DATA NOTE

The genome sequence of the Figwort Mason Wasp, *Symmorphus gracilis* (Brulle, 1832) (Hymenoptera: Vespidae)

[version 1; peer review: 2 approved]

Liam M. Crowley[1], Cian D Williams[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]Department of Biology, University of Oxford, Oxford, England, UK
[2]University of Cambridge Department of Zoology, Cambridge, England, UK

v1 **First published:** 31 Mar 2026, **11**:199
https://doi.org/10.12688/wellcomeopenres.26161.1
Latest published: 31 Mar 2026, **11**:199
https://doi.org/10.12688/wellcomeopenres.26161.1

Abstract

We present a genome assembly from an individual female *Symmorphus gracilis* (Figwort Mason Wasp; Arthropoda; Insecta; Hymenoptera; Vespidae). The assembly contains two haplotypes with total lengths of 220.03 megabases and 217.14 megabases. Most of haplotype 1 (92.39%) is scaffolded into 6 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 19.73 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Symmorphus gracilis, Figwort Mason Wasp, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 31 Mar 2026	✓ view	✓ view

1. **Sharu Paul Sharma**, Iowa State University, Iowa, USA

2. **Ralph S Peters**, Leibniz Institute, Bonn, Germany

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; **Williams CD**: 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, Williams CD, University of Oxford and Wytham Woods Genome Acquisition Lab *et al.* **The genome sequence of the Figwort Mason Wasp, *Symmorphus gracilis* (Brulle, 1832) (Hymenoptera: Vespidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:199 https://doi.org/10.12688/wellcomeopenres.26161.1

First published: 31 Mar 2026, **11**:199 https://doi.org/10.12688/wellcomeopenres.26161.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Vespidae; Eumeninae; *Symmorphus*; *Symmorphus gracilis* (Brulle, 1832) (NCBI:txid1124894).

Background

The Figwort Mason Wasp, *Symmorphus gracilis*, is a common and widespread potter wasp in the family Vespidae, subfamily Eumeninae. Its range extends across most of the Palearctic, from Western Europe to the Himalayas (Cumming, 1989; GBIF.org, 2025). It is common across the south of the UK, although it has been recorded as far north as Yorkshire (Bees, Wasps & Ants Recording Society, 2025; GBIF.org, 2025).

Members of *Symmorphus* can be identified by the presence of a transverse ridge and median longitudinal furrow on the first metasomal tergum (Cumming, 1989). Males have a simple antennal apex, in contrast to other eumenine genera, where it can be hooked or coiled. *S. gracilis* is black with yellow markings between the antennae and on the legs, a medially interrupted yellow band on the dorsal pronotum, and yellow bands on the apical borders of metasomal terga 1 to 5. There is substantial intraspecific colour variation in throughout its range, and some individuals also have yellow markings in other locations, including the clypeus, tegula, and scutellum (Cumming, 1989). It is medium-sized (7–12 mm) for a eumenine wasp, and females tend to be larger than males. It can be distinguished from other similar species of *Symmorphus* by its moderately projected pronotum with an acute humeral angle (Cumming, 1989).

Symmorphus gracilis can be found in a wide range of habitats, although adults prefer damp places close to streams and ditches (Bees, Wasps & Ants Recording Society, 2025). This association may be because it is mainly a specialist predator of *Cionus* weevil larvae (Budrienė, 2003; Wood & Goulson, 2016), which commonly feed on figwort (*Scrophularia* spp.), a plant that prefers damp soil. Females immobilise their prey by stinging, before carrying them to their nest. Like other *Symmorphus*, *S. gracilis* is a tube-nester, and uses existing cavities such as hollow plant stems, vacated burrows of wood-boring insects, or even anthropogenic habitats like straws in thatched roofs and holes in old mortar (Bees, Wasps & Ants Recording Society, 2025; Budrienė, 2003; Nevronytė, 2009). A series of 3 to 4 cells is constructed sequentially, each holding a single egg, stocked with prey, and sealed using clay (Budrienė, 2003). Unlike other *Symmorphus*, *S. gracilis* is not choosy about the diameter of its nests or the length of its cells (Budrienė *et al.*, 2004; Nevronytė, 2009). Offspring hatch and feed on the immobilised prey, before pupating and emerging as adults. In the UK, adults can be seen flying through the summer from May to August, most commonly in June and July (Bees, Wasps & Ants Recording Society, 2025; GBIF.org, 2025).

