The genome sequence of *Anoplius nigerrimus* (Scopoli, 1763), a spider wasp [version 1; peer review: 1 approved with reservations]

Steven Falk¹, 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, Gavin 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 *Anoplius nigerrimus* (Arthropoda; Insecta; Hymenoptera; Pompilidae) of unknown sex. The genome sequence is 624 megabases in span. In total, 45.75% of the assembly is scaffolded into 15 chromosomal pseudomolecules. The mitochondrial genome was also assembled and is 17.5 kilobases in length.

Keywords

Anoplius nigerrimus, spider wasp, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.
Species taxonomy
Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Pompiloidae; Pompilidae; Pompilinae; Anoplius; Anoplius nigerrimus (Scopoli, 1763) (NCBI: txid1667466).

Background
Anoplius nigerrimus is one of 43 species of spider-hunting wasp known from Britain and Ireland, where it is widespread and often common (Edwards, 1998). Unlike close relatives, which are more associated with wetlands, A. nigerrimus is found in a variety of habitats (Day, 1988). Females can excavate vertical burrows ending in a widened cell where the egg is laid, typically under stones, or more frequently will utilise existing burrows or cavities, for example in stems or snail shells. Spiders are hunted in vegetation such as grass clumps (Kurczewski, 2010). Prey spiders are stung on the underside of the cephalothorax. The paralysed prey is then dragged by the spider walking backwards to the nest (Kurczewski, 2010). Hosts are often Lycosidae, although other spider families are preyed upon.

An holarctic species, favouring a temperate climate, A. nigerrimus is univoltine, flying mainly in the summer, with males in particular being regular flower visitors. As the name suggests, this is an entirely black pompilid, characterised by strong bristle-like setae on the female 6th metasomal tergite (a characteristic of the genus Anoplius) and a rather triangular 3rd submarginal cell in the fore wing.

This is the first full genome for a species of Pompilidae and as such will make a valuable contribution towards understanding the evolution of aculeate (stinging) Hymenoptera.

Genome sequence report
The genome was sequenced from a single A. nigerrimus (Figure 1) collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK (latitude 51.769, longitude -1.339). A total of 29-fold coverage in Pacific Biosciences single-molecule long reads and 87-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 18 missing/misjoins and removed 1 haplotype duplication, reducing the assembly size by 0.06% and the scaffold number by 1.31%, and increasing the scaffold N50 by 11.90%.

The final assembly has a total length of 624 Mb in 1202 sequence scaffolds with a scaffold N50 of 2.0 Mb (Table 1). Of the assembly sequence, 45.75% was assigned to 15 chromosomal-level scaffolds (numbered by sequence length) (Figure 2–Figure 5; Table 2). The assembly generated is extremely fragmented,

Table 1. Genome data for Anoplius nigerrimus, iyAnoNige1.1.

| Project accession data          | iyAnoNige1.1 |
|--------------------------------|--------------|
| Assembly identifier            | iyAnoNige1.1 |
| Species                        | Anoplius nigerrimus |
| Specimen                       | iyAnoNige1 |
| NCBI taxonomy ID               | NCBI:txid1667466 |
| BioProject                     | PRJEB464296 |
| BioSample ID                   | SAMEA7746764 |
| Isolate information            | Thorax (genome assembly), head (Hi-C) |

Raw data accessions
PacificBiosciences SEQUEL II ERR6939224
10X Genomics Illumina ERR6363332-ERR6363335
Hi-C Illumina ERR6363336

Genome assembly
Assembly accession GCA_914767735.1
Accession of alternate haplotype GCA_914767785.1
Span (Mb) 624
Number of contigs 1243
Contig N50 length (Mb) 1.7
Number of scaffolds 1202
Scaffold N50 length (Mb) 2.0
Longest scaffold (Mb) 25.2

BUSCO* genome score C:95.0%, S:94.4%, D:0.6%, F:1.2%, M:3.8%, n:5991

*BUSCO scores based on the hymenoptera_odb10 BUSCO set using v5.2.2. 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/iyAnoNige1.1/dataset/CAJZBX01/busc.
making assigning sequence scaffolds to chromosomes very difficult. The assembly has a BUSCO v5.2.2 (Manni et al., 2021)
completeness of 95.0% (single 94.4%, duplicated 0.6%) using the hymenoptera_odb10 reference set (n=5991). While not fully
phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

**Methods**

**Sample acquisition and nucleic acid extraction**
An *A. nigerrimus* of unknown sex was collected from Wytham Woods, Oxfordshire (biological vice-county: Berkshire), UK
(latitude 51.769, longitude -1.339) from woodland by Steven Falk, Independent Researcher, using a net. The samples were
identified by the same individual and snap-frozen on dry ice.

