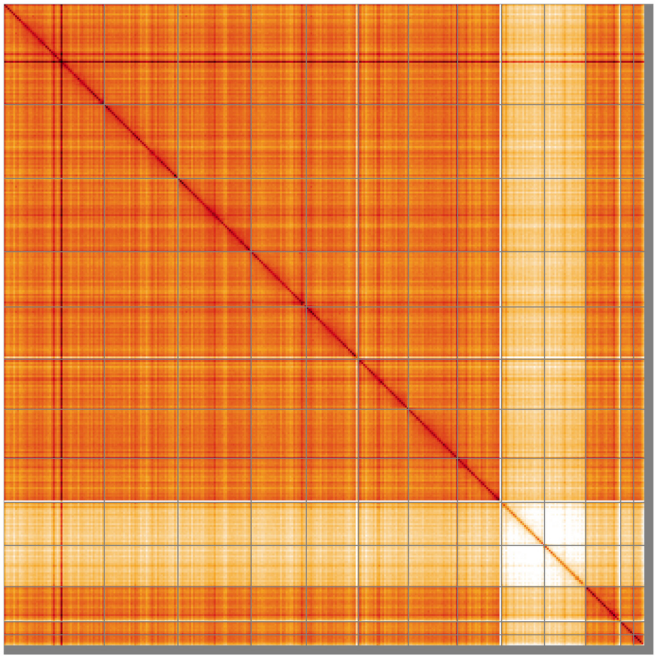

Figure 5. Genome assembly of *Gibbaranea gibbosa:* Hi-C contact map of the qqGibGibb2.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=K4Y5RpnLTA-UQCWSLaiYsQ.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Gibbaranea gibbosa*, qqGibGibb2.

INSDC accession	Name	Length (Mb)	GC%
OZ060745.1	1	435.79	32
OZ060746.1	2	318.48	31.5
OZ060747.1	3	316.74	31.5
OZ060748.1	4	238.9	31.5
OZ060749.1	5	226.11	32
OZ060750.1	6	216.38	31.5
OZ060751.1	7	211.95	32
OZ060752.1	8	190.43	32
OZ060755.1	9	150.93	32
OZ060756.1	10	56.36	32.5
OZ060757.1	11	47.6	32.5
OZ060758.1	MT	0.01	24
OZ060753.1	X1	186.29	31.5
OZ060754.1	X2	178.82	31.5

SPRISelect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, a-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required, determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

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 first assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) 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 (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. Sex chromosomes were identified by Hi-C data coverage. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

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 the primary and alternate haplotypes using the *k*-mer databases ($k = 31$) that were computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these 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), relying on the Conda package manager, the Bioconda initiative

(Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

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. 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

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
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	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
Merqury.FK	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
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	0.25	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc

Software tool	Version	Source
sanger-tol/blobtoolkit	0.5.1	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

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: Gibbaranea gibbosa. Accession number PRJEB73645; https://identifiers.org/ena.embl/PRJEB73645. The genome sequence is released openly for reuse. The *Gibbaranea gibbosa* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. 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

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Archer AF: **Studies in the orbweaving spiders (Argiopidae). 2.** *Am Mus Novit.* 1951; **1502**: 1–34.
Reference Source

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the universal protein knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 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: a field guide.** Second, Princeton University Press, 2020.
Reference Source

Branco VV, Morano E, Cardoso P: **An update to the Iberian spider checklist (Araneae).** *Zootaxa.* 2019; **4614**(2): zootaxa.4614.2.1.
PubMed Abstract | 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

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

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
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

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

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: **Provisional atlas of British spiders (Arachnida, Araneae), Volume 2.** Abbots Ripton, Huntingdon: Biological Records Centre, Centre for Ecology and Hydrology, 2002.
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Lissner J, Bosmans R: **Description of a new *Gibbaranea* (Araneae: Araneidae) from the Western Mediterranean.** *Arachnology Letters.* 2016; **52**: 25–30.
Publisher Full Text

Locket GH, Millidge AF: **British Spiders.** The Ray Society, 1951.
Reference Source

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

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

Nentwig W, Blick T, Bosmans R, *et al.*: **Spiders of Europe.** Version 11.2024, 2024.
Publisher Full Text

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

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

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

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

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

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: ***Gibbaranea ullrichi* (Hahn, 1835) (Araneae, Araneidae) in the Netherlands.** *Nieuwsbrief SPINED.* 1996a; **29**: 6–9.
Reference Source

van Helsdingen PJ: **The county distribution of Irish spiders, incorporating a revised catalogue of the species.** *Irish Naturalist's Journal.* 1996b; **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

Zamani A, Nadolny AA, Esyunin SL, *et al.*: **New data on the spider fauna of Iran (Arachnida: Araneae), Part IX.** *Arachnology.* 2022a; **19**(sp1): 358–384.
Publisher Full Text