Taxonomic relationships within *Symmorphus*, and between *Symmorphus* and the rest of the Eumeninae, are not at all clear from morphology alone (Cumming, 1989). The complete genome sequence for a common member of this genus will help resolve Hymenopteran taxonomy. It will also facilitate investigations of the evolution of sociality, hunting strategies, and reproductive systems, for which the Hymenoptera are an excellent model system.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Symmorphus gracilis* (specimen ID Ox003097, ToLID iySymGrac1; Figure 1), collected from Wytham Woods, Oxfordshire, United Kingdom (latitude 51.767, longitude −1.311) on 2022-07-23. The specimen was collected and identified by Liam Crowley (University of Oxford).

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

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The iySymGrac1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the head and thorax was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Automated MagAttract v2 protocol. We used centrifuge-mediated fragmentation to produce DNA fragments in the 8–10 kb range, following the

Figure 1. Photograph of the *Symmorphus gracilis* (iySymGrac1) specimen used for genome sequencing.

Covaris g-TUBE protocol for ultra-low input (ULI). Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Prior to library preparation, the DNA was fragmented to ~10 kb. Ultra-low-input (ULI) libraries were prepared using the PacBio SMRTbell® Express Template Prep Kit 2.0 and gDNA Sample Amplification Kit. Samples were normalised to 20 ng DNA. Single-strand overhang removal, DNA damage repair, and end-repair/A-tailing were performed according to the manufacturer's instructions, followed by adapter ligation. A 0.85× pre-PCR clean-up was carried out with Promega ProNex beads.

The DNA was evenly divided into two aliquots for dual PCR (reactions A and B), both following the manufacturer's protocol. A 0.85× post-PCR clean-up was performed with ProNex beads. DNA concentration was measured using a Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with the Qubit HS Assay Kit, and fragment size was assessed on an Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit. PCR reactions A and B were then pooled, ensuring a total mass of ≥500 ng in 47.4 μl.

The pooled sample underwent another round of DNA damage repair, end-repair/A-tailing, and hairpin adapter ligation. A 1× clean-up was performed with ProNex beads, followed by DNA quantification using the Qubit and fragment size analysis using the Agilent Femto Pulse. Size selection was performed on the Sage Sciences PippinHT system, with target fragment size determined by Femto Pulse analysis (typically 4–9 kb). Size-selected libraries were cleaned with 1.0× ProNex beads and normalised to 2 nM before sequencing.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head and thorax tissue from the iySymGrac1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/µL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MerquryFK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021).

Manual corrections included 62 breaks and 208 joins. This reduced the scaffold count by 10.7% and increased the scaffold N50 by 4.9%. The curation process is described 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 MerquryFK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

