DATA NOTE

The genome sequence of the Club-Horned Wood Borer Wasp, *Trypoxylon clavicerum* Lepeletier de Saint Fargeau & Audinet-Serville, 1828 (Hymenoptera: Crabronidae)

[version 1; peer review: 2 approved]

Liam M. Crowley[1], Steven Falk[2], Cian D. Williams[3],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]University of Oxford, Oxford, England, UK
[2]Independent researcher, Kenilworth, England, UK
[3]University of Cambridge, Cambridge, England, UK

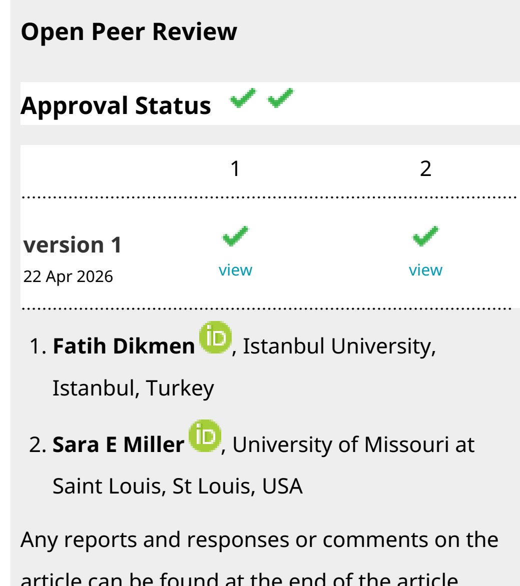
v1 **First published:** 22 Apr 2026, **11**:240
https://doi.org/10.12688/wellcomeopenres.26358.1
Latest published: 22 Apr 2026, **11**:240
https://doi.org/10.12688/wellcomeopenres.26358.1

Abstract

We present a genome assembly from an individual female *Trypoxylon clavicerum* (Club-Horned Wood Borer Wasp; Arthropoda; Insecta; Hymenoptera; Crabronidae). The genome sequence has a total length of 270.89 megabases. Most of the assembly (85.18%) is scaffolded into 9 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 30.97 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces genomes for eukaryotic species found in Britain and Ireland.

Keywords

Trypoxylon clavicerum, Club-Horned Wood Borer Wasp, genome sequence, chromosomal, Hymenoptera



Open Peer Review

Approval Status ✔ ✔

	1	2
version 1 22 Apr 2026	✔ view	✔ view

1. **Fatih Dikmen**, Istanbul University, Istanbul, Turkey
2. **Sara E Miller**, University of Missouri at Saint Louis, St Louis, USA

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

This article is included in the Tree of Life gateway.

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

Author roles: Crowley LM: Investigation, Resources; **Falk S**: 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.

Copyright: © 2026 Crowley LM *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Crowley LM, Falk S, Williams CD *et al.* **The genome sequence of the Club-Horned Wood Borer Wasp, *Trypoxylon clavicerum* Lepeletier de Saint Fargeau & Audinet-Serville, 1828 (Hymenoptera: Crabronidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:240 https://doi.org/10.12688/wellcomeopenres.26358.1

First published: 22 Apr 2026, **11**:240 https://doi.org/10.12688/wellcomeopenres.26358.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Crabronidae; Crabroninae; Trypoxylini; *Trypoxylon*; *Trypoxylon clavicerum* Lepeletier de Saint Fargeau & Audinet-Serville, 1828 (NCBI:txid1124896).

Background

The Club-Horned Wood Borer Wasp, *Trypoxylon clavicerum*, is a solitary wasp in the family Crabronidae. It is relatively small (4.5–8 mm) (Yeo & Corbet, 1995), although females tend to be larger than males. Members of *Trypoxylon* are mostly black all over, and are easily recognised by their long, narrow abdomen, sinuate ("notched") inner margins of the eyes, and distinct neuration of their wings, with only one submarginal cell on the forewing (Saunders, 1896; Yeo & Corbet, 1995). *T. clavicerum* can be identified by reddish-yellow marks on the inner side of its forelegs, although this is not always obvious without a microscope, and the species may be confused with *T. attenuatum.* Antennae are short and distinctly club-shaped, especially in males (Yeo & Corbet, 1995), giving it its common name.