Zamani A, Österman EM, Koponen S: **The first record of *Gibbaranea gibbosa* (Walckenaer, 1802) (Araneae: Araneidae) in Finland.** *Sahlbergia.* 2022b; **28**(1): 23–24.
Reference Source

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 28 March 2025

https://doi.org/10.21956/wellcomeopenres.26199.r119512

✔ **Jessica E Garb**
University of Massachusetts Lowell,, Lowell, MA, USA

This data note reports a chromosome level genome for the spider *Gibbaranea gibbosa*, the only extant representative of this genus in Britain and Ireland. The work follows the protocols of the Wellcome Sanger Trust Darwin Tree of Life project and appears thorough in terms of producing a high-quality genome assembly. This is a nice contribution to the growing genomic resources for spiders.

I have a few comments related to issues I am interested in and perhaps things for the authors to think about going forward.

1. It would be worth collecting at least one more individual, if not an additional male and female from the same locality as the specimens used for the genome sequencing to be deposited in a museum as a morphological voucher. That may become important if there is a need to document some aspect of phenotype in relation to the genome.

2. The use of mitochondrial genotyping of the sequenced specimen and comparing to the BOLD database makes sense for confirming the species identity, but the database is only as reliable as the original input data and how close the match is. The authors could briefly comment on how good the match of their sequenced specimen's barcode is to those that it matches in the database with this species name. The paper only says the species was confirmed but not the level of similarity.

3. The paper states the genome will be annotated using available RNA-Seq data and presented through the Ensemble pipeline at the European Bioinformatics Institute. It is unclear if the available RNA-Seq data will be from other spider species or from this species (*Gibbaranea gibbosa*). It would be ideal and seemingly not much more effort to include additional specimens from the same population for RNA-Seq for annotation. While an annotation based on other spider species RNA/protein sequences may be reasonable and probably could yield good BUSCO scores, most likely it will miss out on finding and/or properly annotating the more interesting and rapidly evolving genes like those involved in venom

and silk, which are what many people in the community would be interested in.

4. The authors say that for the Pacbio HiFi sequencing that samples were required to have an average fragment size greater than 8kb and a total mass over 400 ng to proceed to the low input smrt bell prep kit 3.0 protocol. But then as far as I can tell they never say how much DNA they were able to obtain from the single specimen and whether they definitely did use the low input protocol – it is implied they did. I would like to know the amount of DNA obtained from their specimen before and after size selection.

5. Can the authors confirm whether or not any PCR was involved in generating the primary PacBio sequence data? That would be helpful for me to know.

6. Like one of the other reviewers, I'd also like to read a few more details on how the HiC data provides signal to infer which are the sex chromosomes.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Evolutionary Biology, Spider Biology, Silk and Venom Protein evolution

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 18 March 2025

https://doi.org/10.21956/wellcomeopenres.26199.r120108

?

Guilherme Gainett
Boston Children's Hospital F M Kirby Neurobiology Center, Boston, Massachusetts, USA

This paper provides a chromossome-level genome assembly for a orb-weaving spider, as part of

the Darwin Tree of Life initiative. The genome sequencing was done with state-of-the-art PacBio long-read sequencing and Hi-C scaffolding.

The description is clear and the genome quality measures are high, such as contiguity and completeness (ex: BUSCO).

My one general concern is the HiC data, which does not seem to have resolution to show the chromosomes. It was not clear to me how much the HiC really contributed to the scaffolding of the genome. I think this point needs to be clarified in the text.

Besides this point, I also suggest providing additional details about how the sex chromosomes were inferred. There are also small nomenclature suggestions and minor comments in the **attached pdf**.

This paper provides a great quality resource for the community.

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: Arachnology; Developmental Biology; 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, however I have significant reservations, as outlined above.

Reviewer Report 06 March 2025

https://doi.org/10.21956/wellcomeopenres.26199.r119516

Peng-Yu Jin
University of the Chinese Academy of Sciences, Beijing, Beijing, China

This article is overall excellent, with a well-executed analytical approach and solid data. Notably, the use of PacBio and Hi-C sequencing on a single tiny individual provides valuable insights for studying other small invertebrates of similar size. Incorporating gene annotation files such as GFF would further enhance the study's impact, particularly for research on spider traits like toxin, silk and behaviors. The Hi-C contact map of the chromosomes appears less distinct, as clear block patterns between different chromosomes are not readily visible. It would be helpful if the study could provide a more detailed Hi-C interaction map or discuss potential factors that might have influenced the resolution. Has a karyotype analysis been conducted for this species? If so, incorporating relevant information could further support the chromosome assembly.

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: Single cell RNA sequencing, genome evolution, and nervous system evolution of spider

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.