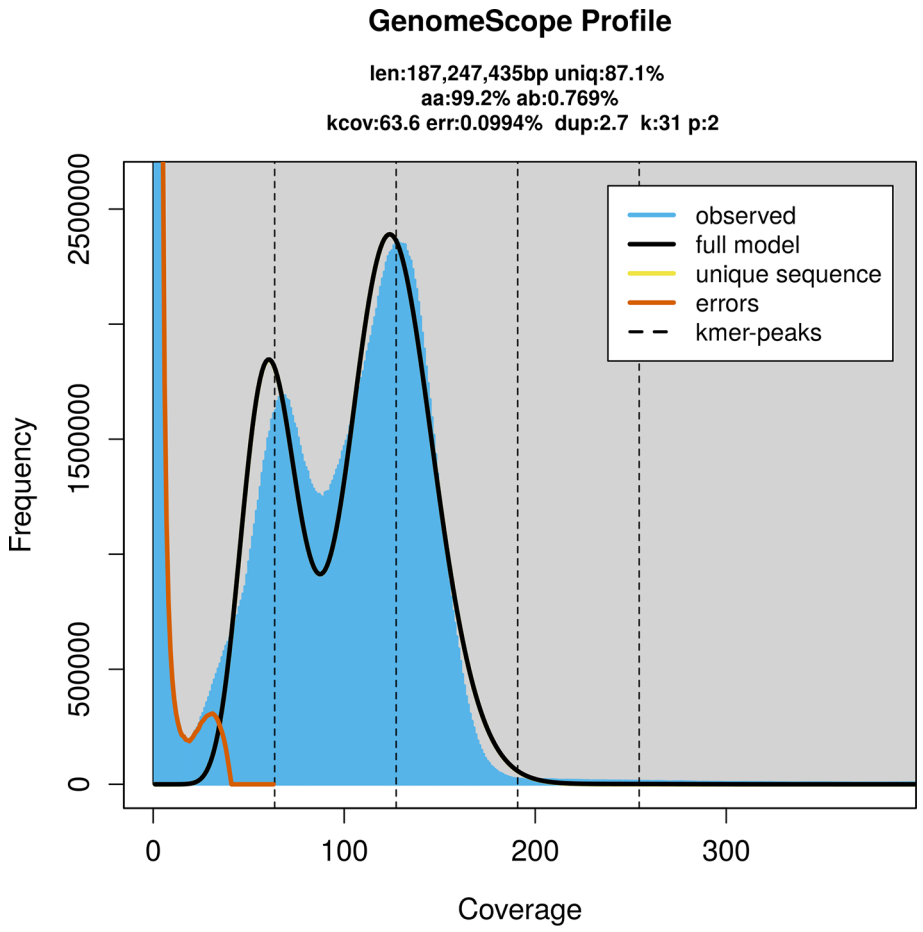

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Genome sequence report

Sequence data

PacBio sequencing of the *Symmorphus gracilis* specimen generated 24.99 Gb (gigabases) from 2.75 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 187.67 Mb, with a heterozygosity of 0.77% and repeat content of 12.86% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 127× coverage. Hi-C sequencing produced 98.11 Gb from 649.71 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 220.03 Mb in 516 scaffolds, with 464 gaps, and a scaffold N50 of 36.1 Mb (Table 2).

Most of the haplotype 1 assembly sequence (92.39%) was assigned to 6 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Scaffolds in the following regions have uncertain order and orientation: Chromosome 1 (~21.7–23.0 Mb), Chromosome 2 (9.3–11.2 Mb), Chromosome 3 (0–1.6 Mb), Chromosome 4 (0–6.8 Mb), Chromosome 5 (4.8–7.3 Mb), and Chromosome 6 (0–1.8 Mb).

Table 1. Specimen and sequencing data for BioProject PRJEB74532.

Platform	PacBio HiFi	Hi-C
ToLID	iySymGrac1	iySymGrac1
Specimen ID	Ox003097	Ox003097
BioSample (source individual)	SAMEA113425693	SAMEA113425693
BioSample (tissue)	SAMEA113425875	SAMEA113425875
Tissue	head and thorax	head and thorax
Instrument	Revio	Illumina NovaSeq 6000
Run accessions	ERR12833644	ERR12861051
Read count total	2.75 million	649.71 million
Base count total	24.99 Gb	98.11 Gb

Table 2. Genome assembly statistics.