Trypoxylon clavicerum is common and widespread throughout the Palaearctic, including the UK, where its recorded range extends as far north as Edinburgh (GBIF.org, 2025). Although not native to the Americas (Sandhouse, 1940), it was first recorded in the Great Lakes region in the 1980s (Coville, 1984), and it has spread as far as the east coast, although it is not recognised as an invasive species. Adults build their nests in deserted insect burrows, holes in dry wood, or hollow plant stems (Bees, Wasps & Ants Recording Society, 2012). Like many solitary wasps, they divide their nests into cells separated by walls made of clay, with one egg per cell. Female *T. clavicerum* lay ~5 eggs per nest, which they provision with small spiders such as *Araneus* or *Linyphia* (Lomholdt, 1975), often preferring juveniles (Udayakumar *et al.*, 2024), possibly because adults may be too large for *T. clavicerum* to successfully subdue (Bees, Wasps & Ants Recording Society, 2012). Adults can be seen on the wing from late May to early September (Bees, Wasps & Ants Recording Society, 2012; GBIF.org, 2025).

Trypoxylon is an astonishingly diverse genus, with over 600 described species (Sabadini *et al.*, 2020). A recent phylogeny using a small number of molecular markers has already elucidated relationships within the genus (Sabadini *et al.*, 2020). The reference genome presented here will be useful for understanding the genetic basis of this diversity. It will also be useful for resolving broader Hymenopteran taxonomy, as well as investigating the evolution of social behaviour, for which the Hymenoptera are an excellent model system.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult *Trypoxylon clavicerum* (specimen ID Ox000704, ToLID iyTryClav1; Figure 1), collected from Wytham Woods, Oxfordshire, UK (latitude 51.764, longitude −1.325) on 2020-07-24. The specimen was collected and identified by Liam Crowley. A second specimen was used for Hi-C sequencing

Figure 1. Photograph of the *Trypoxylon clavicerum* (iyTryClav1) specimen used for genome sequencing.

(specimen ID Ox002815, ToLID iyTryClav2). It was collected from the same location on 2022-07-14 by Liam Crowley and Steven Falk and identified by Liam Crowley.

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 iyTryClav1 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. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by manual 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. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer’s instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

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 tissue from the iyTryClav2 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 (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) 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 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 ten breaks, 60 joins, and removal of five haplotypic duplications. This reduced the scaffold count by 16.0% and increased the scaffold N50 by 22.0%. 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 the primary and alternate 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 *Trypoxylon clavicerum* specimen generated 21.89 Gb (gigabases) from 1.73 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 107.34 Mb, with a heterozygosity of 50.00% and repeat content of 22.35% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 175× coverage. Hi-C sequencing produced 109.47 Gb from 362.50 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 270.89 Mb in 230 scaffolds, with 93 gaps, and a scaffold N50 of 24.81 Mb (Table 2).

