DATA NOTE

The genome sequence of the solitary wasp, *Cerceris ruficornis* (Fabricius, 1793)

[version 1; peer review: 4 approved, 2 approved with reservations]

Olga Sivell[1], Clare Boyes[2], 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

[1]Natural History Museum, London, England, UK
[2]Independent researcher, Welshpool, Wales, UK

v1 **First published:** 27 May 2025, **10**:276
https://doi.org/10.12688/wellcomeopenres.24270.1
Latest published: 27 May 2025, **10**:276
https://doi.org/10.12688/wellcomeopenres.24270.1

Abstract

We present a genome assembly from a female specimen of *Cerceris ruficornis* (solitary wasp; Arthropoda; Insecta; Hymenoptera; Crabronidae). The genome sequence has a total length of 566.08 megabases, of which 65.35% is scaffolded into 14 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 18.07 kilobases. Gene annotation of this assembly on Ensembl identified 11,093 protein-coding genes.

Keywords

Cerceris ruficornis, solitary wasp, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ✓ ? ✓ ✓ ✓

	1	2	3	4	5	6
version 1 27 May 2025	? view	✓ view	? view	✓ view	✓ view	✓ view

1. **Guanliang Meng**, Leibniz Institute, Berlin, Germany
2. **Ariel Chipman**, Hebrew University of Jerusalem School of Pharmacy, Jerusalem, Israel
3. **Sarah Inwood**, University of Otago, Dunedin, New Zealand
4. **Alexandra Cerqueira De Araujo**, University College London, London, UK
5. **Sara E Miller**, Cornell University, Missouri, USA
University of Missouri St. Louis, Saint Louis, USA

6. **Kashif Zahoor**, Government College University Faisalabad, Faisalabad, Pakistan

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: Sivell O: Investigation, Resources; **Boyes C**: Investigation, Resources, 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: Sivell O, Boyes C, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the solitary wasp, *Cerceris ruficornis* (Fabricius, 1793) [version 1; peer review: 4 approved, 2 approved with reservations]** Wellcome Open Research 2025, **10**:276 https://doi.org/10.12688/wellcomeopenres.24270.1

First published: 27 May 2025, **10**:276 https://doi.org/10.12688/wellcomeopenres.24270.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Crabronidae; Philanthinae; Cercerini; *Cerceris*; *Cerceris ruficornis* (Fabricius, 1793) (NCBI:txid2494881)

Background

Cerceris ruficornis is a solitary wasp in the family Crabronidae which is found in scattered locations in Europe, with a concentration of records in Fennoscandia and Denmark (GBIF Secretariat, 2023). In the UK it is restricted to the south of England and Wales (BWARS, 2025). This black and yellow wasp has red legs and a ribbed abdomen. It is on the wing between mid-June and September and occurs on heathlands (BWARS, 2025). There is very little information on the nesting biology. The prey is mainly *Curculionidae* but *Chrysomelidae* are also taken (Lomholdt, 1976).

The genome of *Cerceris ruficornis* 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. Here we present a chromosomally complete genome sequence for *Cerceris ruficornis* based on one female specimen from Buxton Heath, Norfolk, UK (Figure 1).

Genome sequence report

Sequencing data

The genome of a specimen of *Cerceris ruficornis* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 28.04 Gb (gigabases) from 2.58 million reads, which were used to assemble the genome. GenomeScope analysis estimated the haploid genome size at 454.59 Mb, with a heterozygosity of 0.81% and repeat content of 36.64%. These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 59x coverage. Hi-C sequencing produced 82.32 Gb from 545.14 million reads, used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Figure 1. Photograph of the *Cerceris ruficornis* (iyCerRufi2) specimen used for genome sequencing.

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 5 misjoins or missing joins. These interventions decreased the scaffold count by 0.71%, and increased the scaffold N50 by 13.55%. The final assembly has a total length of 566.08 Mb in 559 scaffolds, with 126 gaps, and a scaffold N50 of 20.57 Mb (Table 2).

