DATA NOTE

The genome sequence of the solitary wasp, *Nysson trimaculatus* (Rossi, 1790) (Hymenoptera: Crabronidae)

[version 1; peer review: 3 approved, 1 approved with reservations]

Ryan Mitchell[1], Idris Adams[2], Gavin R. Broad[3],
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]Independent researcher, Sligo, County Sligo, Ireland
[2]University College London, London, England, UK
[3]Natural History Museum, London, England, UK

v1 **First published:** 02 May 2026, **11**:257
https://doi.org/10.12688/wellcomeopenres.26493.1
Latest published: 02 May 2026, **11**:257
https://doi.org/10.12688/wellcomeopenres.26493.1

Abstract

We present a genome assembly from an individual female *Nysson trimaculatus* (solitary wasp; Arthropoda; Insecta; Hymenoptera; Crabronidae). The assembly contains two haplotypes with total lengths of 301.31 megabases and 403.46 megabases. Most of haplotype 1 (86.44%) is scaffolded into 28 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 25.2 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

Nysson trimaculatus, 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
version 1 02 May 2026	✓ view	✓ view	✓ view	? view

1. **Vahap Eldem**, Istanbul University, Istanbul, Turkey
2. **Sharu Paul Sharma**, Iowa State University, Iowa, USA
3. **Marta Coronado-Zamora**, Universitat Autònoma de Barcelona - UAB Barcelona, Bellaterra, Spain

Vicente Chillida, Universitat Autonoma de Barcelona, Bellaterra, Spain
Autonomous University of Barcelona
Department of Genetics and Microbiology, Bellaterra, Spain

4. **Din Syafruddin**, Universitas Hasanuddin, Makassar, Indonesia

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: Mitchell R: Investigation, Resources; **Adams I**: Writing – Original Draft Preparation; **Broad GR**: Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Mitchell R, Adams I, Broad GR *et al.* **The genome sequence of the solitary wasp, *Nysson trimaculatus* (Rossi, 1790) (Hymenoptera: Crabronidae) [version 1; peer review: 3 approved, 1 approved with reservations]** Wellcome Open Research 2026, **11**:257 https://doi.org/10.12688/wellcomeopenres.26493.1

First published: 02 May 2026, **11**:257 https://doi.org/10.12688/wellcomeopenres.26493.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Crabronidae; Bembicinae; Nyssonini; *Nysson*; *Nysson trimaculatus* (Rossi, 1790) (NCBI: txid2495129).

Background

Nysson trimaculatus is a small to medium-sized (6–8 mm) solitary wasp in the family Bembicidae (for identification, see Richards (1980)), occurring across much of Europe from Scandinavia to the Mediterranean. In the UK, it is found throughout southern England and Wales with occasional records expanding northwards (BWARS, 2026). Flying from June to September, *N. trimaculatus* occurs in a variety of open habitats such as heathlands, scrub and quarries.

Like other *Nysson* species, this yellow and black wasp is a kleptoparasite of other Hymenoptera, with hosts including *Gorytes laticinctus* (Wolf, 1951) and *G. quadricinctus* (Nemkov, 2012), *Lestiphorus bicinctus* (Benno, 1966), and *Oryttus concinnus* (Nemkov, 2012). Host nests are located by scent, with an egg laid at the bottom of the infiltrated cell; once hatched, the *N. trimaculatus* larva will destroy the host egg and consume the provisioned Hemipteran prey (Evans, 1966). Bitsch *et al*. (2020) note that adults have been observed on inflorescences of Apiaceae, while other observations have been made on leaves of the sycamore maple *Acer pseudoplatanus* (BWARS, 2026).

The complete genome of *N. trimaculatus* presented here will help facilitate further study on the evolution of kleptoparasitism, prey capture, and hunting strategies in the Bembicidae, and on Hymenopteran taxonomy. The assembly was produced using the Tree of Life pipeline using a specimen from Buxton Heath, England, UK (Figure 1).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Nysson trimaculatus* (specimen ID NHMUK015059101, ToLID iyNysTrim1; Figure 1), collected from Buxton Heath, England, UK (latitude 52.75, longitude 1.22) on 2022-07-04. The specimen was collected and identified by Ryan Mitchell.