Assembly name	iySymGrac1.hap1.1	iySymGrac1.hap2.1
Assembly accession	GCA_965620125.1	GCA_965620105.1
Assembly level	chromosome	scaffold
Span (Mb)	220.03	217.14
Number of chromosomes	6	scaffold-level
Number of contigs	980	924
Contig N50	0.68 Mb	0.66 Mb
Number of scaffolds	516	417
Scaffold N50	36.1 Mb	36.01 Mb
Longest scaffold length (Mb)	52.98	-
Organelles	Mitochondrion: 19.73 kb	-

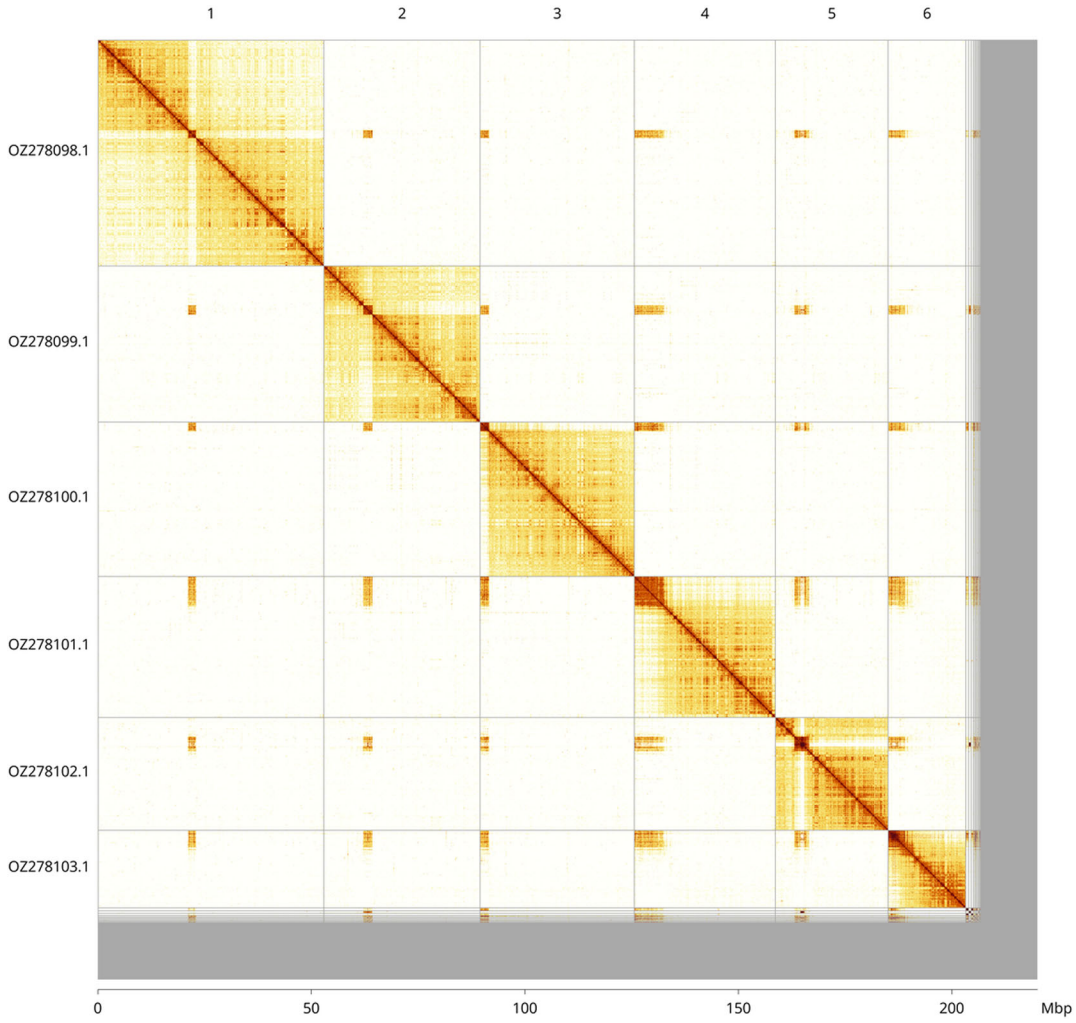

Figure 3. Hi-C contact map of the *Symmorphus gracilis* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Symmorphus gracilis* iySymGrac1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ278098.1	1	52.98	39.50
OZ278099.1	2	36.58	38
OZ278100.1	3	36.10	38.50
OZ278101.1	4	33.04	39.50
OZ278102.1	5	26.38	39
OZ278103.1	6	18.20	39.50

