DATA NOTE

The genome sequence of the ichneumonid wasp, *Sussaba pulchella* (Holmgren, 1858) (Hymenoptera: Ichneumonidae)

[version 1; peer review: 2 approved]

Ian Barnes[1], Chris Fletcher[1], Inez Januszczak[1], Gavin R. Broad[1], Liam M. Crowley[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]University of Oxford, Oxford, England, UK

v1 **First published:** 10 Oct 2025, **10**:564
https://doi.org/10.12688/wellcomeopenres.24968.1
Latest published: 10 Oct 2025, **10**:564
https://doi.org/10.12688/wellcomeopenres.24968.1

Abstract

We present a genome assembly from an individual male *Sussaba pulchella* (ichneumonid wasp; Arthropoda; Insecta; Hymenoptera; Ichneumonidae). The genome sequence has a total length of 299.91 megabases. Most of the assembly (81.88%) is scaffolded into 15 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 34.69 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

Sussaba pulchella, ichneumonid wasp, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 10 Oct 2025	✓ view	✓ view

1. **Erinn Fagan-Jeffries**, The University of Adelaide, Adelaide, Australia

2. **Taslima Sheikh**, Baba Ghulam Shah Badshah University, Rajouri, India
Baba Ghulam Shah Badshah University (Ringgold ID: 209367), Jammu, India

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: Barnes I: Investigation, Resources; **Fletcher C**: Investigation, Resources; **Januszczak I**: Investigation, Resources; **Broad GR**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Crowley LM**: Investigation, Resources;

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).
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: Barnes I, Fletcher C, Januszczak I *et al.* **The genome sequence of the ichneumonid wasp, *Sussaba pulchella* (Holmgren, 1858) (Hymenoptera: Ichneumonidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:564 https://doi.org/10.12688/wellcomeopenres.24968.1

First published: 10 Oct 2025, **10**:564 https://doi.org/10.12688/wellcomeopenres.24968.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Ichneumonoidea; Ichneumonidae; Diplazontinae; *Sussaba*; *Sussaba pulchella* (Holmgren, 1858) (NCBI:txid662494)

Background

Sussaba pulchella is a small (forewing ~4 mm) ichneumonid wasp of the subfamily Diplazontinae. As with all diplazontines for which host associations are known, *S. puchella* is a parasitoid of hoverfly (Diptera: Syrphidae) larvae. The biology of *S. pulchella* is poorly known. Although there are published host records from aphid-feeding hoverflies of the genus *Sphaerophoria*, such as *S. scripta* (Linnaeus) (e.g., Scott, 1939), these are probably due to misidentifications of other *Sussaba* species, particularly *S. flavipes*, which had previously been misidentified as *S. pulchella* (Diller, 1980). One species of *Sussaba* has been reared from larvae of *Pipizella* feeding on root-dwelling aphids (Fitton & Rotheray, 1982), and the laterally compressed metasomas of *Sussaba* suggest that many of their hosts might be concealed in such a way, given that diplazontines which attack more exposed syrphid larvae tend to be frequently reared. Diplazontines emerge from the host puparium but oviposit in either the egg of the syrphid host (in the developing embryo), early or mid-instar larva. Host stage preference does not seem to be known for *Sussaba* but might be expected to be into the egg or early instar host, given the preferences of the most closely related genera with known host associations (Klopfstein *et al.*, 2011).

Amongst Diplazontinae, *Sussaba* species can be recognised by the placement of the spiracle on the second metasomal segment, where it is on the laterotergite rather than the dorsal part of the tergite, as in all other Diplazontinae. The face of *Sussaba* species has two vertical grooves and is often unsculptured, thus resembling *Promethes*. Klopfstein (2014) provided keys to European Diplazontinae, including *Sussaba*, but identification of *Sussaba* species can be tricky. Males are more straightforward as the colour and shape of tyloids on the antennae tend to be species-specific. These tyloids are raised, ovoid areas of cuticle on the antennal flagellomeres and are involved in coiling the antennae around the female's antennae, which is part of the courtship process in some clades of Diplazontinae (Klopfstein *et al.*, 2010). Females can be separated from other species with the metasoma weakly laterally compressed by the colour pattern of the legs (coxae and trochanters all partly black), the distinct yellow 'shoulder marks' on the mesoscutum and the shining hind coxa. Males have uniformly dark brown tyloids, yellow and orange on the metasoma and fairly long antennae (Klopfstein, 2014).

