DATA NOTE

The genome sequence of the common wasp, *Vespula vulgaris* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Liam Crowley

University of Oxford and Wytham Woods Genome Acquisition Lab, 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, Darwin Tree of Life Consortium

1 Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

**Abstract**

We present a genome assembly from an individual female *Vespula vulgaris* (the common wasp; Arthropoda; Insecta; Hymenoptera; Vespidae). The genome sequence is 188 megabases in span. The majority of the assembly is scaffolded into 25 chromosomal pseudomolecules.

**Keywords**

Vespula vulgaris, common wasp, genome sequence, chromosomal

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 L: Investigation, Resources, 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 (206194) and the Darwin Tree of Life Discretionary Award (218328). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Crowley L 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 L, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective et al. The genome sequence of the common wasp, Vespula vulgaris (Linnaeus, 1758) [version 1; peer review: awaiting peer review] Wellcome Open Research 2021, 6:232 https://doi.org/10.12688/wellcomeopenres.17205.1

First published: 14 Sep 2021, 6:232 https://doi.org/10.12688/wellcomeopenres.17205.1
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidae; Vespidae; Vespinae; Vespu; Vespula vulgaris (Linnaeus, 1758) (NCBI:txid7454).

Introduction
The Common wasp, *Vespula vulgaris*, is one of the most widespread species of social wasp in the UK. This species is eusocial, living in colonies with a reproductive queen, sterile workers and reproductive males. Colonies are annual in the UK, typically producing up to around 10,000 workers (*Archer, 1981*). Nests are commonly constructed underground, particularly within old rodent holes, but can also be frequently found in aerial situations such as roof spaces, sheds, outhouses etc. Nests are constructed out of a paper-like substance produced from macerated wood fibres mixed with saliva. The nest consists of hexagonal cells arranged into combs, covered by a nest envelope. The nest envelope of this species is yellow to brown in colour due to the mix of partially rotted wood used to make the pulp.

Overwintered queens emerge from early March, found a nest around May, and produce the first workers from early June (*Archer, 2008*). Worker numbers build up throughout the summer, with males and new queens produced around September. Nests usually last until late October, although exceptionally, some may last through the winter to February.

This species is a generalist predator, with workers preying on a wide range of insect and other arthropod species, which are killed, butchered and malaxed before being carried back to the nest to be fed to the developing brood. Adults feed on carbohydrate rich substances including nectar, sap, honey-dew and secretions from the larvae. The propensity of adults to visit flowers, particularly shallow blooms, means this species may act as an important pollinator. We note the recent spread species of social wasp in the UK. This species is eusocial, living in colonies with a reproductive queen, sterile workers and reproductive males. Colonies are annual in the UK, typically producing up to around 10,000 workers (*Archer, 1981*). Nests are commonly constructed underground, particularly within old rodent holes, but can also be frequently found in aerial situations such as roof spaces, sheds, outhouses etc. Nests are constructed out of a paper-like substance produced from macerated wood fibres mixed with saliva. The nest consists of hexagonal cells arranged into combs, covered by a nest envelope. The nest envelope of this species is yellow to brown in colour due to the mix of partially rotted wood used to make the pulp.

Overwintered queens emerge from early March, found a nest around May, and produce the first workers from early June (*Archer, 2008*). Worker numbers build up throughout the summer, with males and new queens produced around September. Nests usually last until late October, although exceptionally, some may last through the winter to February.

This species is a generalist predator, with workers preying on a wide range of insect and other arthropod species, which are killed, butchered and malaxed before being carried back to the nest to be fed to the developing brood. Adults feed on carbohydrate rich substances including nectar, sap, honey-dew and secretions from the larvae. The propensity of adults to visit flowers, particularly shallow blooms, means this species may act as an important pollinator. We note the recent production of a high-quality genome assembly for *V. vulgaris* and two other species of this genus (*Harrop et al., 2020*), and believe the sequence described here, generated as part of the Darwin Tree of Life project, will further aid understanding of the biology and ecology of the common wasp.