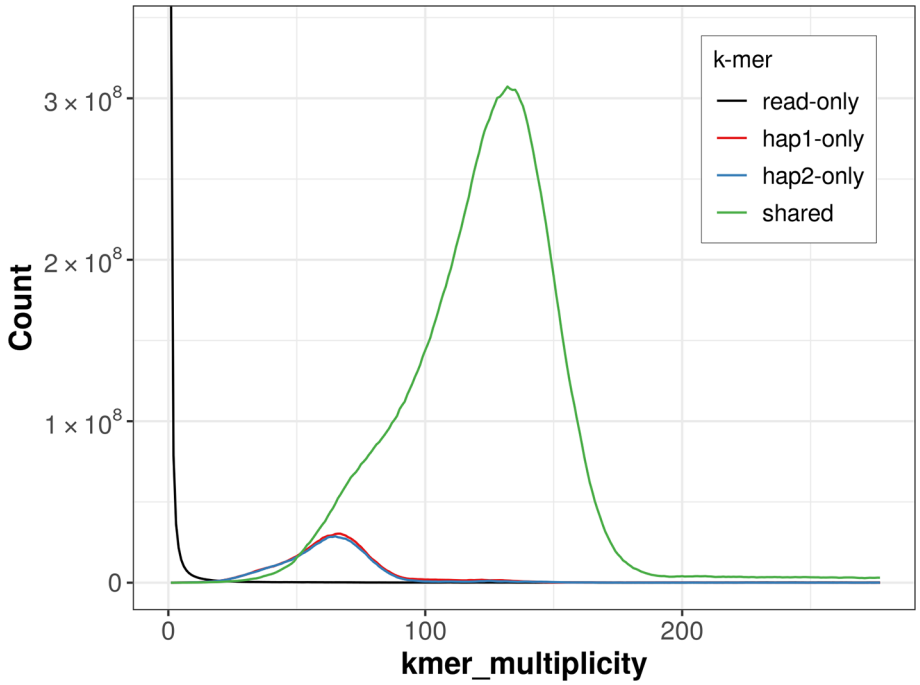

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

The mitochondrial genome was also assembled (length 19.73 kb, OZ278104.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 57.1, and for haplotype 2, 57.4. When the two haplotypes are combined, the assembly achieves an estimated QV of 57.2. The *k*-mer completeness is 90.17% for haplotype 1, 89.55% for haplotype 2, and 99.06% for the combined haplotypes (Figure 4).

BUSCO analysis using the hymenoptera_odb10 reference set (n = 5 991) identified 97.2% of the expected gene set (single = 96.3%, duplicated = 0.9%) in haplotype 1. For haplotype 2, BUSCO v.5.8.3 analysis identified 96.1% of the expected gene set (single = 95.2%, duplicated = 0.9%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **5.C.Q57**.