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

Most of the assembly sequence (65.35%) was assigned to 14 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

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 combined primary and alternate assemblies achieve an estimated QV of 64.8. The *k*-mer completeness is 81.92% for the primary haplotype and 76.17% for the alternate haplotype; and 98.83% for the combined primary and alternate assemblies. BUSCO v.5.5.0 analysis using the hymenoptera_odb10 reference set (n = 5,991) identified 96.5% of the expected gene set (single = 96.3%, duplicated = 0.1%).

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The primary assembly achieves the EBP reference standard of **6.7.Q65**.

Genome annotation report

The *Cerceris ruficornis* genome assembly (GCA_963989415.1) was annotated externally by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 18,816 transcribed mRNAs from 11,093 protein-coding and 1,434 non-coding genes. The average transcript length is 11,558.03 bp. There are 1.50 coding transcripts per gene and 6.22 exons per

Table 1. **Specimen and sequencing data for *Cerceris ruficornis*.**

Project information			
Study title	Cerceris ruficornis		
Umbrella BioProject	PRJEB66742		
Species	*Cerceris ruficornis*		
BioSpecimen	SAMEA112964329		
NCBI taxonomy ID	2494881		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	iyCerRufi2	SAMEA112975494	whole organism
Hi-C sequencing	iyCerRufi2	SAMEA112975494	whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR12102416	5.45e+08	82.32
PacBio Sequel IIe	ERR12102451	2.58e+06	28.04

Table 2. **Genome assembly data for *Cerceris ruficornis*.**

Genome assembly		
Assembly name	iyCerRufi2.1	
Assembly accession	GCA_963989415.1	
Alternate haplotype accession	*GCA_963989395.1*	
Assembly level for primary assembly	chromosome	
Span (Mb)	566.08	
Number of contigs	685	
Number of scaffolds	559	
Longest scaffold (Mb)	66.28	
Assembly metric	**Measure**	***Benchmark***
Contig N50 length	3.2 Mb	*≥ 1 Mb*
Scaffold N50 length	20.57 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 65.4; alternate: 64.1; combined: 64.8	*≥ 40*
k-mer completeness	Primary: 81.92%; alternate: 76.17%; combined: 98.83%	*≥ 95%*
BUSCO*	C:96.5%[S:96.3%,D:0.1%], F:0.6%,M:2.9%,n:5,991	*S > 90%; D < 5%*
Percentage of assembly assigned to chromosomes	65.35%	*≥ 90%*
Sex chromosomes	None	*localised homologous pairs*
Organelles	Mitochondrial genome: 18.07 kb	*complete single alleles*

* BUSCO scores based on the hymenoptera_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