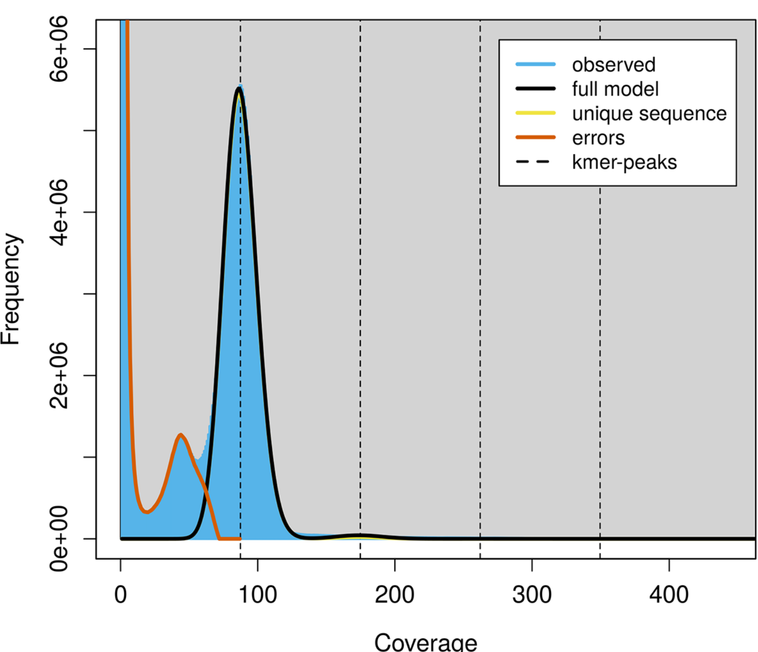


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB55982.

Platform	PacBio HiFi	Hi-C
ToLID	iyTryClav1	iyTryClav2
Specimen ID	Ox000704	Ox002815
BioSample (source individual)	SAMEA7701565	SAMEA113425450
BioSample (tissue)	SAMEA7701763	SAMEA113425547
Tissue	whole organism	whole organism
Instrument	Sequel	Illumina NovaSeq 6000
Run accessions	ERR10224918	ERR12318562
Read count total	1.73 million	362.50 million
Base count total	21.89 Gb	109.47 Gb

Table 2. **Genome assembly data for *Trypoxylon clavicerum*.**

Genome assembly	Primary assembly
Assembly name	iyTryClav1.1
Assembly accession	GCA_965263705.1
Alternate haplotype accession	GCA_965263745.1
Assembly level	chromosome
Span (Mb)	270.89
Number of chromosomes	9
Number of contigs	323
Contig N50	10.06 Mb
Number of scaffolds	230
Scaffold N50	24.81 Mb
Organelles	Mitochondrion: 30.97 kb