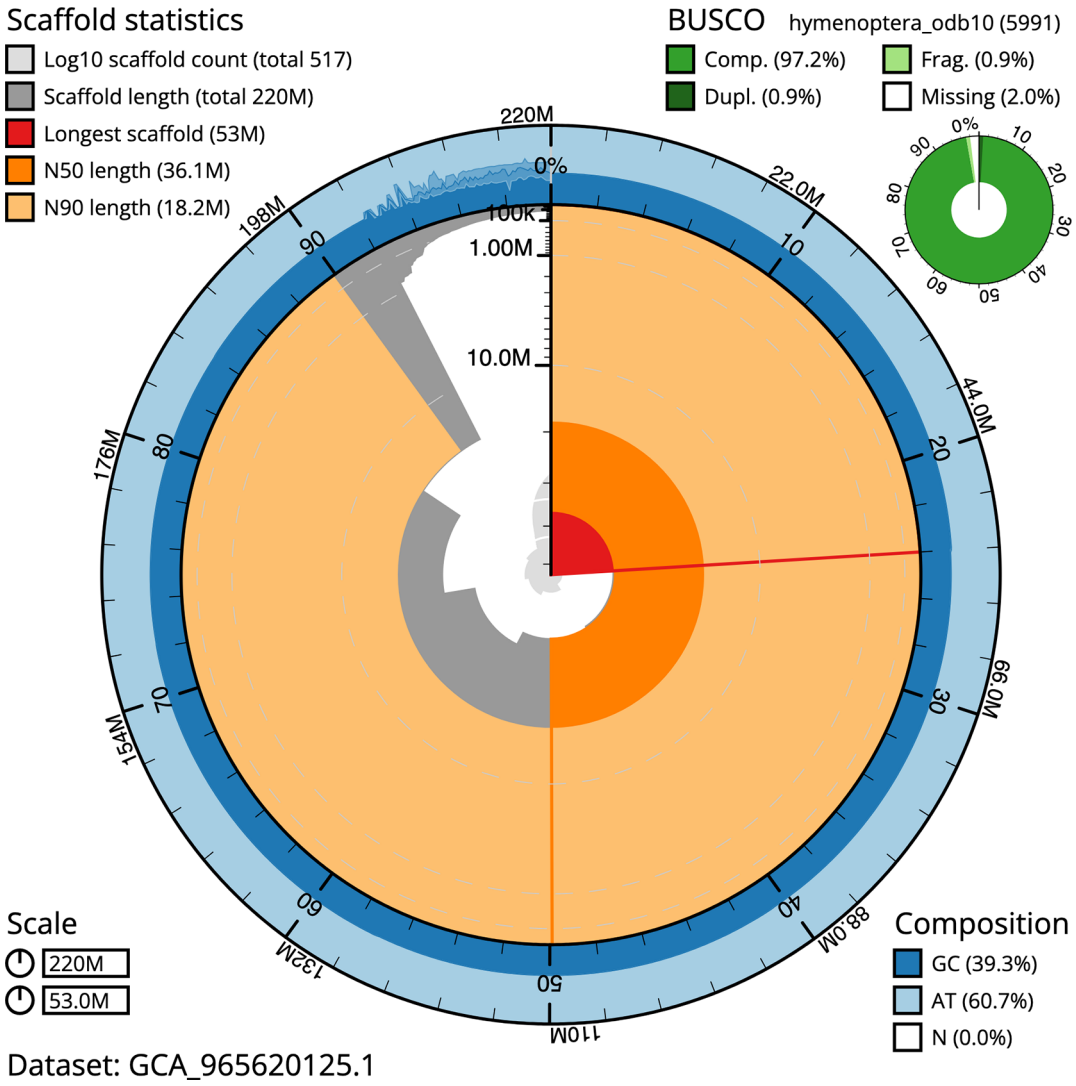

Figure 5. Assembly metrics for iySymGrac1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