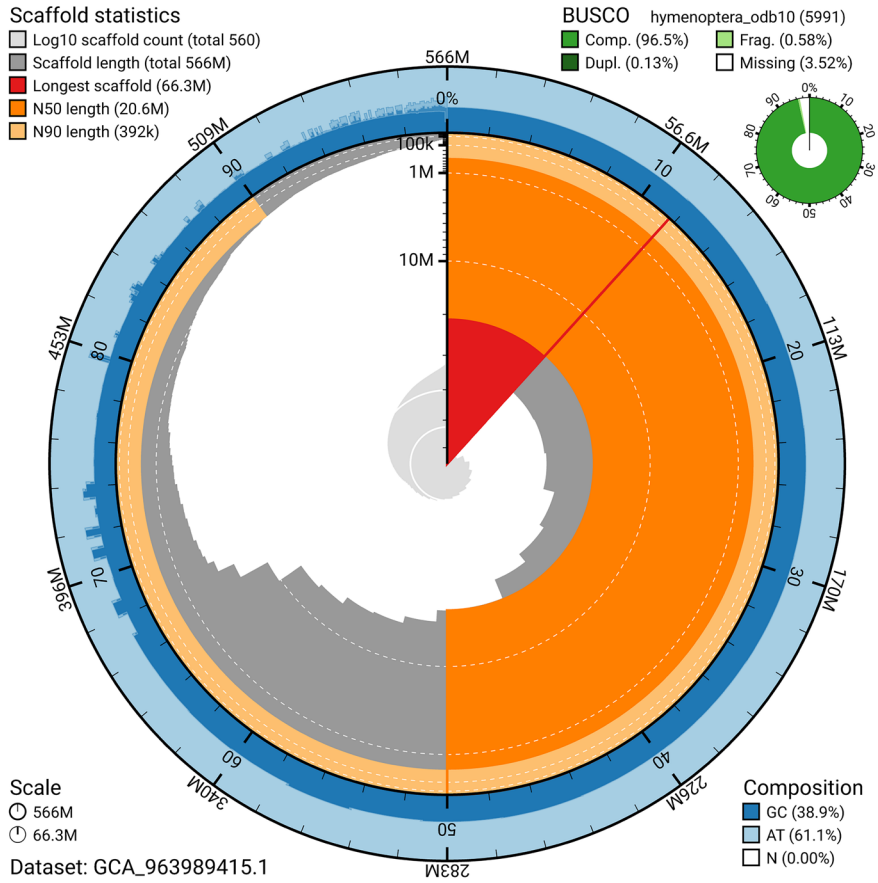

Figure 2. Genome assembly of *Cerceris ruficornis*, iyCerRufi2.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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 hymenoptera_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963989415.1/dataset/GCA_963989415.1/snail.

transcript. For further information about the annotation, please refer to https://beta.ensembl.org/species/060ab491-5335-4f09-95ba-32ffa082ec0b.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Cerceris ruficornis* (specimen ID NHMUK015059139, ToLID iyCerRufi2), collected from Buxton Heath, England, United Kingdom (latitude 52.75, longitude 1.22) on 2022-07-04. The specimen was collected by Olga Sivell (Natural History Museum) and identified by Clare Boyes (Dipterists Forum) and preserved by dry freezing (–80 °C).

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 identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io (Beasley *et al.*, 2023).

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

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification (Howard *et al.*, 2025). Detailed protocols are