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute. The iyAnoNige1 sample was weighed and
dissected on dry ice with tissue set aside for Hi-C sequencing. Thorax tissue was disrupted using a Nippi Powermasher fitted
with a BioMasher pestle. 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

---

**Figure 2. Genome assembly of *Anoplius nigerrimus*, iyAnoNige1.1: metrics.** The BlobToolKit Snailplot 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 623,801,046 bp assembly. The distribution of chromosome lengths is shown in dark grey with the plot radius scaled to the longest chromosome present in the assembly (25,181,928 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 chromosome lengths (2,031,098 and 210,717 bp), respectively. The pale grey spiral shows the cumulative chromosome 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/iyAnoNige1.1/dataset/CAJZBX01/snail](https://blobtoolkit.genomehubs.org/view/iyAnoNige1.1/dataset/CAJZBX01/snail).
Qiagen 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 read cloud sequencing libraries were constructed according to the manufacturers’ instructions. Sequencing was performed by the Scientific Operations core at the Wellcome Sanger Institute.

**Figure 3. Genome assembly of Anoplius nigerrimus, iyAnoNige1.1. 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/iyAnoNige1.1/dataset/CAJZBX01/blob](https://blobtoolkit.genomehubs.org/view/iyAnoNige1.1/dataset/CAJZBX01/blob)
on Pacific Biosciences SEQUEL II and Illumina NovaSeq 6000 instruments. Hi-C data were generated from head tissue of iyAnoNige1 using the Arima v2.0 kit and sequenced on an Illumina NovaSeq 6000 instrument.

**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 described previously (Howe et al., 2021). Manual curation (Howe et al., 2021) 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.

---

**Figure 4. Genome assembly of Anoplius nigerrimus, iyAnoNige1.1: 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/iyAnoNige1.1/dataset/CAJZBX01/cumulative.
Table 2. Chromosomal pseudomolecules in the genome assembly of *Anoplius nigerrimus*, iyAnoNige1.1.

| INSDC accession | Chromosome | Size (Mb) | GC%  |
|-----------------|------------|-----------|------|
| OU612077.1      | 1          | 17.15     | 44.5 |
| OU612078.1      | 2          | 25.18     | 46.6 |
| OU612079.1      | 3          | 23.71     | 46.4 |
| OU612080.1      | 4          | 22.51     | 47.2 |
| OU612081.1      | 5          | 21.05     | 43.0 |
| OU612082.1      | 6          | 19.00     | 48.0 |
| OU612083.1      | 7          | 18.71     | 46.1 |
| OU612084.1      | 8          | 18.01     | 47.6 |
| OU612085.1      | 9          | 17.72     | 48.2 |
| OU612086.1      | 10         | 17.69     | 43.7 |
| OU612087.1      | 11         | 16.52     | 42.3 |
| OU612088.1      | 12         | 16.10     | 45.9 |
| OU612089.1      | 13         | 16.06     | 46.3 |
| OU612090.1      | 14         | 15.78     | 43.8 |
| OU612091.1      | 15         | 11.82     | 41.7 |
| OU612092.1      | MT         | 0.02      | 20.5 |
| -               | Unplaced   | 346.78    | 41.6 |

Figure 5. Genome assembly of *Anoplius nigerrimus*, iyAnoNige1.1: Hi-C contact map. Hi-C contact map of the iyAnoNige1 assembly, visualised in HiGlass. Chromosomes are shown in size order from left to right and top to bottom. The interactive Hi-C map can be viewed here.
### Table 3. Software tools used.

| Software tool       | Version | Source                                                                 |
|---------------------|---------|------------------------------------------------------------------------|
| Hifiasm             | 0.15.2  | 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       | https://github.com/marcelauliano/MitoHiFi                             |
| HiGlass             | 1.11.6  | Kerpedjiev et al., 2018                                               |
| PretextView         | 0.2.x   | https://github.com/wtsi-hpag/PretextView                              |
| BlobToolKit         | 3.0.5   | Challis et al., 2020                                                  |

### Ethics/compliance issues

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 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: Anoplus nigerrimus. Accession number PRJEB46296; https://identifiers.org/ena.embl/PRJEB46296.