Sussaba pulchella is frequently collected and widespread across Britain, usually found by sweep netting. Its global range encompasses much of the Palaearctic, Nearctic and India (Dasch, 1964; Klopfstein, 2014; Morley, 1913).

We present a chromosome-level genome sequence for *Sussaba pulchella*, the first high-quality genome for the genus *Sussaba* (data obtained via NCBI datasets, O'Leary *et al.*, 2024). The assembly was produced using the Tree of Life pipeline from a specimen collected in Wytham Woods, England, UK (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Blaxter *et al.*, 2022).

Methods

Sample acquisition and D.NA barcoding

The specimen used for genome sequencing was an adult *Sussaba pulchella* (specimen ID NHMUK014451610, ToLID iySusPulc1; Figure 1), collected from Wytham Woods, Bert's Pheasant Pen, England, United Kingdom (latitude 51.77, longitude –1.31) on 2021-09-02. The specimen was collected by Ian Barnes, Chris Fletcher, Inez Januszczak, Gavin Broad and Liam Crowley. For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

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

Figure 1. Photograph of the *Sussaba pulchella* (iySusPulc1) specimen used for genome sequencing.

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 iySusPulc1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism 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 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 using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the iySusPulc1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. 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 (Cheng *et al.*, 2021) with the --primary and --l0 options. The latter option is to switch off internal hifiasm purging for a haploid assembly. 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) with 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. 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 58 breaks, 124 joins, and removal of four duplications. 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) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality using the *k*-mer databases (k = 31) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

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

Genome sequence report

Sequence data

PacBio sequencing of the *Sussaba pulchella* specimen generated 22.63 Gb (gigabases) from 2.55 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 233.58 Mb, with a heterozygosity of 0.64% and repeat content of 21.52% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 86× coverage. Hi-C sequencing produced 70.97 Gb from 469.99 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

This assembly is from a haploid male specimen. The final assembly has a total length of 299.91 Mb in 1 379 scaffolds, with 1 639 gaps, and a scaffold N50 of 14.66 Mb (Table 2).

Most of the haploid assembly sequence (81.88%) was assigned to 15 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

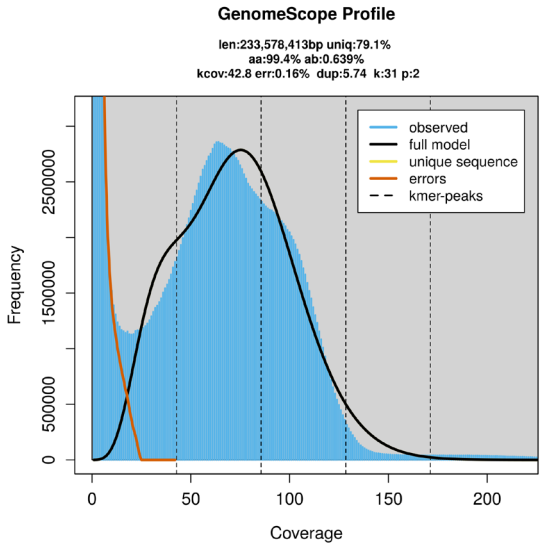

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB63414.

Platform	PacBio HiFi	Hi-C
ToLID	iySusPulc1	iySusPulc1
Specimen ID	NHMUK014451610	NHMUK014451610
BioSample (source individual)	SAMEA14448316	SAMEA14448316
BioSample (tissue)	SAMEA14448514	SAMEA14448514
Tissue	whole organism	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000
Run accessions	ERR11593791	ERR11606303
Read count total	2.55 million	469.99 million
Base count total	22.63 Gb	70.97 Gb

Table 2. Genome assembly statistics.