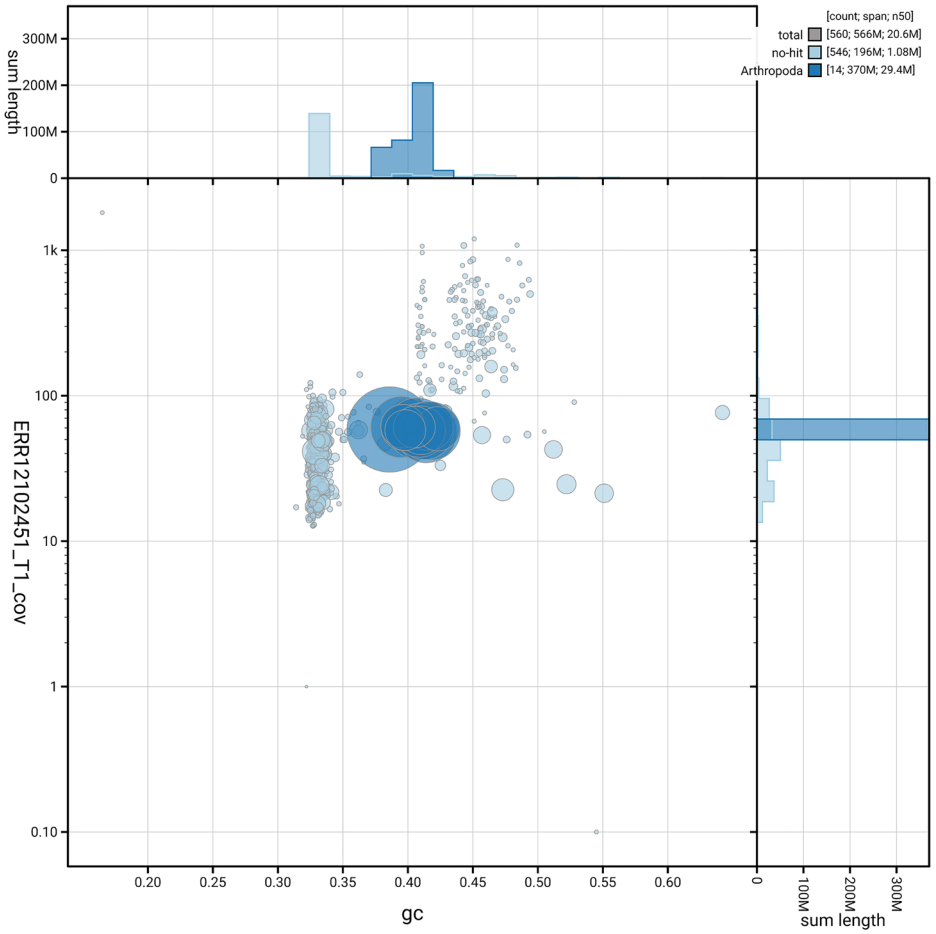

Figure 3. Genome assembly of *Cerceris ruficornis*, iyCerRufi2.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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_963989415.1/dataset/GCA_963989415.1/blob.

available on protocols.io (Denton *et al.*, 2023b). The iyCerRufi2 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the whole organism 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 and crosslinking

Hi-C data were generated from the whole organism of the iyCerRufi2 sample using the Arima-HiC v2 kit (Arima Genomics) with 20–50 mg of frozen tissue (stored at –80 °C). As per manufacturer's instructions, tissue was fixed, and the DNA crosslinked using a TC buffer with a final formaldehyde concentration of 2%. The tissue was then homogenised using the Diagnocine Power Masher-II. The crosslinked DNA was digested using a restriction enzyme master mix, then biotinylated and ligated. A clean up was performed with SPRIselect beads prior to library preparation. DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit, and sample biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Library preparation and sequencing

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

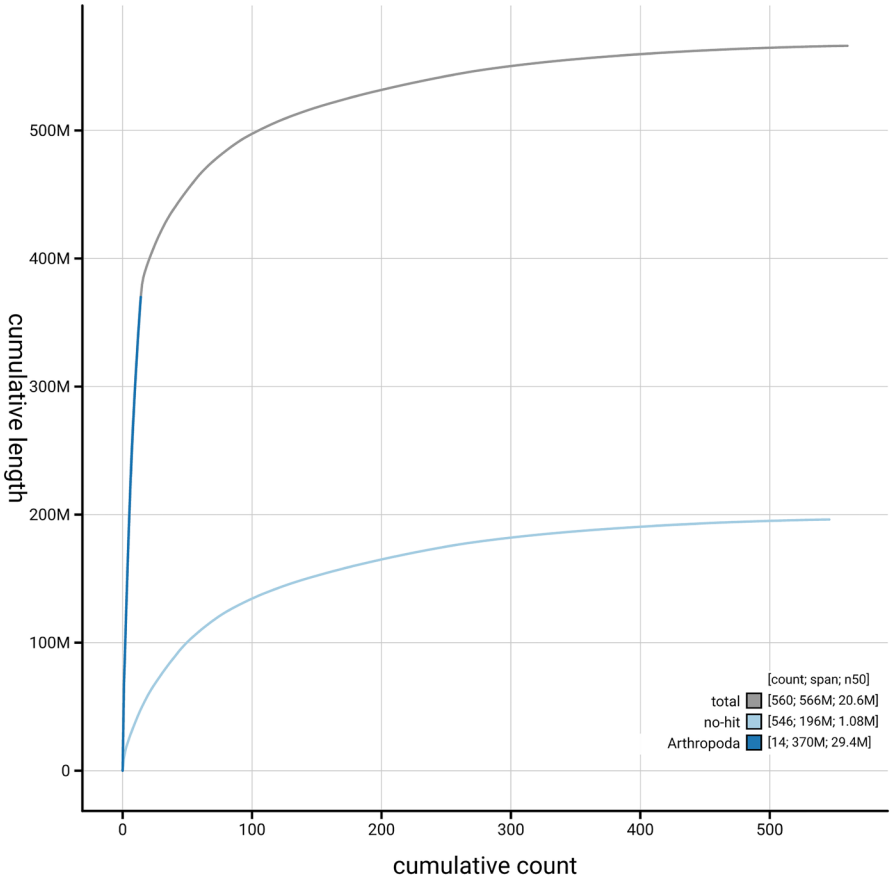

Figure 4. Genome assembly of *Cerceris ruficornis*, iyCerRufi2.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963989415.1/dataset/GCA_963989415.1/cumulative.