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

Detailed protocols for nucleic acid extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The iyNysTrim1 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. High molecular weight (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.

Figure 1. Photograph of the *Nysson trimaculatus* (iyNysTrim1) specimen used for genome sequencing.

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 iyNysTrim1 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.

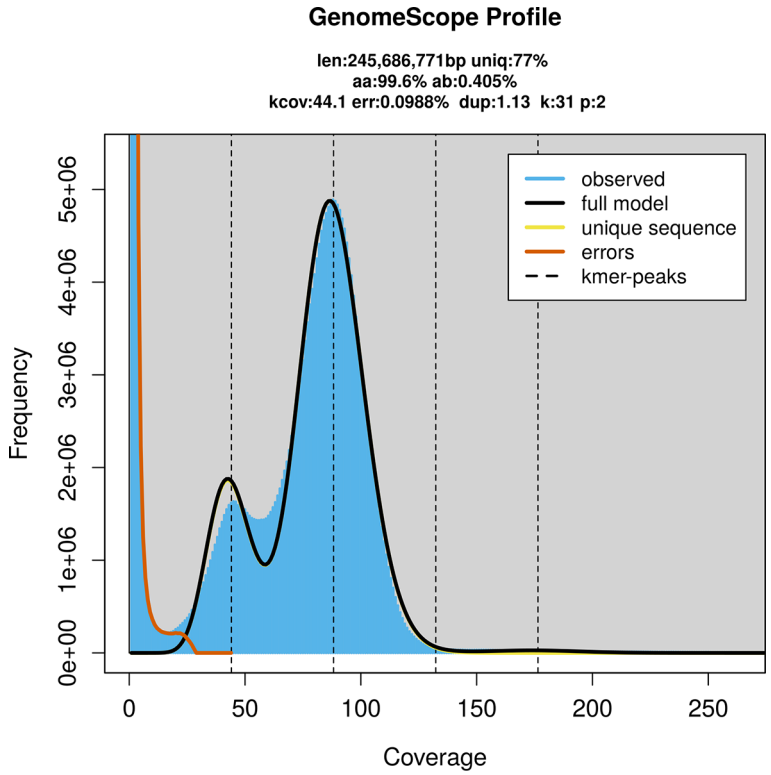

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB83146.

Platform	PacBio HiFi	Hi-C
ToLID	iyNysTrim1	iyNysTrim1
Specimen ID	NHMUK015059101	NHMUK015059101
BioSample (source individual)	SAMEA112964094	SAMEA112964094
BioSample (tissue)	SAMEA112975225	SAMEA112975225
Tissue	whole organism	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000
Run accessions	ERR14048090	ERR14056191
Read count total	1.99 million	346.17 million
Base count total	22.94 Gb	104.54 Gb

Table 2. **Genome assembly statistics.**

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	iyNysTrim1.hap1.1	iyNysTrim1.hap2.1
Assembly accession	GCA_975301485.1	GCA_975301255.1
Assembly level	chromosome	scaffold
Span (Mb)	301.31	403.46
Number of chromosomes	28	scaffold-level
Number of contigs	797	2 657
Contig N50	1.9 Mb	0.86 Mb
Number of scaffolds	596	2 473
Scaffold N50	10.25 Mb	4.21 Mb
Longest scaffold length (Mb)	17.05	-
Organelles	Mitochondrion: 25.2 kb	-