Assembly name	iySusPulc1.1
Assembly accession	GCA_963971145.1
Assembly level	chromosome
Span (Mb)	299.91
Number of chromosomes	15
Number of contigs	3 018
Contig N50	0.2 Mb
Number of scaffolds	1 379
Scaffold N50	14.66 Mb
Sex chromosomes	N/A
Organelles	Mitochondrion: 34.69 kb

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.

The haploid assembly achieves an estimated QV of 55.3. The *k*-mer completeness is 98.89% for the haploid assembly (Figure 4).

BUSCO v.5.5.0 analysis using the hymenoptera_odb10 reference set (n = 5 991) identified 92.5% of the expected gene set (single = 92.1%, duplicated = 0.4%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haploid assembly, is **5.7.Q55**.

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

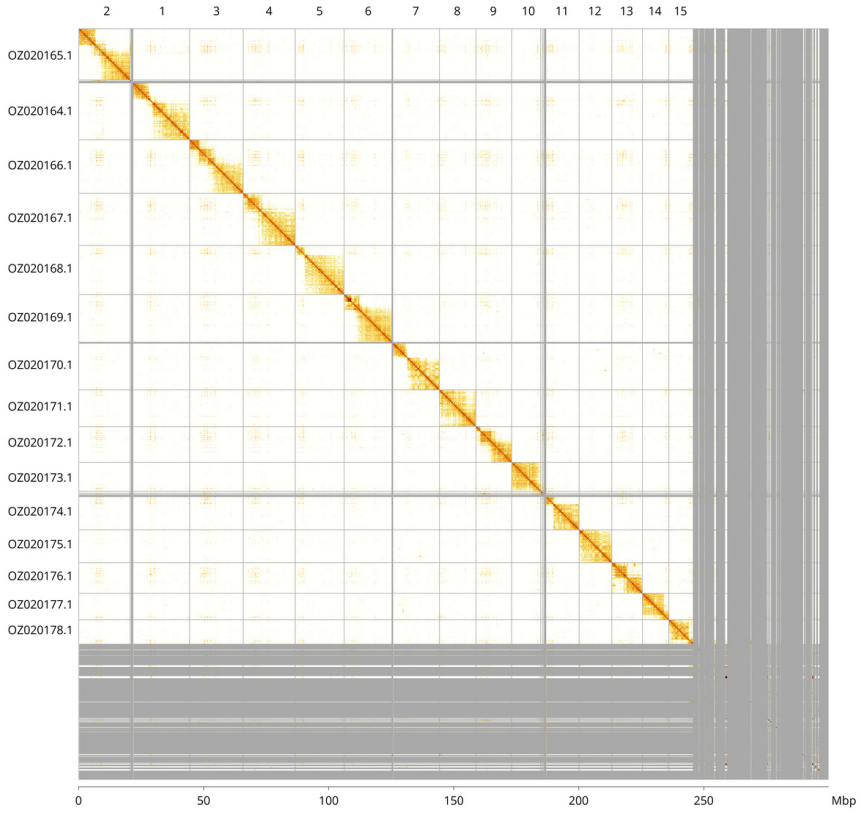

Figure 3. Hi-C contact map of the *Sussaba pulchella* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haploid genome assembly of *Sussaba pulchella* iySusPulc1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ020164.1	1	21.77	39.50
OZ020165.1	2	22.63	39
OZ020166.1	3	21.43	40.50
OZ020167.1	4	20.76	40.50
OZ020168.1	5	19.49	40
OZ020169.1	6	19.49	40.50
OZ020170.1	7	18.54	39
OZ020171.1	8	14.66	39.50
OZ020172.1	9	14.22	38.50
OZ020173.1	10	13.74	39.50
OZ020174.1	11	13.18	40
OZ020175.1	12	13.08	39
OZ020176.1	13	12.23	40
OZ020177.1	14	10.62	38.50
OZ020178.1	15	9.71	38

