DATA NOTE

The genome sequence of the tree wasp, *Dolichovespula sylvestris* Scopoli, 1763 [version 1; peer review: 2 approved]

Steven Falk¹, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Gavin R. Broad², Darwin Tree of Life Consortium

¹Independent Researcher, Kenilworth, Warwickshire, UK
²Department of Life Sciences, Natural History Museum, London, UK

Abstract
We present a genome assembly from an individual male *Dolichovespula sylvestris* (the tree wasp; Arthropoda; Insecta; Hymenoptera; Vespidae). The genome sequence is 233 megabases in span. The majority of the assembly (95.56%) is scaffolded into 26 chromosomal pseudomolecules. The mitochondrial genome was also assembled and is 21.3 kilobases in length.

Keywords
Dolichovespula sylvestris, tree wasp, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status
✓ ✓

version 1
28 Mar 2022

1. **Xu Wang**, Auburn University, Auburn, USA

HudsonAlpha Institute for Biotechnology, Huntsville, USA

2. **Xinhai Ye**, Zhejiang University, Zhejiang, China

Any reports and responses or comments on the article can be found at the end of the article.
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Vespidae; Vespinae; Dolichovespula; Dolichovespula sylvestris Scopoli, 1763 (NCBI:txid85444).

Background
The tree wasp, Dolichovespula sylvestris, is a eusocial vespine wasp that builds relatively small paper nests with a grey envelope in concealed locations, often in cavities or underground, belying its name, although it will build in protected sites in trees (Spredberry, 1973). Nests typically have three or four layers of comb and have a maximum of a little over 200 workers at their peak (Archer, 2012). The colony cycle is short, typically from late May to the end of September, with males produced from late July (Edwards, 1980). Males form mating aggregations (Edwards, 1980), but queens typically mate with one or very few males (Foster & Ratnieks, 2001) meaning that most workers in a nest share the same parentage and are more closely related to their siblings than to the offspring of a fellow worker; worker reproduction is responsible for 50% of male eggs laid in a nest, but almost all of these are destroyed by fellow workers (‘worker policing’) or by the queen (Wenseleers et al., 2005). As with other social wasps in temperate climates, colonies are annual and only the mated queens over-winter, workers and males dying in the autumn. The new nest is constructed by the queen and the cycle begins again.

Dolichovespula sylvestris has a wide range throughout the Palaearctic, reaching 66°N (Archer, 2012). It is a widespread species in Britain and Ireland, including on islands, but is thought to be declining (Edwards & Telfer, 2002). Nests can be usurped throughout much of its range by the socially parasitic Dolichovespula amissa, but this species has not been found in Britain yet. As with other vespines, a wide variety of food is taken, which is mostly made up of insects but can include carrion. Adult wasps are frequent flower visitors and are considered to be important pollinators of common figwort (Scrophularia nodosa) (Proctor et al., 1996).

The evolution of social behaviour in the Hymenoptera is a huge area of research. Until recently, there had been no Vespidae genomes available, so the publication of genomes of species of Vespa (Crowley et al., 2022), Vespula (Crowley et al., 2021) and now Dolichovespula complement the existing genomes of Polistinae (Patalano et al., 2015; Standage et al., 2016), the other vespid subfamily with eusocial societies.

Genome sequence report
The genome was sequenced from a single male D. sylvestris (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.770, longitude -1.331). A total of 27-fold coverage in Pacific Biosciences single-molecule long reads and 175-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 158 missing/misjoins, increasing the assembly size by 3.47%, the scaffold number by 40.00% and the scaffold N50 by 70.68%.

The final assembly has a total length of 233 Mb in 224 sequence scaffolds with a scaffold N50 of 9.6 Mb (Table 1). Of the assembly sequence, 95.56% was assigned to 26 chromosomal-level scaffolds (numbered by sequence length) (Figure 2–Figure 5; Table 2). The assembly has a BUSCO v5.2.2 (Manni et al., 2021) completeness of 96.1% (single 94.4%, duplicated 1.7%) using the hymenoptera_odb10 reference set (n=5991).

Methods
Sample acquisition and DNA extraction
A single male D. sylvestris (iyDolSylv1) was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.770, longitude -1.331) by Steven Falk, independent researcher, from woodland using a net. The sample was identified by the same individual and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The iyDolSylv1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Thorax tissue was cryogenically disrupted to a fine powder using a Covaris cryoPREP Automated Dry Pulveriser, receiving multiple impacts. Fragment size analysis of 0.01–0.5 ng of DNA was then performed using an Agilent FemtoPulse. High molecular weight (HMW) DNA was extracted using the Qiagen Plant MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 200-ng aliquot of extracted DNA using 0.8X AMPure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size between 12–20 kb in a Megaruptor 3 system with
speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8X ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

**Sequencing**

Pacific Biosciences HiFi circular consensus and 10X Genomics Chromium read cloud sequencing libraries were constructed according to the manufacturers’ instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were generated from head tissue of iyDolSylv1 using the Arima v2 kit and sequenced on NovaSeq 6000.