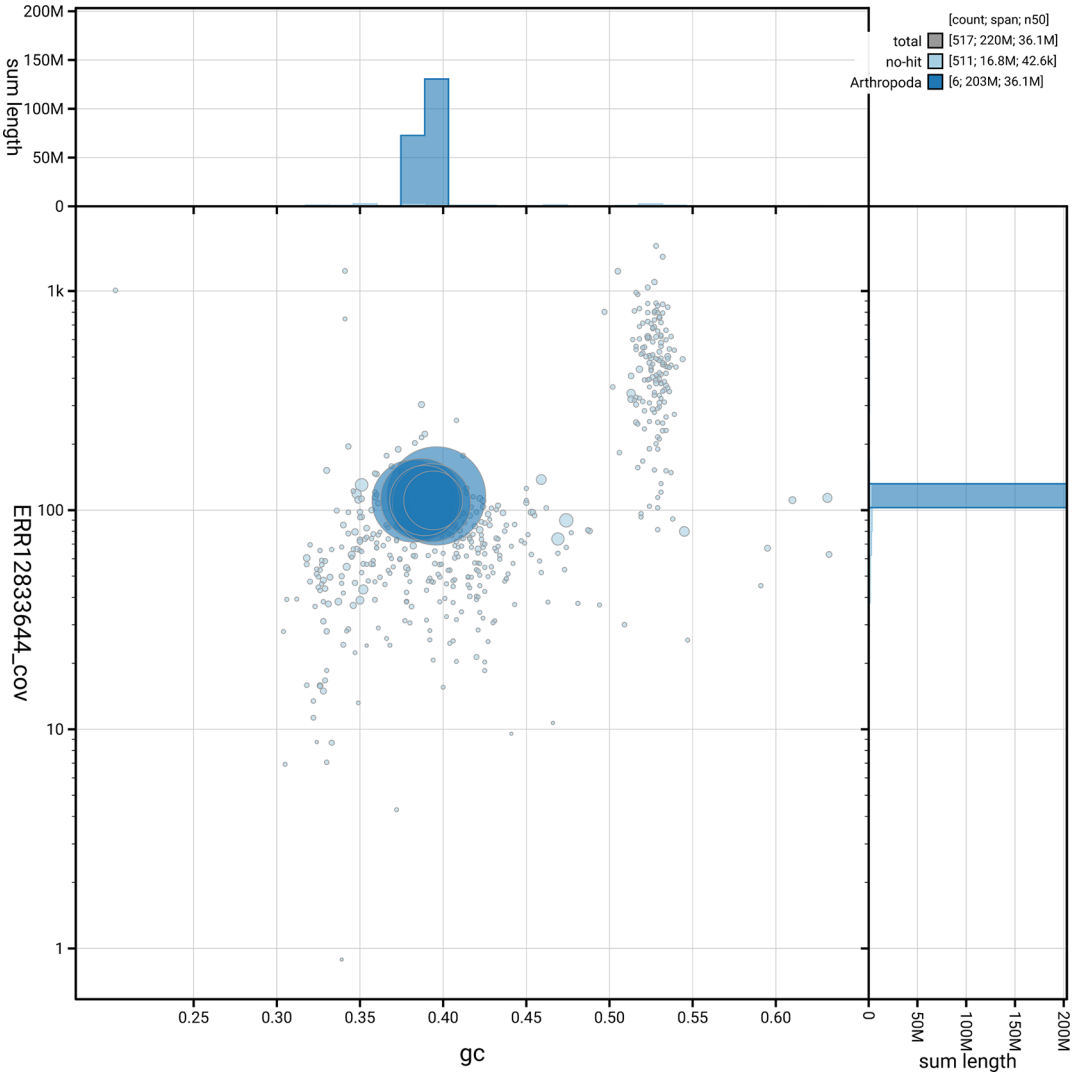

Figure 6. BlobToolKit blob plot for iySymGrac1.hap1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Symmorphus gracilis* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	5.C.Q57	6.C.Q40
Contig N50 length	0.68 Mb	≥ 1 Mb
Scaffold N50 length	36.10 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 57.1; haplotype 2: 57.4; combined: 57.2	≥ 40
k-mer completeness	Haplotype 1: 90.17%; Haplotype 2: 89.55%; combined: 99.06%	≥ 95%
BUSCO	C:99.0% [S:98.1%, D:0.9%], F:0.3%, M:0.8%, n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	92.39%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues.

Author information

Contributors are listed at the following links:

- Members of the University of Oxford and Wytham Woods Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Wellcome Sanger Institute – Legal and governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards 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: Symmorphus gracilis (figwort mason wasp). Accession number PRJEB74532. The genome sequence is released openly for reuse. The *Symmorphus gracilis* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Tables 1 and 2.