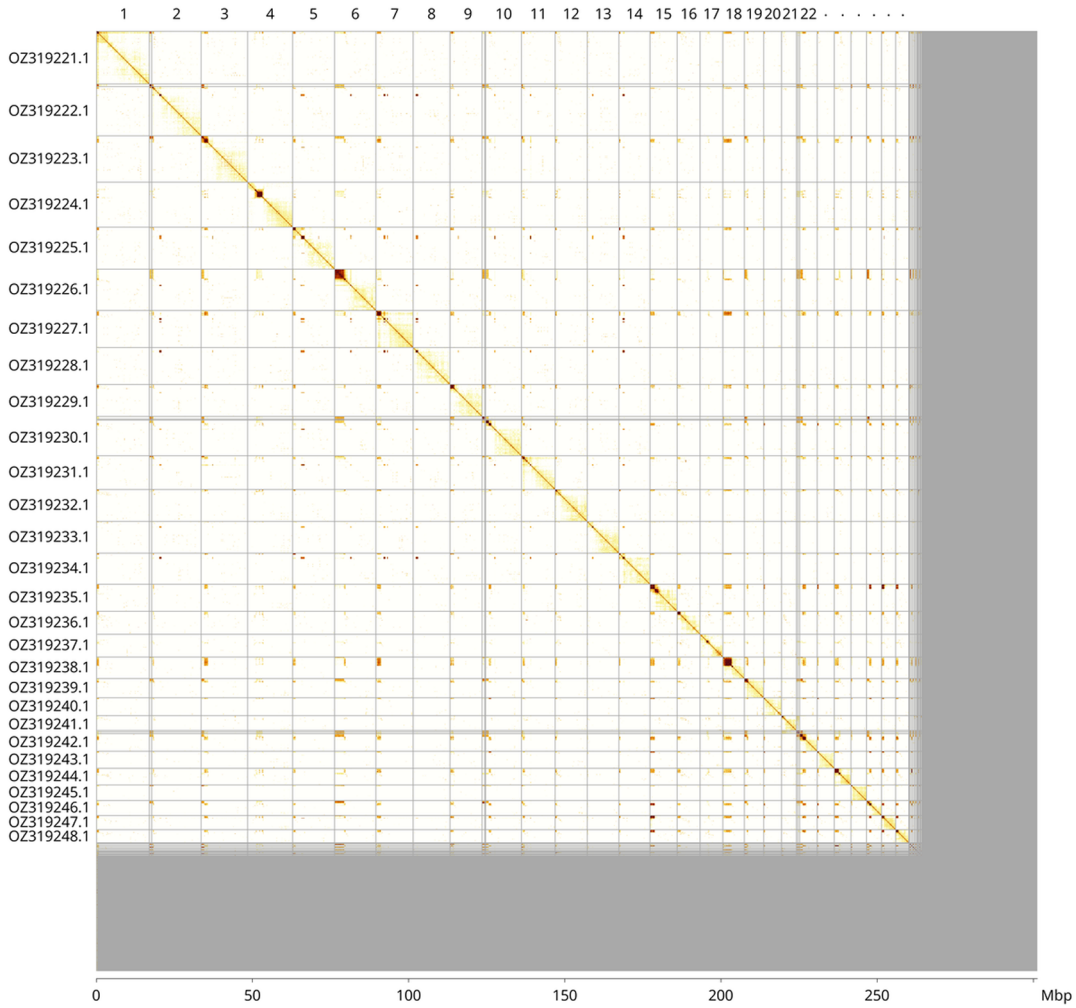

Figure 3. **Hi-C contact map of the *Nysson trimaculatus* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

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 OATK (Zhou *et al.*, 2025).

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Nysson trimaculatus* iyNysTrim1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ319221.1	1	17.80	38.50
OZ319222.1	2	15.84	38.50
OZ319223.1	3	14.86	38.50
OZ319224.1	4	14.37	39.50
OZ319225.1	5	13.46	38.50
OZ319226.1	6	13.24	39.50
OZ319227.1	7	11.91	39.50
OZ319228.1	8	11.84	39
OZ319229.1	9	11.43	39
OZ319230.1	10	11.39	39.50
OZ319231.1	11	10.83	38.50
OZ319232.1	12	10.25	39
OZ319233.1	13	10.18	38
OZ319234.1	14	9.94	39
OZ319235.1	15	8.64	41.50
OZ319236.1	16	7.39	40
OZ319237.1	17	7.28	39.50
OZ319238.1	18	6.90	39.50
OZ319239.1	19	6.17	39
OZ319240.1	20	5.74	38
OZ319241.1	21	5.74	39
OZ319242.1	22	5.66	39.50
OZ319243.1	23	5.46	38.50
OZ319244.1	24	5.34	40
OZ319245.1	25	4.98	39.50
OZ319246.1	26	4.78	38.50
OZ319247.1	27	4.61	41
OZ319248.1	28	4.41	40