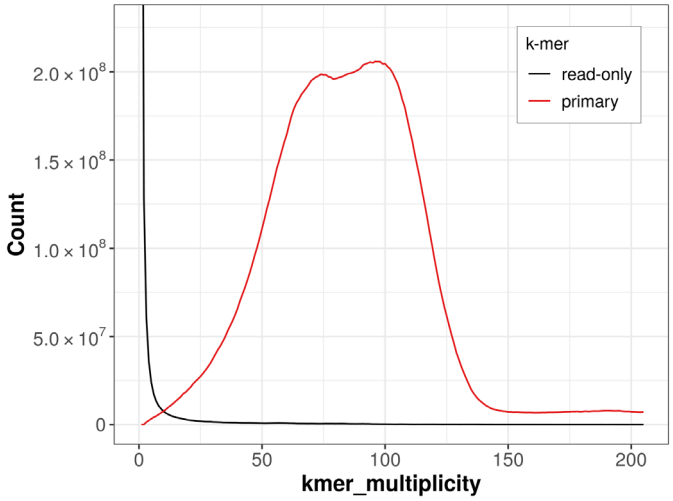

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assembly. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The red curve corresponds to *k*-mers present in the haploid assembly.

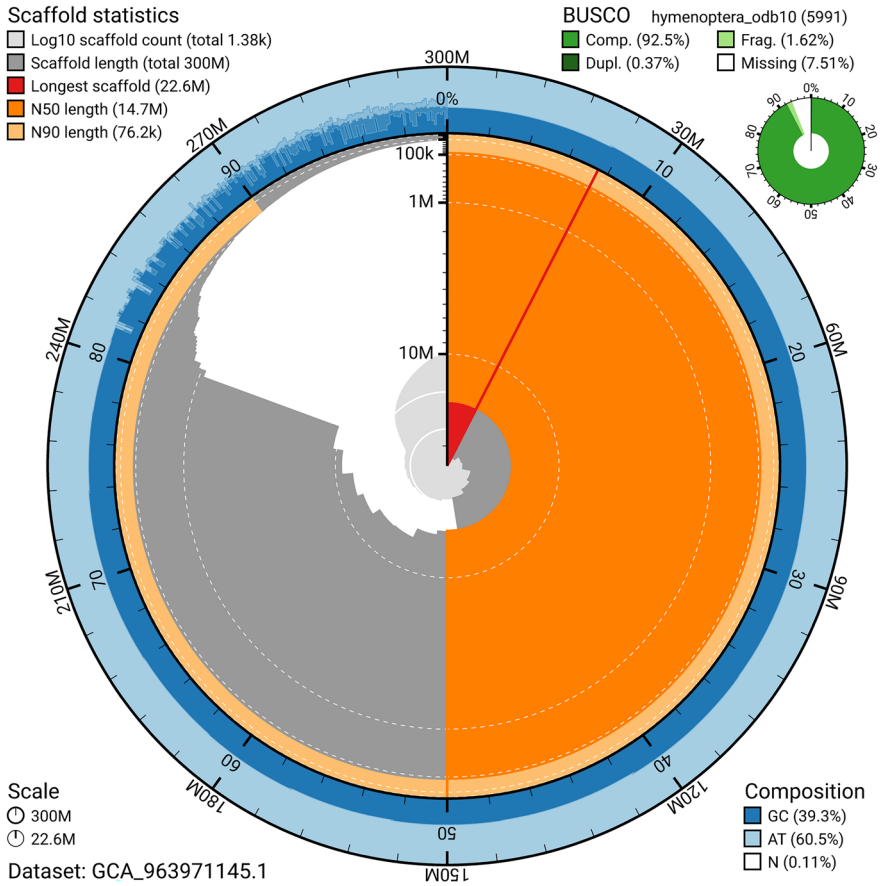

Figure 5. Assembly metrics for iySusPulc1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 hymenoptera_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