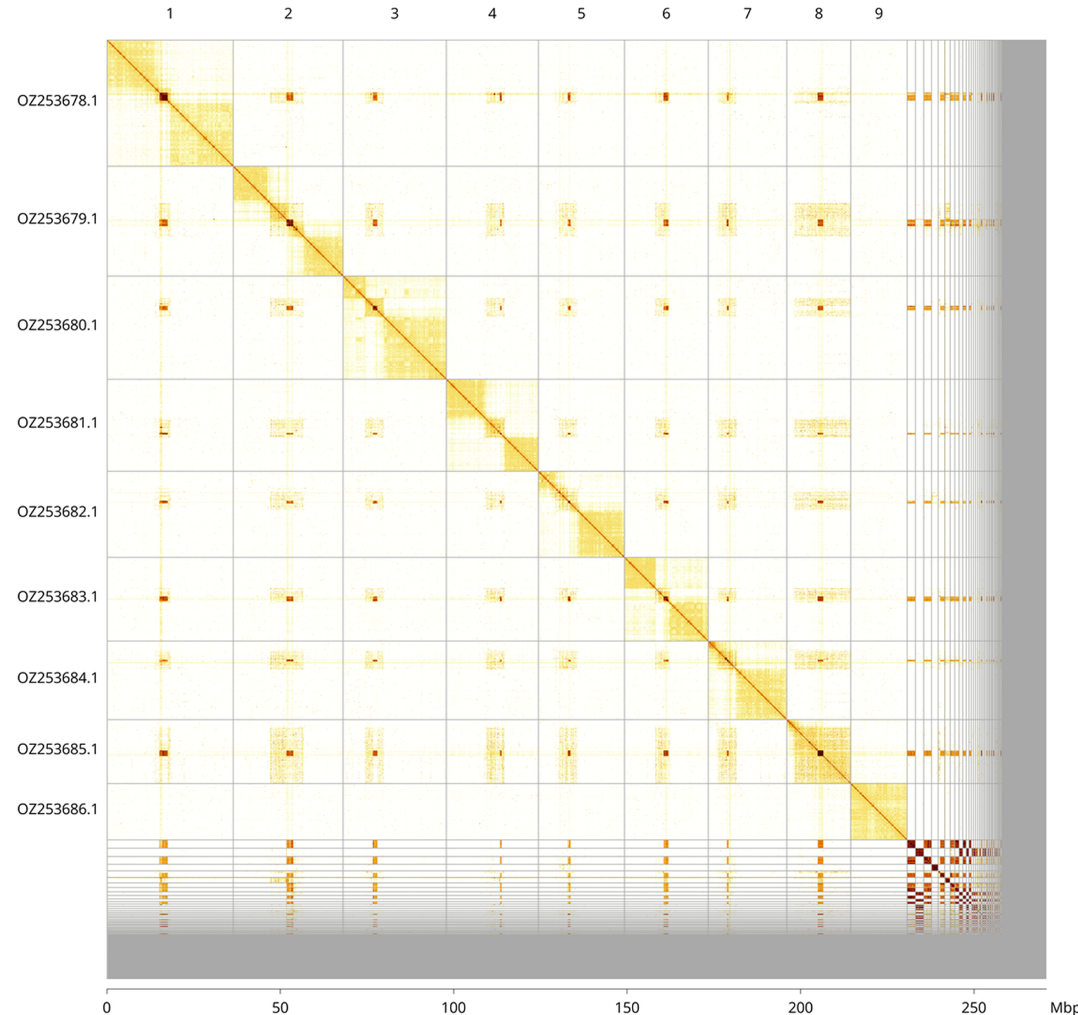


Figure 3. **Hi-C contact map of the *Trypoxylon clavicerum* genome assembly.** Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Trypoxylon clavicerum* iyTryClav1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ253678.1	1	36.52	42.50
OZ253679.1	2	31.61	43
OZ253680.1	3	29.78	43
OZ253681.1	4	26.54	43.50
OZ253682.1	5	24.81	43
OZ253683.1	6	24.16	44
OZ253684.1	7	22.63	42.50
OZ253685.1	8	18.47	43
OZ253686.1	9	16.22	42

Most of the assembly sequence (85.18%) was assigned to 9 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The exact order and orientation of the scaffolds are uncertain in the following regions: chromosome 1 from ~15.2–17.5 Mb, chromosome 2 from ~15.3–17.2 Mb, and chromosome 8 from ~8.9–10.5 Mb.

The mitochondrial genome was also assembled (length 30.97 kb, OZ253687.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 59.2. The *k*-mer completeness is 93.68% for the primary assembly, 72.39% for the alternate haplotype, and 98.73% for the combined assemblies (Figure 4).