**Genome assembly**

Assembly was carried out with Hifiasm (Cheng et al., 2021). Haplotypic duplication was identified and removed with purge_dups (Guan et al., 2020). One round of polishing was performed by aligning 10X Genomics read data to the assembly with longranger align, calling variants with freebayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao et al., 2014) using SALSA2 (Ghurye et al., 2019). The assembly was checked for contamination as
Figure 2. Genome assembly of Dolichovespula sylvestris, iyDolSylv1.2: metrics. The BlobToolKitSnailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 232,601,616 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (25,955,835 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (9,596,145 and 4,826,817 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the hymenoptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/iyDolSylv1.2/dataset/CAKKNJ02/snail.
Figure 3. Genome assembly of *Dolichovespula sylvestris*, iyDolSylv1.2: GC coverage. BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/iyDolSylv1.2/dataset/CAKKNJ02/blob.

Manual curation was performed using HiGlass (Kerpedjiev et al., 2018) and Pretex. The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva et al., 2021), which performed annotation using MitoFinder (Allio et al., 2020). The genome was analysed and BUSCO scores generated within the BlobToolKit environment (Challis et al., 2020). Table 3 contains a list of all software tool versions used, where appropriate.

Ethics/compliance issues
The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission
**Figure 4.** Genome assembly of *Dolichovespula sylvestris*, iyDolSylv1.2: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the busco/genes/axrules. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/iyDolSylv1.2/dataset/CAKKNJ02/cumulative.

of materials by a Darwin Tree of Life Partner is subject to the Darwin Tree of Life Project Sampling Code of Practice. 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
Figure 5. Genome assembly of *Dolichovespula sylvestris*, iyDolSylv1.2: HI-C contact map. HI-C contact map of the iyDolSylv1.2 assembly, visualised in HiGlass. Chromosomes are shown in size order from left to right and top to bottom. An interactive version of this map is available [here](#).

Table 2. Chromosomal pseudomolecules in the genome assembly of *Dolichovespula sylvestris*, iyDolSylv1.2.

| INSDC accession | Chromosome | Size (Mb) | GC% |
|-----------------|------------|-----------|-----|
| OU964961.1      | 1          | 25.96     | 36.4|
| OU964962.1      | 2          | 13.13     | 36.1|
| OU964963.1      | 3          | 11.49     | 36.2|
| OU964964.1      | 4          | 11.01     | 36.7|
| OU964965.1      | 5          | 10.78     | 37.0|
| OU964966.1      | 6          | 10.61     | 37.6|
| OU964967.1      | 7          | 10.22     | 36.6|
| OU964968.1      | 8          | 9.99      | 36.7|
| OU964969.1      | 9          | 9.98      | 36.9|
| OU964970.1      | 10         | 9.60      | 35.6|
| OU964971.1      | 11         | 9.35      | 36.9|
| OU964972.1      | 12         | 8.38      | 36.0|
| OU964973.1      | 13         | 7.81      | 35.3|
| OU964974.1      | MT         | 5.96      | 35.6|
| OU964975.1      | 14         | 5.75      | 36.7|
| OU964976.1      | 15         | 5.33      | 36.8|
| OU964977.1      | 16         | 5.05      | 36.8|
| OU964978.1      | 17         | 4.84      | 36.8|
| OU964979.1      | 18         | 4.63      | 36.8|
| OU964980.1      | 19         | 4.42      | 36.8|
| OU964981.1      | 20         | 4.22      | 36.8|
| OU964982.1      | 21         | 4.03      | 36.8|
| OU964983.1      | 22         | 3.84      | 36.8|
| OU964984.1      | 23         | 3.65      | 36.8|
| OU964985.1      | 24         | 3.46      | 36.8|
| OU964986.1      | 25         | 3.27      | 36.8|
| OU964987.1      | 26         | 3.08      | 36.8|
| -               | MT         | 0.03      | 17.8|

Unplaced 10.30 36.4
supplied to the Darwin Tree of Life Project. 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**

Underlying data

European Nucleotide Archive: Dolichovespula sylvestris (tree wasp). Accession number PRJEB46852; https://identifiers.org/ena.embl/PRJEB46852.