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), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 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. Size-selection and clean-up were carried out using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using the Qubit Fluorometer v4.0 (ThermoFisher 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 the 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, the biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size-selected to 400–600 bp using SPRISelect beads. DNA was then enriched using Arima-HiC v2 Enrichment beads. The NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) was used for end repair, A-tailing, and adapter ligation, following a modified protocol in which library preparation is carried out while the DNA remains bound to the enrichment beads. PCR amplification was performed using KAPA HiFi HotStart mix and custom dual-indexed adapters (Integrated DNA Technologies) in a 96-well plate format. Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, samples were amplified for 10–16 PCR cycles. Post-PCR clean-up was carried out using SPRISelect beads. The libraries were quantified using the Accuclear Ultra

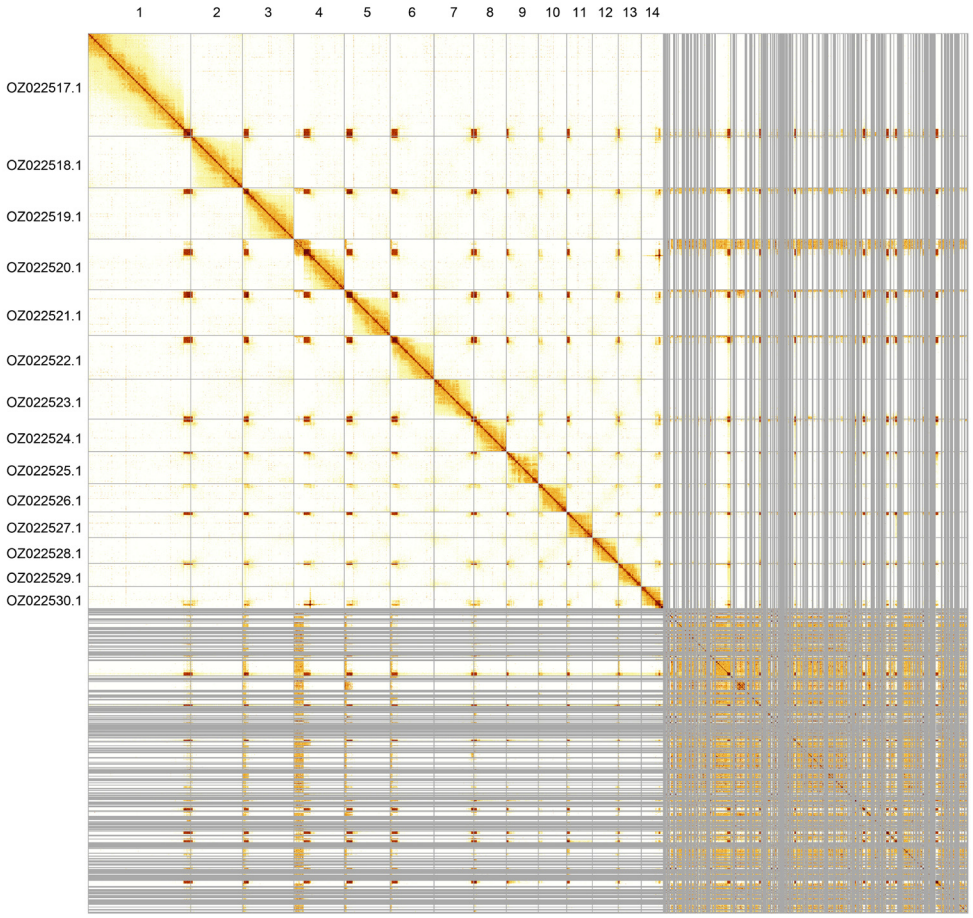

Figure 5. Genome assembly of *Cerceris ruficornis*. Hi-C contact map of the iyCerRufi2.1 assembly, generated using PretextSnapshot. Chromosomes are shown in order of size and labelled with chromosome numbers (top) and chromosome accession numbers (left).