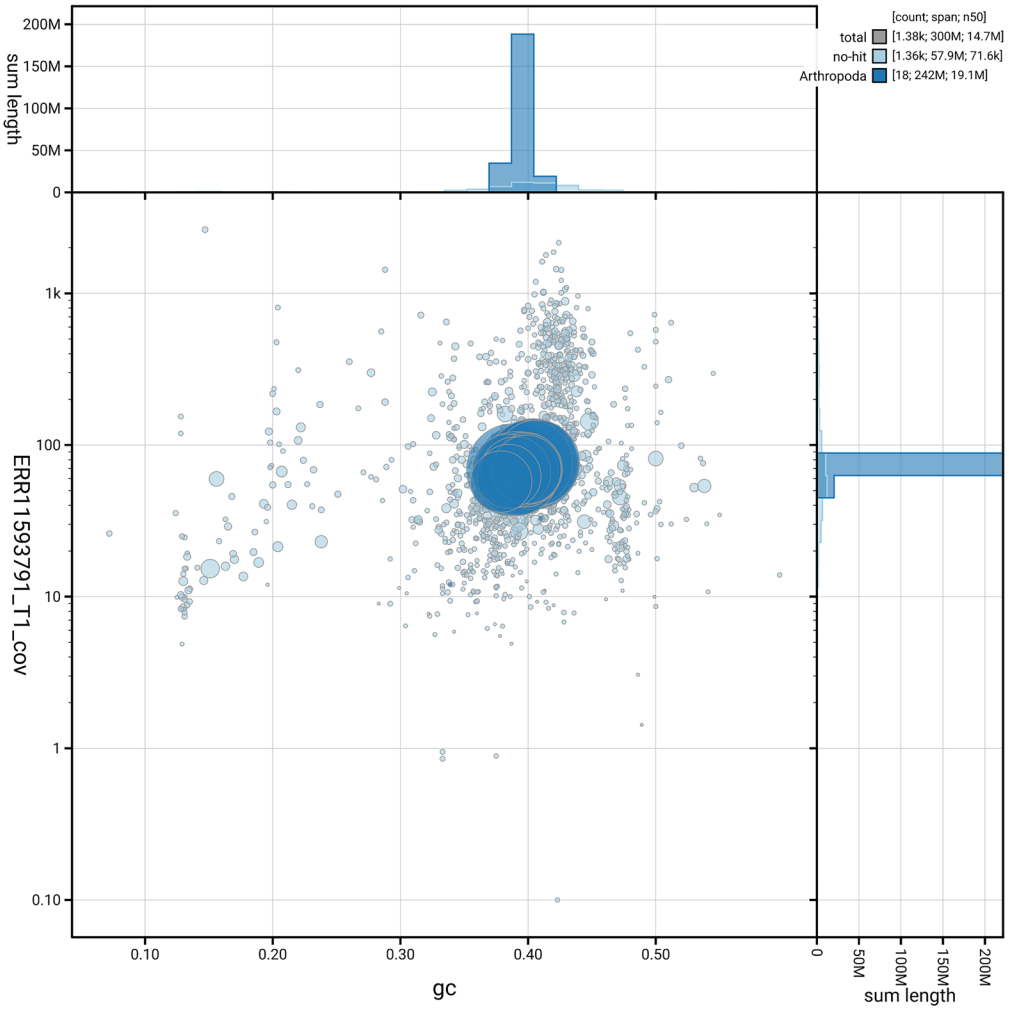

Figure 6. BlobToolKit GC-coverage plot for iySusPulc1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Sussaba pulchella* assembly.

Measure	Value	Benchmark
EBP summary (primary)	5.7.Q55	6.C.Q40
Contig N50 length	0.20 Mb	≥ 1 Mb
Scaffold N50 length	14.66 Mb	= chromosome N50
Consensus quality (QV)	55.3	≥ 40
k-mer completeness	98.89%	95%
BUSCO	C:92.5% [S:92.1%; D:0.4%]; F:1.6%; M:5.9%; n:5 991	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	81.88%	≥ 90%

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: Sussaba pulchella. Accession number PRJEB63414. The genome sequence is released openly for reuse. The *Sussaba pulchella* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

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

Author information

Contributors are listed at the following links:

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

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	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	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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.16.1	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	2.24-r1122	https://github.com/lh3/minimap2

Software	Version	Source
MitoHiFi	3.01	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	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.4.0	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	1.1a.2	https://github.com/c-zhou/yahs

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

Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **Sequence locally, think globally: the Darwin Tree of Life Project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free 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, 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

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

Dasch CE: **Ichneumon-flies of America North of Mexico: 5. Subfamily Diplazontinae**. American Entomological Institute, 1964; 3.
Reference Source