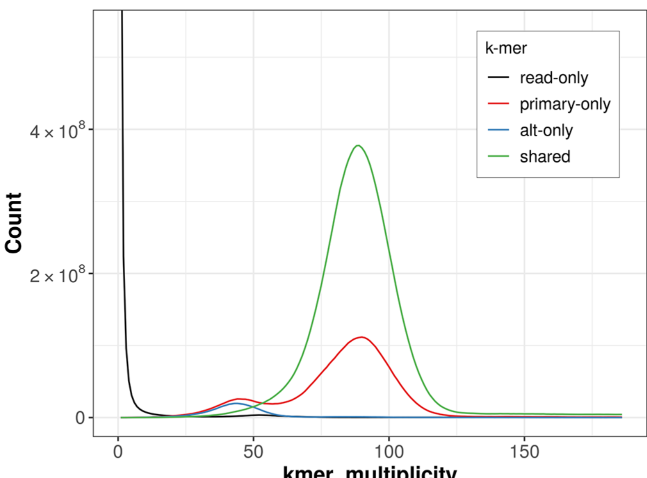


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

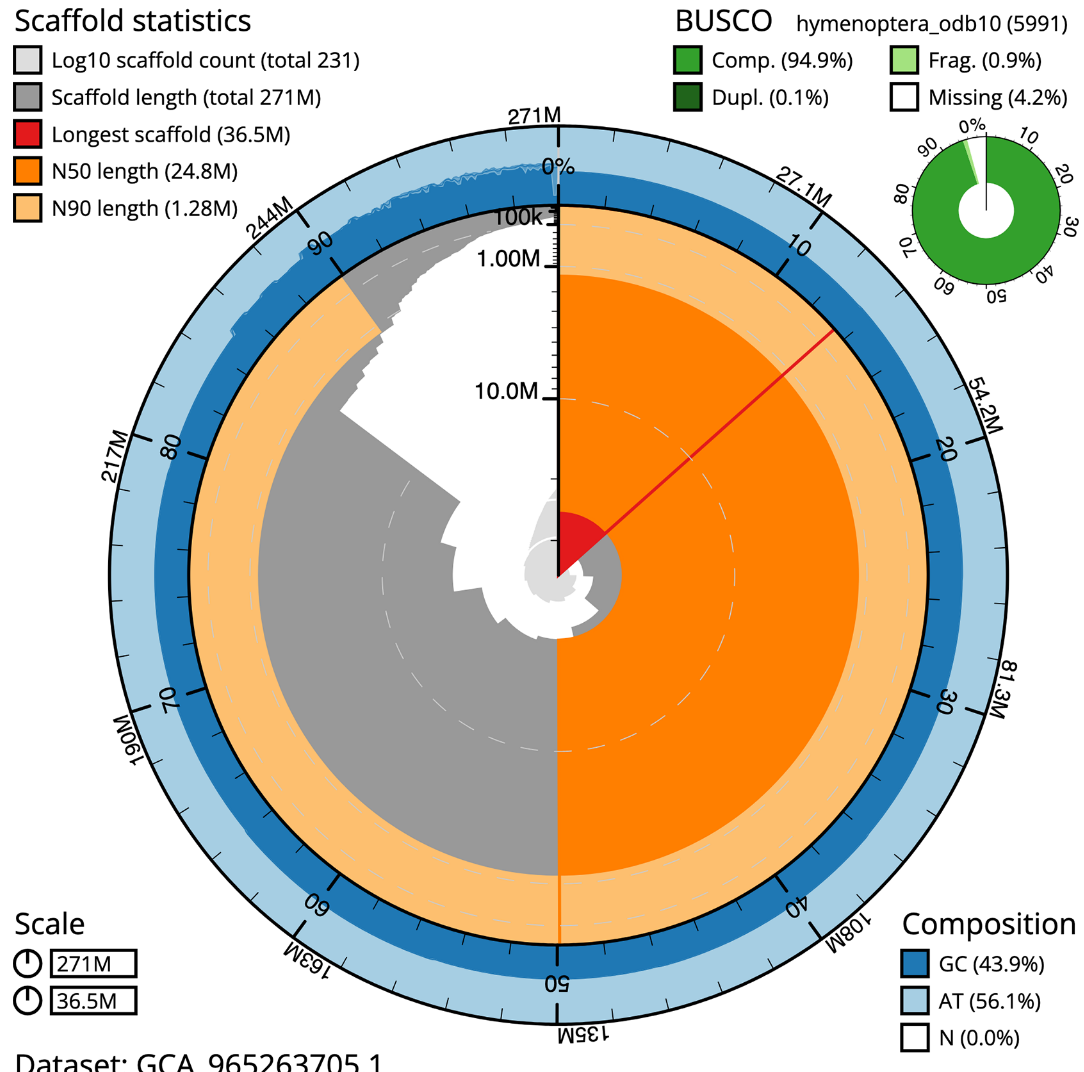


Figure 5. Assembly metrics for iyTryClav1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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.

BUSCO v.5.7.1 analysis using the hymenoptera_odb10 reference set ($n = 5\,991$) identified 94.9% of the expected gene set (single = 94.8%, duplicated = 0.1%). 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) and the Earth BioGenome Project Report on Assembly Standards January 2026. The EBP metric, calculated for the primary assembly, is **7.7.Q59**.