Table 3. Chromosomal pseudomolecules in the genome assembly of *Cerceris ruficornis*, iyCerRufi2.

INSDC accession	Name	Length (Mb)	GC%
OZ022517.1	1	66.28	38.5
OZ022518.1	2	33.17	39.5
OZ022519.1	3	33.0	40.5
OZ022520.1	4	32.52	41.5
OZ022521.1	5	29.41	42
OZ022522.1	6	28.13	41.5
OZ022523.1	7	25.65	41
OZ022524.1	8	20.91	41.5
OZ022525.1	9	20.57	41
OZ022526.1	10	18.12	39.5
OZ022527.1	11	16.7	42.5
OZ022528.1	12	16.39	39.5
OZ022529.1	13	14.89	40.5
OZ022530.1	14	14.21	40
OZ022531.1	MT	0.02	16.5

High Sensitivity dsDNA Standards Assay kit (Biotium) and normalised to 10 ng/μL before sequencing. Hi-C sequencing was performed on the 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. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

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

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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 ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed in the blobtoolkit pipeline, a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools (Danecek *et al.*, 2021) 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

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	666652151335353eef2fcd58880bcef5bc2928e1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0

Software tool	Version	Source
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow
PretextSnapshot	-	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.4.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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 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: Cerceris ruficornis. Accession number PRJEB66742; https://identifiers.org/ena.embl/PRJEB66742. The genome sequence is released openly for reuse. The *Cerceris ruficornis* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Author information

Members of the Natural History Museum Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.12159242.

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.14870789.

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

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

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

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode 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

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

BWARS: ***Cerceris ruficornis*.** BWARS, 2025.
Reference Source

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

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

GBIF Secretariat: ***Cerceris ruficornis* (Fabricius, 1793)**. 2023.
Reference Source

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

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
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

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

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

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

Lomholdt O: **The Sphecidae (Hymenoptera) of Fennoscandia and Denmark.** 1st ed. Klampenborg: Scandinavian Science Press Ltd, 1976.
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

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

Pereira L, Sivell O, Sivess L, *et al.*: **DToL Taxon-specific Standard Operating Procedure for the terrestrial and freshwater arthropods working group**. 2022.
Publisher Full Text

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

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

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

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

Rhie A, Walenz BP, Koren S, *et al.*: **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

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

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

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ? ✓ ? ✓ ✓ ✓

Version 1

Reviewer Report 05 November 2025

https://doi.org/10.21956/wellcomeopenres.26767.r126007

✓ **Kashif Zahoor**
Government College University Faisalabad, Faisalabad, Punjab, Pakistan

Its really huge work, and well presented.
However, I've few comments:

"Most of the assembly sequence (65.35%) was assigned to 14 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size"

Must cite some references?
How assembly sequence was assigned to 14 chromosomal-level scaffolds??
How genetic data was assigned to respective chromosomes?
How chromosomes were annotated with numbering? How chromosome mapping can be verified? its only *in silico* which is possible at this stage?.... which protocol was followed for assigning/annotation of chromosomes? Its Rhie et al., 2021??
Mitochondrial gene COI was used only or other marker like COII, ND4, etc was used? If Mitochondria complete genome is 18.07kb?
Which "Tissue" was used?
GC% should be verified? or only If BUSCO scores are sufficient when compared with other Hymenoptera?

I recommend this manuscript for indexing after incorporating to my comments

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: Genetics, Molecular Biology, Entomology

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 08 September 2025

https://doi.org/10.21956/wellcomeopenres.26767.r130386

Sara E Miller
[1] Cornell University, Missouri, USA
[2] Department of Biology, University of Missouri St. Louis, Saint Louis, MO, USA

The authors present a new high quality nuclear and mitochondrial genome assembly for the solitary wasp, *Cerceris ruficornis* (Hymenoptera: Crabronidae). This is the first genome assembly for this species and adds to the available genomes for this large family of wasps.