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 14 breaks and 88 joins. This reduced the scaffold count by 7.9%, increased the scaffold N50 by 12.8%, and increased the total assembly length by 2.9%. 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 database (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 *Nysson trimaculatus* specimen generated 22.94 Gb (gigabases) from 1.99 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 245.95 Mb, with a heterozygosity of 0.40% and repeat content of 23.05% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 88× coverage. Hi-C sequencing produced 104.54 Gb from 346.17 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

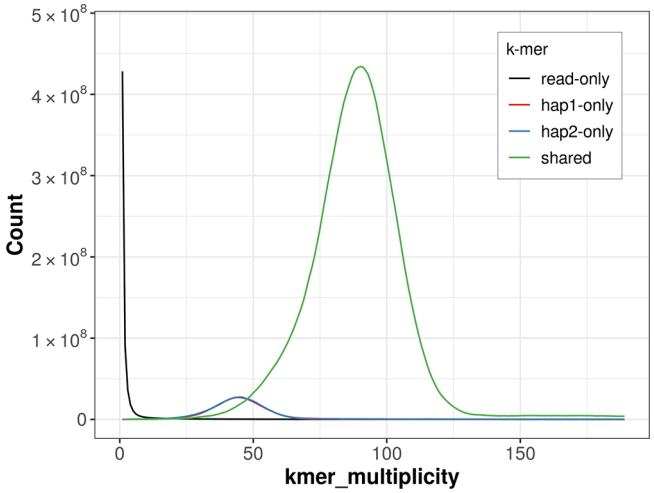

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

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 301.31 Mb in 596 scaffolds, with 201 gaps, and a scaffold N50 of 10.25 Mb (Table 2).

Most of the haplotype 1 assembly sequence (86.44%) was assigned to 28 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome was also assembled (length 25.2 kb, OZ319249.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