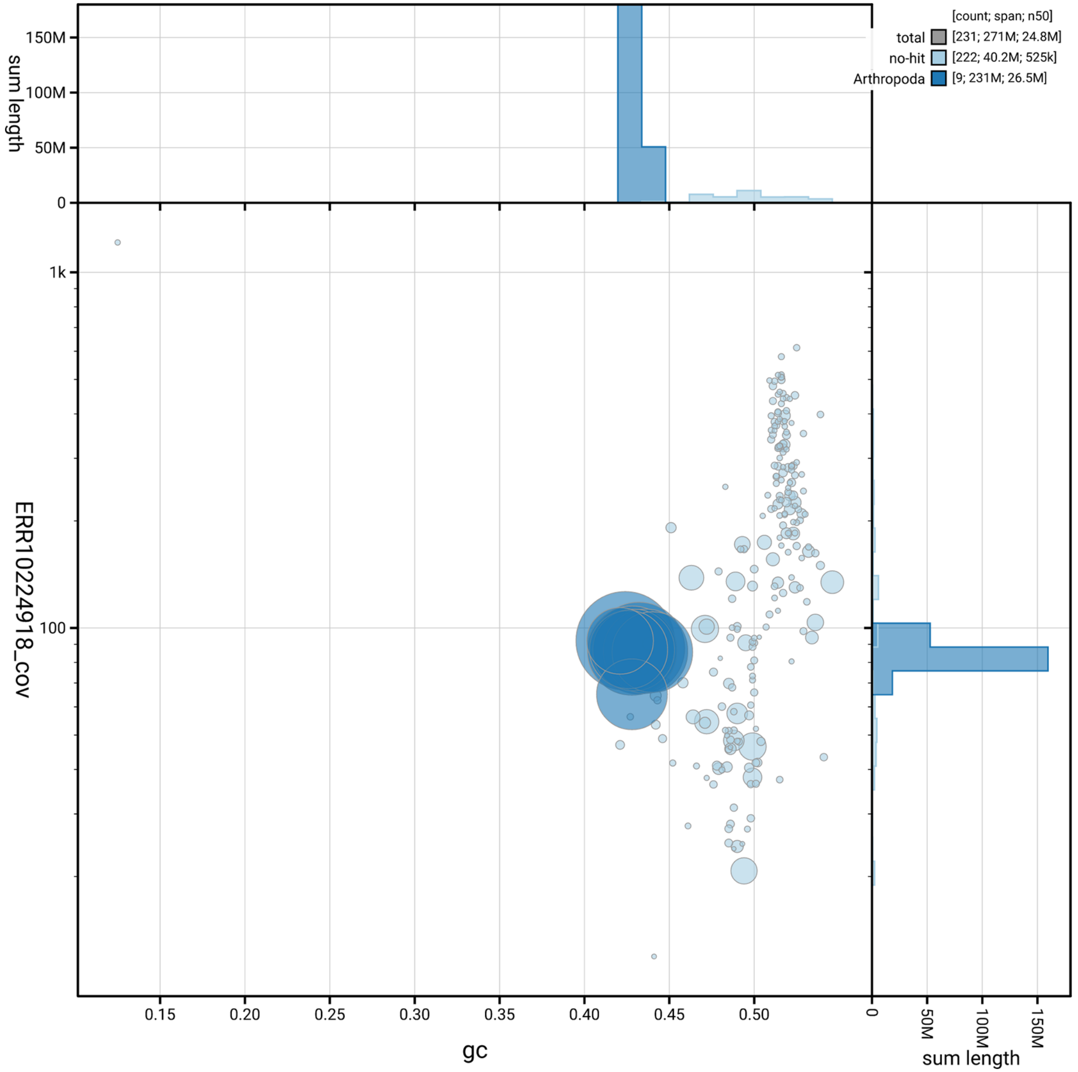


Figure 6. BlobToolKit blob plot for iyTryClav1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total 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 *Trypoxylon clavicerum* assembly.