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

Fitton MG, Rotheray GE: **A key to the European genera of diplazontine ichneumon-flies, with notes on the British fauna.** *Syst Entomol.* 1982; **7**(4): 311–20.
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

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

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

Klopfstein S: **Revision of the Western Palaearctic Diplazontinae (Hymenoptera, Ichneumonidae).** *Zootaxa.* 2014; **3801**: 1–143.
PubMed Abstract | Publisher Full Text

Klopfstein S, Quicke DLJ, Kropf C: **The evolution of antennal courtship in diplazontine parasitoid wasps (Hymenoptera, Ichneumonidae, Diplazontinae).** *BMC Evol Biol.* 2010; **10**(218): 218.
PubMed Abstract | Publisher Full Text | Free Full Text

Klopfstein S, Quicke DLJ, Kropf C, *et al.*: **Molecular and morphological phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae).** *Zool Scr.* 2011; **40**: 379–402.
Publisher Full Text

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

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

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

Morley C: **The Fauna of British India, including Ceylon and Burma**. Hymenoptera - Vol.III. Ichneumonidae. London: Taylor & Francis; 1913.

O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the Tree of Life with NCBI datasets.** *Sci Data.* 2024; **11**(1): 732.
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**

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

Scott EI: **An account of the developmental stages of some aphidophagous Syrphidae (Dipt.) And their parasites (Hymenopt.).** *Ann Appl Biol.* 1939; **26**(3): 509–32.
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 18 November 2025

https://doi.org/10.21956/wellcomeopenres.27503.r138539

✔ **Taslima Sheikh**
[1] Baba Ghulam Shah Badshah University, Rajouri, Jammu and Kashmir, India
[2] zoology, Baba Ghulam Shah Badshah University (Ringgold ID: 209367), Jammu, India

This data note reports a high-quality genome assembly and annotation for the ichneumonid wasp *Sussaba pulchella*. The study is well executed and makes a useful addition to genomic resources within the Hymenoptera. The methods used (PacBio HiFi sequencing, Hi-C scaffolding, and mitochondrial genome assembly) are appropriate and produce robust results.
Only a few **minor revisions** are needed:

1. Correct the typographical error where the species name appears once as *S. puchella*.
2. Clarify that the male and female morphological characters described in the Background refer specifically to *S. pulchella*, and mention whether the combination is diagnostic.
3. Provide a short explanation for the unusually large mitochondrial genome (34.69 kb) ; for example, whether it represents true duplication, tandem repeats, or an assembly artifact and confirm with coverage evidence if available.
4. Add the tissue type used for DNA extraction and COI barcoding, and if possible, include the approximate DNA input quantity for the library.

With these small additions, the article will be fully sound and suitable for indexing
Recommendation: Accept after minor revision.

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: Entomology; Lepidoptera; Orthoptera; Hymenoptera

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 November 2025

https://doi.org/10.21956/wellcomeopenres.27503.r136058

Erinn Fagan-Jeffries
The University of Adelaide, Adelaide, Australia

This article gives a clear introduction to the taxonomy and biology of *Sussaba pulchella*, the ichneumonid wasp for which a genome assembly is presented. The DNA sequencing and bioinformatic pipelines are included and referenced appropriately, and summary statistics of the genome assembly are presented in legible figures.

Specific suggestions for minor improvements to the article text, none of which are required for the article to be scientifically sound:
1) where the article introduction states "Females can be separated from other species with the metasoma weakly laterally compressed..." I assume this is talking about *S. pulchella*, but the sentence would be less ambiguous if that was included (i.e., "Females of *S. pulchella* can be separated from other species of *Sussaba* with the metasoma..."). Similarly for the last sentence of this paragraph that starts "Males have uniformly..." it would be helpful to explicitly include if the listed characters in combination are diagnostic of the species or not.
2) In the methods where it states "a small sample was dissected from the specimen" - it is unusual not to see the body part listed (e.g., is this a leg, part of the abdomen, etc.). I guess not necessary, but useful for future reference perhaps?

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: Systematics of Ichneumonoidea

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.