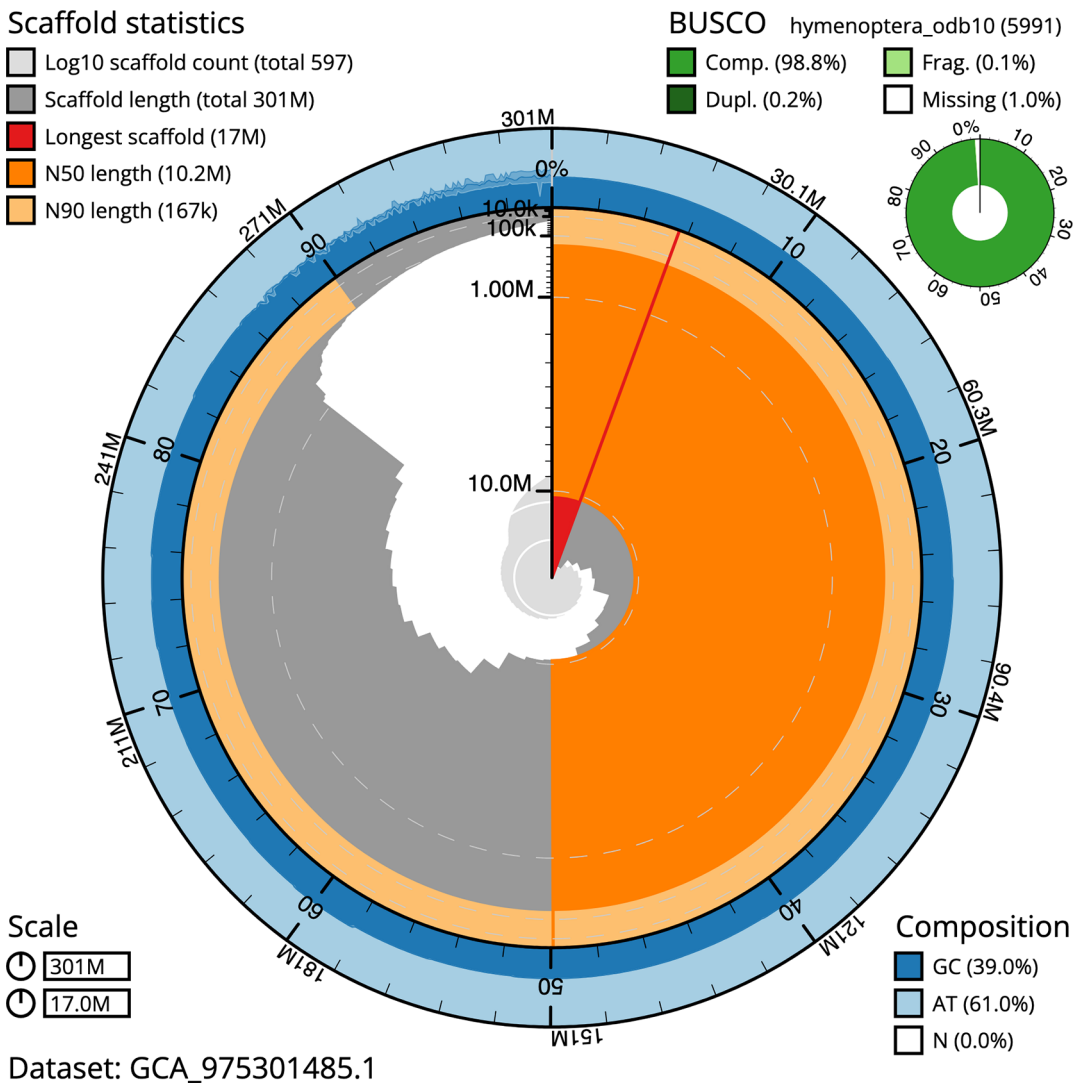

Figure 5. Assembly metrics for iyNysTrim1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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.

Assembly quality metrics

For haplotype 1, the estimated QV is 59.6, and for haplotype 2, 56.4. When the two haplotypes are combined, the assembly achieves an estimated QV of 57.5. The *k*-mer completeness is 92.50% for haplotype 1, 92.91% for haplotype 2, and 99.57% for the combined haplotypes (Figure 4).

BUSCO analysis using the hymenoptera_odb10 reference set (n = 5 991) identified 98.8% of the expected gene set (single = 98.7%, duplicated = 0.2%) in haplotype 1. For haplotype 2, BUSCO v.6.0.0 analysis identified 98.9% of the expected gene set (single = 98.1%, duplicated = 0.8%). 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 January 2026. The EBP metric, calculated for the haplotype 1, is **6.7.Q59**, meeting the recommended reference standard.