Genome sequencing and assembly were completed using standard protocols of the Darwing Tree of Life Consortium. The genome was constructed using single female using Pacific Biosciences HiFi circular consensus DNA libraries, sequenced on the Biosciences SEQUEL II, and scaffolded with Hi-C data. Raw data and assembly methods are available. The genome assembly is contiguous and high quality with a scaffold N50 of 20.57 Mb and 96.5% coverage of BUSCO genes. An annotation is provided with this genome.

The final assembly is on the larger side for Hymenopteran genomes, at 566.08 Mb, and only 65.35% of the genome was able to be assigned to chromosomal-level scaffolds. Presumably the increase of genome size and difficulty resolving portions of this genome into chromosomal-level scaffolds is the result of expansion of repetitive sequences. However, the number and distribution of repetitive elements is not directly addressed by the authors and feels like a missed opportunity here.

Overall, this is a standard genome report and can be indexed as is.

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: Hymenoptera genetics and genomics, evolutionary ecology

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 05 September 2025

https://doi.org/10.21956/wellcomeopenres.26767.r126582

Alexandra Cerqueira De Araujo
University College London, London, UK

The authors present here a chromosome level *de novo* assembly of the genome of *Cerceris ruficornis* (Crabronidae, Hymenoptera), first assembly available for this species. This assembly is composed of 14 chromosomes and one mitochondrial chromosome, with an assembly total length of 566.08Gb and a N50 of 20.57Mb. This assembly achieves a EBP reference standard of 6.7.Q65 and contains most of the orthologous genes found in Hymenoptera (i.e. 96.5% of the complete BUSCO sequences). Altogether, these values reflect an assembly quality that is above the minimum reference standard for genome assemblies.

Although I deem this paper adequate for indexing, however I would like to raise a few comments that should be addressed before indexing:

1. In the “Sequencing data” paragraph and Table 1, the read count for ERR12102416 (i.e. 5.45e+08) doesn’t match the read count reported on NCBI (2.7e+08 reads reported by NCBI).
2. The authors have assembled the genome of *C. ruficornis* to chromosome level. They managed to obtained 14 chromosomes in total, plus extra sequences (i.e. 559 scaffolds). Is there any reference supporting this number of chromosomes for this species?
3. The assembly span for the genome of *C. ruficornis* is bigger than the span of most hymenopteran genomes (mean span = 338Gb, median span = 287Gb). There is also only 65% of the total assembly contained in the 14 chromosomes. How many genes, especially BUSCO genes, are present in these 14 chromosomes? I found in total 11093 sequences with the attribute “gene” in the gff provided in https://beta.ensembl.org/species/060ab491-5335-4f09-95ba-32ffa082ec0b, 11073 of them were found in the 14 chromosomes. I believe

that the author should emphasis this results as it shows that almost the entirety of coding regions is contained in the assembled 14 chromosomes. The genome size (in GB) could be smaller than the span mentioned in the manuscript. Of the remaining scaffolds, how many of them align to other scaffold in the assembly? i.e. how many of them are repeated?

4. I see some discrepancies in the “Genome annotation report” section:
    1. The average number of coding transcript per gene is 1.50 in the manuscript, but the number found in https://beta.ensembl.org/species/060ab491-5335-4f09-95ba-32ffa082ec0b is 1.57.
    2. The average transcript length is 11,558.03 bp in the manuscript, but this value cannot be found in https://beta.ensembl.org/species/060ab491-5335-4f09-95ba-32ffa082ec0b. Using AGAT v1.4.1 (agat_sp_statistics.pl), I found: mean mRNA length = 12404 bp, mean gene length = 12138 bp, mean number of exons per mRNA = 6.6 (but 6.2 in the manuscript).
5. The methods and protocols used to build and analyse the assembly are I believe appropriate. However, in its current state, it is not possible to build the exact same assembly built by the authors, as parameter information is missing. I do not see any link to code repository to generate this assembly and all data associated to it (did I miss it?). If the authors would prefer to not upload the code to an online repository, it is important that they mention the parameters used for all programs and tools mentioned in the manuscript (I believe they could be added to Table 4). If I did miss the link, please, make it more visible. I did see this link https://zenodo.org/records/10047654 , which is associated to this sentence “Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023).” in the “Assembly curation” section, but it is not clear whether this pipeline has been used only for this section or not.