Measure	Value	Benchmark
EBP summary (primary)	7.7.Q59	6.C.Q40
Contig N50 length	10.06 Mb	≥ 1 Mb
Scaffold N50 length	24.81 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.2; alternate: 59.1; combined: 59.2	≥ 40
k-mer completeness	Primary: 93.68%; alternate: 72.39%; combined: 98.73%	≥ 95%
BUSCO	C:94.9% [S:94.8%, D:0.1%], F:0.9%, M:4.2%, n:5 991	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	85.18%	≥ 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: Trypoxylon clavicerum (club horned wood borer wasp). Accession number PRJEB55982. The genome sequence is released openly for reuse. The *Trypoxylon clavicerum* 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	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	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	0.16.1	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	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3.2	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
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	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

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: ***Trypoxylon clavicerum.*** 2012.
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.
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

Coville RE: **The occurrence of *Trypoxylon clavicerum* in North America (Hymenoptera: Sphecidae).** *Pan Pac. Entomol.* 1984; **60**(3).
Reference Source

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).
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

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

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

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

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

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

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**

Lomholdt O: *The Sphecidae (Hymenoptera) of Fennoscandia and Denmark.* Scandinavian Science Press; 1975.

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

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**

Sabadini CP, Machado CB, Vilhena PDS, *et al.*: **Species delimitation and phylogenetic relationships in the genus *Trypoxylon* (Hymenoptera: Crabronidae) using molecular markers: An alternative to taxonomic impediment.** *Syst. Biodivers.* 2020; **18**(4): 315–327.
Publisher Full Text

Sandhouse GA: **A review of the Nearctic wasps of the genus *Trypoxylon* (Hymenoptera: Sphecidae).** *Am. Midl. Nat.* 1940; **24**(1): 133.
Publisher Full Text

Saunders E: **The Hymenoptera Aculeata of the British Islands: A Descriptive Account of the Families, Genera, and Species Indigenous to Great Britain and Ireland, with Notes as to Habits, Localities, Habitats.** 1896.
Reference Source

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 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 Res.* 2024; **9**(339).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Udayakumar A, Shamappa VH, Kandan A, *et al.*: **First report of *Trypoxylon clavicerum* (Crabronidae: Hymenoptera) from the Indian subcontinent: Its nesting behaviour and prey preference.** *Anim. Biol.* 2024; **74**(3): 309–322.
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 | **PubMed Abstract** | **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; pp. 314–324.
Publisher Full Text

Yeo PF, Corbet SA: *Solitary Wasps.* Richmond Publishing Co.; 1995.

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 04 June 2026

https://doi.org/10.21956/wellcomeopenres.29029.r155524

© **2026 Miller S.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sara E Miller
University of Missouri at Saint Louis, St Louis, Missouri, USA

The authors present a new high quality nuclear and mitochondrial genome assembly for the solitary Club-Horned Wood Borer wasp, *Trypoxylon clavicerum* (Hymenoptera: Crabronidae).

Genome sequencing and assembly were completed using standard Darwin Tree of Life Consortium protocols from a single specimen and scaffolded with Hi-C data from a second specimen. The resulting genome has a length of 270.99 Mb, scaffolded into 9 chromosomes, and is contiguous and high quality with a scaffold N50 of 24.81 Mb and 94.9% coverage of BUSCO genes. No annotation is provided. The final assembly exceed Earth BioGenome Project Assembly Standards benchmarks.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Hymenoptera genetics and genomics, evolutionary ecology

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

Reviewer Report 03 June 2026

https://doi.org/10.21956/wellcomeopenres.29029.r155526

© 2026 Dikmen F. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fatih Dikmen
Istanbul University, Istanbul, Turkey

The Club-Horned Wood Borer Wasp, *Trypoxylon clavicerum*, is a solitary wasp, widespread throughout the Palaearctic and a small bee species belonging to Apoidea. Sequencing the genome of this species may provide an evolutionary baseline for comparative bee genomics, serving as a critical phylogenetic bridge to decode the origins of the clade Anthophila. Because Crabronidae represents one fo the the closest living relatives to modern bees within the superfamily Apoidea, the *de novo* assembly of this species' genome offers a powerful outgroup to trace the profound metabolic and morphological shifts required to transition from a carnivorous, hunting lifestyle to an obligate pollen- and nectar-based herbivory. Furthermore, as an exclusively solitary organism, its genetic architecture serves as a vital reference point for uncovering the molecular roots of eusociality, allowing researchers to isolate the specific gene duplications and regulatory networks that drove the evolution of complex social behaviors in modern bees.

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: Bee taxonomy, genomics, zoology.

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.