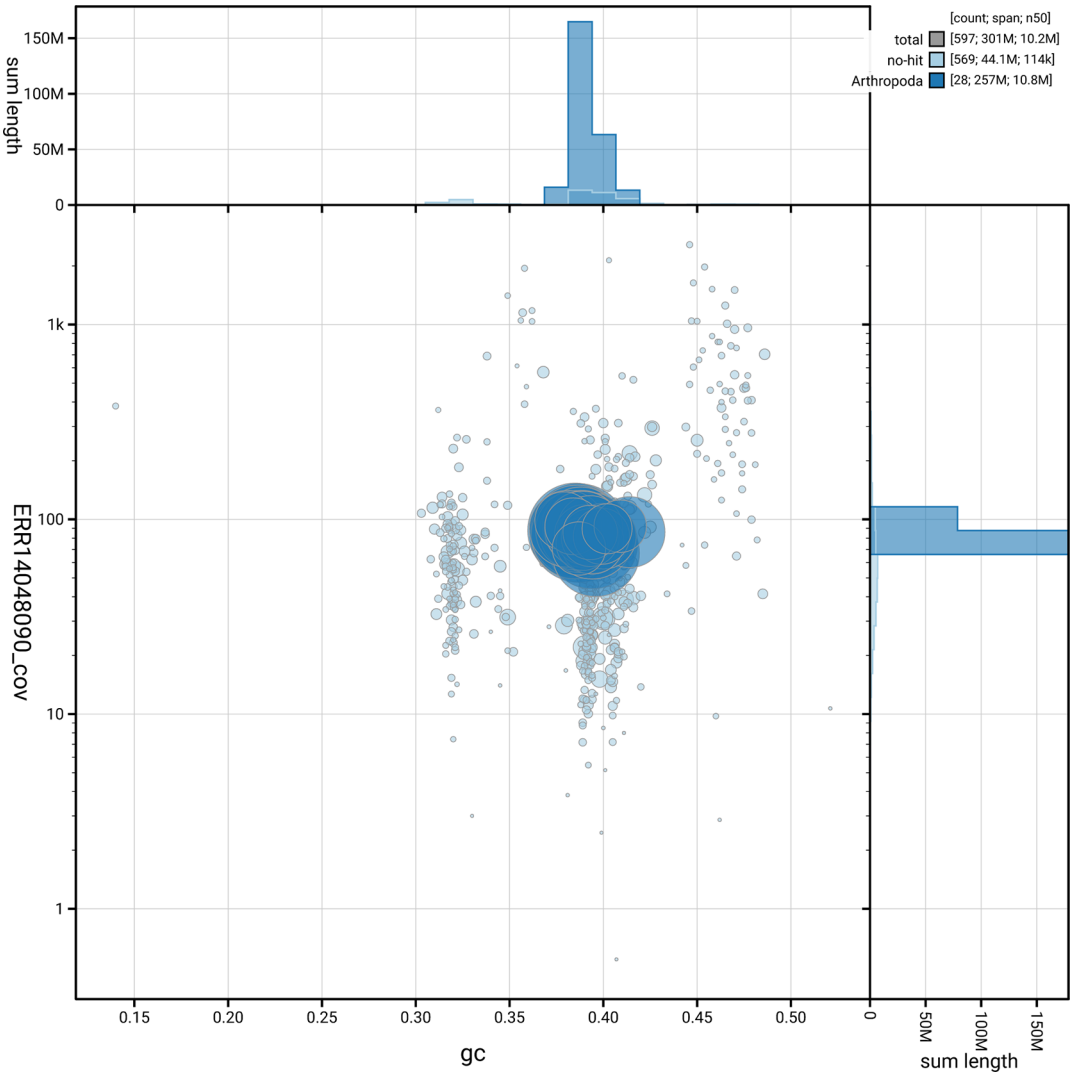

Figure 6. BlobToolKit blob plot for iyNysTrim1.hap1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the 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 *Nysson trimaculatus* assembly.**

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.7.Q59	6.C.Q40
Contig N50 length	1.90 Mb	≥ 1 Mb
Scaffold N50 length	10.25 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 59.6; haplotype 2: 56.4; combined: 57.5	≥ 40
k-mer completeness	Haplotype 1: 92.50%; Haplotype 2: 92.91%; combined: 99.57%	≥ 95%
BUSCO	C:98.8% [S:98.7%, D:0.2%], F:0.1%, M:1.0%, n:5 991	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	86.44%	≥ 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

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	6.0.0	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	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
Oatk	1.0	https://github.com/c-zhou/oatk
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	v0.9.0	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

Data availability

European Nucleotide Archive: Nysson trimaculatus. Accession number PRJEB83146. The genome sequence is released openly for reuse. The *Nysson trimaculatus* 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.

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

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.

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

Benno P: **Enige aantekeningen bij de fenologie van *Lestiphorus bicinctus* (Rossi) en zijn koekoekswesp, *Nysson trimaculatus* (Rossi) (Hym., Sphecidae: Nyssoninae).** *Entomologische Berichten.* 1966; **26**: 7–11.

Bitsch J, Barbier Y, Gayubo SF, *et al.*: *Hyménoptères Spheciformes d'Europe.* Paris: Fédération Française des Sociétés de Sciences Naturelles; 2020. **1**. Faune de France n° 101.

Buchfink B, Reuter K, Drost H-G: **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: *Nysson trimaculatus.* Bees, Wasps & Ants Recording Society; 2026. Accessed 2026-04-13.
Reference Source

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

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat. Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial Arthropod fauna.** *Wellcome Open Research.* 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).
PubMed Abstract | Publisher Full Text | Free Full Text

Evans HE: *The Comparative Ethology and Evolution of the Sand Wasps.* Cambridge, MA: Harvard University Press; 1966.
Publisher Full Text

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

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

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

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.
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

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).
Publisher Full Text

Nemkov PG: **Biological features of the digger wasps of the subfamily Bembicinae (Hymenoptera, Crabronidae).** *A. I. Kurentsov's Annual Memorial Meetings.* 2012; **23**: 114–32.