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

Yes

Are the protocols appropriate and is the work technically sound?

Partly

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

Partly

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative genomics

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

Reviewer Report 04 September 2025

https://doi.org/10.21956/wellcomeopenres.26767.r130387

?

Sarah Inwood
University of Otago, Dunedin, New Zealand

The authors have used state-of-the-art sequencing and assembly methods to generate a genome for the solitary wasp *Cerceris ruficornis.* The genome was generated from a single individual of presumably unknown sex (this is not stated) using both HiFi and Hi-C data with BUSCO scores suggestive of a high-quality assembly. The authors also produced a mitochondrial genome assembly.

Minor comments:

Only 65.35% of the genome was scaffolded into chromosomal pseudomolecules and given the stated benchmark for this is above 90%, the lack of discussion of this result was surprising. Figure 3 shows that all of the scaffolded chromosomal pseudomolecules had hits to Arthropoda while none of the smaller unscaffolded contigs do, with these appearing to be in a couple of fairly distinct clusters. Given they could potentially be contaminants not detected by the ASCC pipeline, it would be of interest to know whether any of the unscaffolded contigs contained BUSCO hits or any other data to suggest they definitely belong in the genome assembly of *Cerceris ruficornis,* or whether the only data to suggest they belong in the assembly was them not being filtered out by the ASCC pipeline.

The Hi-C contact map in Figure 5 also looks a bit odd, with strong contacts between the ends of most chromosomal pseudomolecules. This result would also benefit from acknowledgement and brief discussion, whether it can be linked to something else e.g. is it repeat sequence, a potential artifact, or does it have no explanation? Chromosome 4 also seems to have interactions with a number of the unscaffolded contigs; this too could benefit from discussion - are these thought to be the result of incomplete scaffolding, or is there evidence to support them not being scaffolded?

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, host-parasite interactions, parasitoid wasps, insects, viruses, biocontrol

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 30 August 2025

https://doi.org/10.21956/wellcomeopenres.26767.r130391

Ariel Chipman
Hebrew University of Jerusalem School of Pharmacy, Jerusalem, Jerusalem District, Israel

Sivell et al. report on the genome sequence of the solitary wasp *Cerceris ruficornis*, a species known from much of Northern Europe. The specimen they sequenced was collected in Norfolk, England. The sequencing and assembly were done according to the standard protocols and metrics of the Darwin Tree of Life Consortium. The quality of the assembly is good but falls short of the benchmark in some of the metrics, specifically in the percentage of the sequence successfully assembled into chromosomes. BUSCO and completeness scores are high.
The report is fairly standard. It is written clearly and according to the guidelines for such a report. I see no reason not to accept the report as it is.

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: Arthropods evolution, evo-devo comparative genomics.

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

Reviewer Report 12 July 2025

https://doi.org/10.21956/wellcomeopenres.26767.r125320

?

Guanliang Meng
Leibniz Institute, Berlin, Germany

The authors use PacBio HiFi, Hi-C data to assemble the genome of *Cerceris ruficornis*, anchoring the scaffolds into 14 chromosomes, providing haplotype-level assemblies. Gene completeness analysis such as BUSCO scores suggest that the genome is high quality. My major concerns include:
(1) The differences between estimated genome size (454.59 Mb) and the assembled size (566.08 Mb);
(2) only 65.35% of the assembly was assigned to the 14 chromosomes. What are the true number of chromosomes for the species? What are the un-assigned scaffolds? Do they belong to contaminants (perhaps the ASCC pipeline was not able to identify these contaminants)? Especially that there are several clouds of dots in Figure 3;
(3) The intensive links between the ends of chromosomes (Figure 5). Are these due to assembly error, or due to wrong placements of the scaffolds, or due to repeats such as telomeres?

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, Transcriptomics, Bioinformatics, Evolutionary Biology

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.