The genome sequence is released openly for reuse. The *D. sylvestris* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using the 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.

| Software tool | Version | Source |
|---------------|---------|--------|
| Hifiasm       | 0.15.1-r334 | Cheng et al., 2021 |
| purge_dups    | 1.2.3 | Guan et al., 2020 |
| SALSA2        | 2.2 | Ghurye et al., 2019 |
| longranger_align | 2.2.2 | https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines |
| freebayes     | v1.3.1-17-gaa2ace8 | Garrison & Marth, 2012 |
| MitoHiFi      | 2.0 | Uliano-Silva et al., 2021 |
| HiGlass       | 1.11.6 | Kerpedjiev et al., 2018 |
| PretextView   | 0.1.x | https://github.com/wtsi-hpg/PretextView |
| BlobToolKit   | 3.0.5 | Challis et al., 2020 |

**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): 892–902. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Archer ME: Vespine Wasps of the World: Behaviour, Ecology & Taxonomy of the Vespinae. *Siri Scientific Press.* 2012; 4. [Reference Source](#)

Challis R, Richards E, Rojan J, et al.: BlobToolKit–Interactive Quality Assessment of Genome Assemblies. *G3 (Bethesda).* 2020; 10(4): 1361–74. [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): 170–75. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

**Author information**

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: [https://doi.org/10.5281/zenodo.5746938](https://doi.org/10.5281/zenodo.5746938).

Members of the Darwin Tree of Life Barcoding collective are listed here: [https://doi.org/10.5281/zenodo.5744972](https://doi.org/10.5281/zenodo.5744972).

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: [https://doi.org/10.5281/zenodo.6125027](https://doi.org/10.5281/zenodo.6125027).

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: [https://doi.org/10.5281/zenodo.5746904](https://doi.org/10.5281/zenodo.5746904).

Members of the Tree of Life Core Informatics collective are listed here: [https://doi.org/10.5281/zenodo.6125046](https://doi.org/10.5281/zenodo.6125046).

Members of the Darwin Tree of Life Consortium are listed here: [https://doi.org/10.5281/zenodo.5638618](https://doi.org/10.5281/zenodo.5638618).
Open Peer Review

Current Peer Review Status: ✓ ✓

Version 1

Reviewer Report 17 November 2022

https://doi.org/10.21956/wellcomeopenres.19682.r53341

© 2022 Ye X. 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.

Xinhai Ye
Zhejiang University, Zhejiang, China

This genome report presents a chromosome-level genome assembly of the tree wasp, Dolichovespula sylvestris, which is a eusocial wasp in the family Vespidae. The quality of this genome is very high, with a scaffold N50 of 9.6 Mb and 96% BUSCO score. This genomic resource will facilitate comparative genomics studies of insects in the future, especially in the research field of social insects. The manuscript is well-written and the expression is clear. I only have two minor comments:

1. The authors used the SALSA algorithm for Hi-C scaffolding - this method does not require a specified chromosome number, therefore, I'm wondering if there's any other evidence (such as chromosomic staining) of the 26 scaffolded chromosomes in this study.

2. A single male, not female, wasp was used for genome sequencing. I believe that all experts on Hymenoptera know the reason for the authors' choice; however, I'm not sure if the researchers in other fields know the reason as well. Thus, if possible, I suggest a one-sentence note in the Methods.

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 genomics

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

---

Reviewer Report 07 June 2022

https://doi.org/10.21956/wellcomeopenres.19682.r50702

© 2022 Wang X. 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.

**Xu Wang**

1 Department of Pathobiology, College of Veterinary Medicine, Auburn University, Auburn, AL, USA
2 HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

In this manuscript, Falk, Broad, and the Darwin Tree of Life Consortium sequenced and assembled the tree wasp (*Dolichovespula sylvestris*) genome. The authors combined PacBio long-read data, 10x Genomics linked-read data, as well as Hi-C data to achieve super high quality and continuity of the tree wasp genome. This assembly represents the first Vespinae genome and is an excellent genome resource for comparative genomics and evolution in Hymenopterans.

Please find specific comments below:

1. Please add BUSCO 5.2.2 to Table 3. Software tools used.

2. “10X” should be “10x”.

3. “Qiagen Plant MagAttract HMW DNA extraction kit” – I was not aware of a plant-specific version of the MagAttract HMW DNA extraction kit (cat no 67563). Could it be a typo? If not, please justify the use of a plant kit for an insect sample.

4. The 26 pseudochromosomes provide valuable information on hymenopteran chromosome evolution. Could the authors provide some information about the karyotype of the tree wasp? How many chromosomes are there for a haploid male genome?

5. The mitochondrial genome assembly accession number is missing. Please provide in the main text or the data table.

**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?**
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Insect evolution, parasitoid wasps, genomics

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