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

Richards OW: *Scolioidea, Vespoidea and Sphecoidea. Hymenoptera, Aculeata.* London: Royal Entomological Society of London; 1980; **6**.

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

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project.** *Wellcome Open Research.* 2024; **9**: 339.
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.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–24.
Publisher Full Text

Wolf H: **Die parasitische lebensweise der grabwespengattung Nysson Latr. (Hym. Crabronidae).** *Nachrichten der Sammelstelle für Schmarotzerbestimmung.* 1951; **33**: 77–80.

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: A de novo assembly tool for complex plant organelle genomes.** *Genome Biol.* 2025; **26**(235).
Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔ ?

Version 1

Reviewer Report 05 June 2026

https://doi.org/10.21956/wellcomeopenres.29174.r155343

?

Din Syafruddin
Universitas Hasanuddin, Makassar, Indonesia

The manuscript is well written and presented, except for detail chromosomal content (for the chromosomal genome), and gene content for the mitochondrial genome. (it may be accessible through supplementary files, but would probably be informative to summarize it in the abstract. Arrangement of the chromosomes and mitochondrial genome should also be presented as map in graphical abstract

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: Molecular parasitology and 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, however I have significant reservations, as outlined above.

Reviewer Report 03 June 2026

https://doi.org/10.21956/wellcomeopenres.29174.r155344

Marta Coronado-Zamora
Departament de Genetica i de Microbiologia, Universitat Autònoma de Barcelona - UAB Barcelona, Bellaterra, Catalonia, Spain

Vicente Chillida
[1] Departament de Genetica i de Microbiologia, Universitat Autonoma de Barcelona, Bellaterra, Catalonia, Spain
[2] Autonomous University of Barcelona Department of Genetics and Microbiology, Bellaterra, Catalonia, Spain

This data note describes a high quality chromosome assembly for the solitary wasp (*Nysson trimaculatus*). The genome data are available at NCBI BioProject, and data is accessible as checked in May 2026. The genome assembly is highly contiguous, with overall very good assembly metrics. The mitochondrial genome is also included.

Comments:
- The genome currently lacks gene annotations, although it is stated that it will be annotated with available RNA-seq. However, it is not clear if the RNA-seq was produced within this study or it was already available.
- Regarding haplotype 2, why was it assembled to the scaffold level while haplotype 1 was able to be assembled to the chromosome level?
- fasta_windows is mentioned in table 5 but not in the main text
- As DNA concentration was measured with Qubit and Nanodrop, it would be nice for the authors to report 260/280 and 260/230 ratios for DNA purity from Nanodrop

Typos:
- Extra white space in “(Figure 2).“

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: Population genomics, bioinformatics, transposable elements

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

Reviewer Report 01 June 2026

https://doi.org/10.21956/wellcomeopenres.29174.r155345

Sharu Paul Sharma
Iowa State University, Iowa, USA

This manuscript presents a high-quality genome assembly for the solitary wasp (Nysson trimaculatus), generated as part of the Darwin Tree of Life project. The study is clearly written and provides a useful genomic resource for future work.
A 25.2 kb mitochondrial assembly is longer than typical insect mitochondrial genome, so a brief note on completeness, duplicated/control-region content, or circularization would be useful.
Overall, the manuscript is clear and concise with metrics expected of a high quality assembly including high BUSCO score.

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.

Reviewer Report 22 May 2026

https://doi.org/10.21956/wellcomeopenres.29174.r155341

Vahap Eldem
Istanbul University, Istanbul, Turkey

The manuscript presents a chromosome level genome assembly for the solitary wasp Nysson trimaculatus, generated using PacBio HiFi and Hi-C data. Overall, the assembly appears to be of high quality, with high BUSCO completeness and strong consensus accuracy.

The manuscript would benefit from clarifying a minor taxonomic inconsistency. The title, NCBI taxonomy, and species taxonomy section appear to place N. trimaculatus in Crabronidae, whereas the background text states that the species belongs to Bembicidae. This should be clarified.

Overall, this is a valuable genome data note and appears suitable for indexing after minor clarification.

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, Bioimformatics

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.