Genome sequence report
The genome was sequenced from a single female *V. vulgaris* collected from Wytham Woods, Oxfordshire, UK (latitude 51.774, longitude -1.332). A total of 87-fold coverage in Pacific Biosciences single-molecule long reads and 190-fold coverage in 10X Genomcs read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 18 missing/misjoins, reducing the assembly length by 0.001% and the scaffold number by 34.15%, and increasing the scaffold N50 by 2.74%. The final assembly has a total length of 188 Mb in 28 sequence scaffolds with a scaffold N50 of 9 Mb (Table 1). Of the assembly sequence, 99.5% was assigned to 25 chromosomal-level scaffolds (numbered by sequence length)

| Table 1. Genome data for *Vespula vulgaris*, iyVesVulg1.1. |
|-------------------------------------------------------------|
| **Project accession data**                                   |
| Assembly identifier | iyVesVulg1.1 |
| Species           | Vespula vulgaris |
| Specimen           | iyVesVulg1 |
| NCBI taxonomy ID   | NCBI:txid7454 |
| BioProject         | PRJE433808 |
| BioSample ID       | SAMEA7520502 |
| Isolate information | Female, head/thorax |

| **Raw data accessions**                                               |
|------------------------------------------------------------------------|
| PacificBiosciences SEQUEL II | ERR6635597 |
| 10X Genomics Illumina                                                  | ERR6054669-ERR6054672 |
| Hi-C Illumina                                                         | ERR6054673-ERR6054675 |
| Illumina PolyA RNA-Seq                                               | ERR6286719 |

| **Genome assembly**                                      |
|-----------------------------------------------------------|
| Assembly accession | GCA_905475345.1 |
| Accession of alternate haplotype | GCA_905404185.1 |
| Span (Mb)         | 188 |
| Number of contigs | 50 |
| Contig N50 length (Mb) | 9 |
| Number of scaffolds | 28 |
| Scaffold N50 length (Mb) | 9 |
| Longest scaffold (Mb) | 20 |
| BUSCO* genome score | C:96.4%, S:96.1%, D:0.3% | J: 0.8%, M:2.8%, n:5991 |

*BUSCO scores based on the hymenoptera_odb10 BUSCO set using v5.1.2. C= complete [S= single copy, D=duplicated], F=fragmented, M=missing, n=number of orthologues in comparison. A full set of BUSCO scores is available at [https://blobtoolkit.genomehubs.org/view/Vespula%20vulgaris/dataset/CAjQ501/busco](https://blobtoolkit.genomehubs.org/view/Vespula%20vulgaris/dataset/CAjQ501/busco). (Figure 1–Figure 4; Table 2). The assembly has a BUSCO completeness of 96.4% using the hymenoptera_odb10 reference set. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

Methods
A single female *V. vulgaris* was collected from Wytham Woods, Oxfordshire, UK (latitude 51.774, longitude -1.332) by Liam Crowley, University of Oxford, using a net. The sample was snap-frozen using dry ice and stored in a CoolRack.

DNA was extracted from head/thorax tissue of iyVesVulg1 at the Wellcome Sanger Institute (WSI) Scientific Operations...
core from the whole organism using the Qiagen MagAttract HMW DNA kit, according to the manufacturer’s instructions. RNA (also from the whole organism) was extracted in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer’s instructions. RNA was then eluted from the same tissue in 50 μl RNAse-free water and its concentration RNA assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay. A further sample from the same tissue was prepped for Hi-C sequencing in the Tree of Life Laboratory using the Arima v2.0 kit.

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud sequencing libraries were constructed according
Figure 2. Genome assembly of *Vespula vulgaris*, lyVesVulg1.1: 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/vespula%20vulgaris/dataset/CAJQFS01/blob.

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) was carried out with SALSA2 (Ghurye et al., 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow et al., 2016) as described previously (Howe et al., 2021). Manual curation was performed using gEVAL, HiGlass.
Figure 3. Genome assembly of *Vespula vulgaris*, iVesVulg1.1: 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 buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Vespula%20vulgaris/dataset/CAJQFS01/cumulative.

(Kerpedjiev et al., 2018) and Pretex. The mitochondrial genome was assembled with MitoHiFi (Uliano-Silva et al., 2021). 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.