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

Table 5. Software versions and sources.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.8.3	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	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	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	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.8.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin M-J, 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

Bees, Wasps & Ants Recording Society: ***Symmorphus gracilis.*** 2025.
Reference Source

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
Publisher Full Text

Budrienė A, Budrys E, Nevronytė Ž: **Solitary Hymenoptera *Aculeata* inhabiting trap-nests in Lithuania: Nesting cavity choice and niche overlap.** *Latvijas Entomologs.* 2004.
Reference Source

Budrienė A: **Prey of *Symmorphus* wasps (Hymenoptera: Eumeninae) in Lithuania.** *Acta Zool Lituan.* 2003; **13**: 306–310.
Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 Genes|Genomes|Genetics.* 2020; **10**(4): 1361–1374.
Publisher 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.
Publisher Full Text

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat. Biotechnol.* 2022; **40**(9): 1332–1335.
Publisher Full Text

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial Arthropod fauna.** *Wellcome Open Res.* 2023; **8**: 123.
Publisher Full Text

Cumming JM: **CLASSIFICATION AND EVOLUTION OF THE EUMENINE WASP GENUS *SYMMORPHUS* WESMAEL (HYMENOPTERA: VESPIDAE).** *Mem Entomol Soc Can.* 1989; **121**: 5–168.
Publisher Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2).
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.
Publisher 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.
Publisher Full Text

GBIF Secretariat: **GBIF occurrence download.** 2025.
Publisher Full Text

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).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
Publisher Full Text

Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
Publisher Full Text

Li H: **Minimap2: Pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
Publisher Full Text

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

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239).
Publisher Full Text

Nevronytė Ž: **Development and nesting behaviour of trap-nesting wasps (Hymenoptera: Vespidae: Eumeninae).** 2009.
Reference Source

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

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
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.** *Wellcome Open Res.* 2024; **9**: 339.
Publisher 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.
Publisher Full Text

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–324.
Publisher Full Text

Wood T, Goulson D: **Aculeate bee and wasp survey report 2015/16 for the Knepp Wildland Project.** 2016.
Reference Source

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1).
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 07 May 2026

https://doi.org/10.21956/wellcomeopenres.28810.r153842

✔ **Ralph S Peters**
Leibniz Institute, Bonn, Germany

The manuscript accompanies publication of a chromosome-level genome of a representative of the solitary Vespidae subfamily Eumeninae (potter wasps, mason wasps). The manuscript and genome publication merely provides a resource for future research and provides no own studies on taxonomy, evolutionary biology or others. This, however, is no problem, as genome resource papers are common, and a high quality genome (even including two haplotypes from the same individual female) from a broad representation of the megadiverse Hymenoptera is highly welcome, including the mason wasps Symmorphus.
The background section with some info and intro to mason wasps and the genus Symmorphus provides information maybe even in too much detail (for example, in the description of differences between S. gracilis with congeners that is not really necessary here and also not ideally structured). Also, the end of the background section indicating some possible applications of the provided genome is quite generic (though not wrong) and could use some more specific ideas, for example, for studies within Vespidae. Ergo, the authors should consider shortening the section on diagnosis of S. gracilis (and rather refer to the literature) and extending the section on possible applications of the provided genome.
The methods and results (there is no discussion which is fine for a genome resource paper) are sufficient and reflect the current state-of-the-art of producing high-quality genomes.

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 diversity and evolution. phylogenomics, sometimes 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 30 April 2026

https://doi.org/10.21956/wellcomeopenres.28810.r153205

Sharu Paul Sharma
Iowa State University, Iowa, USA

The manuscript presents Figwort Mason Wasp (*S. gracilis*) genome assembly as a part of the Darwin Tree of Life project. The study provides a potentially useful genomic resource for future work in this and related taxa.
Overall, the manuscript is clear and concise with metrics expected of a high quality assembly. There are a few points that need clarification before publication.
- The authors should briefly discuss why the final assembly span is substantially larger than the GenomeScope2 haploid genome size estimate.
- The assembly meets most high-level quality criteria, but the contig N50 falls below the benchmark reported in Table 4. The authors should briefly acknowledge this and clarify whether it affects downstream use of the assembly.
- The authors should clarify how users should interpret the lower k-mer completeness of each individual haplotype relative to the benchmark, especially since the combined haplotypes achieve high completeness. This is important for users who may rely primarily on haplotype 1.

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: Bioinformatics

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