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. 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. Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered...
Figure 4. Genome assembly of *Vespula vulgaris*, iyVesVulg1.1: Hi-C contact map. Hi-C contact map of the iyVesVulg1.1 assembly, visualised in HiGlass.

| INSDC accession | Chromosome | Size (Mb) | GC%  |
|-----------------|------------|-----------|------|
| FR997668.1      | 1          | 20.09     | 35.3 |
| FR997669.1      | 2          | 19.27     | 36.3 |
| FR997670.1      | 3          | 13.10     | 35.2 |
| FR997671.1      | 4          | 10.43     | 34.6 |
| FR997672.1      | 5          | 9.59      | 35.2 |
| FR997673.1      | 6          | 9.17      | 35   |
| FR997674.1      | 7          | 9.11      | 34.5 |
| FR997675.1      | 8          | 8.75      | 33.9 |
| FR997676.1      | 9          | 8.61      | 35.1 |
| FR997677.1      | 10         | 8.52      | 35   |
| FR997678.1      | 11         | 7.65      | 33.9 |
| FR997679.1      | 12         | 5.68      | 34.2 |
| FR997680.1      | 13         | 5.44      | 33.1 |
### Table 3. Software tools used.

| Software tool | Version | Source |
|---------------|---------|--------|
| Hifiasm       | 0.12    | 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      | 1.0     | https://github.com/marcelauliano/MitoHiFi |
| gEVAL         | N/A     | Chow et al., 2016 |
| HiGlass       | 1.11.6  | Kerpedjiev et al., 2018 |
| PretextView    | 0.1.x   | https://github.com/wtsi-hpag/PretextView |
| BlobToolKit   | 2.6.1   | Challis et al., 2020 |

### Data availability

European Nucleotide Archive: Vespula vulgaris (common wasp). Accession number PRJEB43808; https://identifiers.org/ena.embl:PRJEB43808.

The genome sequence is released openly for reuse. The *V. vulgaris* 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.
Acknowledgements

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4789929.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893704.

Members of the Wellcome Sanger Institute Tree of Life programme collective are listed here: https://doi.org/10.5281/zenodo.5377053.

References

Archer ME: A Simulation Model for the Colonial Development of Paravespula Vulgaris (Linnaeus) and Dolichovespula Sylvestris (Scopoli) (Hymenoptera: Vespidae). Melanderia. 1981; 36: 1.
Reference Source

Archer ME: Taxonomy, Distribution and Nesting Biology of Species of the Genus Paravespula or the Vespula Vulgaris Species Group (Hymenoptera, Vespidae). Entomologist's Monthly Magazine (United Kingdom). 2008; 144: 5–29.

Challis R, Richards E, Rajan 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

Chow W, Brugger K, Caccamo M, et al.: gEVAL - a Web-Based Browser for Evaluating Genome Assemblies. Bioinformatics. 2016; 32(16): 2508–10. PubMed Abstract | Publisher Full Text | Free Full Text

Garrison E, Marth G: Haplotype-Based Variant Detection from Short-Read Sequencing. arXiv: 1207.3907. 2012. Reference Source

Ghurye J, Rhee A, Walenz BP, et al.: Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly. PLoS Comput Biol. 2019; 15(8): e1007272. PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, et al.: Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies. Bioinformatics. 2020; 36(9): 2896–98. PubMed Abstract | Publisher Full Text | Free Full Text

Harrop TWR, Guhlin J, McLaughlin GM, et al.: High-Quality Assemblies for Three Invasive Social Wasps from the Vespula Genus. G3 (Bethesda). 2020; 10(10): 3479–88. PubMed Abstract | Publisher Full Text | Free Full Text

Howe K, Chow W, Collins J, et al.: Significantly Improving the Quality of Genome Assemblies through Curation. GigaScience. 2021; 10(1): giaa153. PubMed Abstract | Publisher Full Text | Free Full Text

Kerpedjiev P, Abdennur N, Lekschas F, et al.: HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps. Genome Biol. 2018; 19(1): 125. PubMed Abstract | Publisher Full Text | Free Full Text

Rao SS, Huntley MH, Durand NC, et al.: A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014; 158(7): 1665–80. PubMed Abstract | Publisher Full Text | Free Full Text

Simão FA, Waterhouse RM, Ioannidis P, et al.: BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. Bioinformatics. 2015; 31(19): 3210–12. PubMed Abstract | Publisher Full Text

Uliano-Silva M, Nunes JG, Krasheninnikova K, et al.: marcelauliano/MitoHiFi: mitohifi_v2.0. 2021. PubMed Full Text