The genome sequence is released openly for reuse. The A. nigerrimus 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 and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

### Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: 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.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: 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.

Members of the Tree of Life Core Informatics collective are listed here: 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.

### References

Allio R, Schomaker-Bastos A, Romiguier J, et al.: **MitoFinder: Efficient Automated Large-Scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics.** Mol Ecol Resour. 2020; 20(4): 892–905. PubMed Abstract | Publisher Full Text | Free Full Text

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 Nova**
Assembly Using Phased Assembly Graphs with Hiiasm. Nat Methods. 2021; 18(2): 170–75.

PubMed Abstract | Publisher Full Text | Free Full Text

Day MC: Handbooks for the Identification of British Insects. Royal Entomological Society, 1988.

Reference Source

Edwards R: Provisional Atlas of the Aculeate Hymenoptera of Britain and Ireland: Part 2. Institute of Terrestrial Ecology, 1998.

Reference Source

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

Reference Source

Ghurye J, Rhie A, Walenz BP, et al.: Integrating Hi-C Links with Assembly Graphs for Chromosome-Scale Assembly. PLoS Comput Biol. 2019; 15(8):

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):

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, Abadpour N, Lekitsch F, et al.: HiGlass: Web-Based Visual Exploration and Analysis of Genome Interaction Maps. Genome Biology. 2018; 19(1): 125.

PubMed Abstract | Publisher Full Text | Free Full Text

Kurczewski FE: Prey and Nesting Behavior of Some North American Spider Wasps (Hymenoptera: Pompilidae). Northeast Nat. 2010; 17(1): 115–24.

Publisher Full Text

Manni M, Berkeley MR, Segpey 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): 4647–54.

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; 158(7): 1665–80.

PubMed Abstract | Publisher Full Text | Free Full Text

Uliano-Silva M, Nunes JGF, Krasheninnikova K, et al.: mitohifi_v2.0. 2021.

Publisher Full Text
Open Peer Review

Current Peer Review Status:  

© 2022 Michell C. 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.

Craig Michell  
Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland

Firstly, thank you for the opportunity to review this manuscript. In the manuscript titled “The genome sequence of *Anoplius nigerrimus* (Scopoli, 1763), a spider wasp” the authors, Falk and Broad, present the genome of *Anoplius nigerrimus* (Scopoli, 1763) as produced by utilising several advance sequencing technologies.

Discussion of the manuscript

Background:  
I have no major issues with the background discussion of the insect. I would prefer that the final sentence is expanded, I feel that “… will make a valuable contribution towards understanding the evolution of aculeate Hymenoptera” is relatively ambiguous. What type of contribution do the authors mean? For example, the evolution of venom? This type of discussion would make it easier to understand the rationale behind sequencing this species. Currently it feels like it was in the freezer and available.

Genome sequence report:  
The bare facts of the samples collection, genome sequencing and genome statistics are stated. The data availability information is listed in Table 1 as required. When was the insect collected? Why was it not sexed? Not knowing the sex can make sequencing more difficult in Hymenoptera as many employ a haplodiploid sex determination system. Resulting in many Hymenopteran genome projects using haploid males as they are easier to assemble. I am concerned by “the assembly generated is extremely fragmented”. I would like further information about this, for example, why is the genome extremely fragmented? is it because of repetitive regions? Failed library preparation? Small insert size for the long reads / 10X reads? and so on. This is especially concerning when sequencing projects with similar genome coverage have much better contiguity. In Figure 1., the insect is out of focus. I would recommend using a better image in its place of the actual insect and not the in-focus collection vial.
Methods:
What year was the insect collected? Why was it not sexed? How was it identified? Now I understand that many of the genomes are being published by the DTOL consortia and that there is likely a certain recipe for the manuscripts, but I do not like that they are basically word for word copies (just observed by going through other submitted articles) between each manuscript with only the species and sample accession replaced. Not only is this the case but it seems to perpetuate systematic errors. Meaning that although the bare minimum of the methods is reported it is difficult to understand and lacks any substance.

Reading the DNA extraction section was confusing. The Authors use three different technologies to generate this genome assembly, PacBio HiFi reads, 10X genomics and Hi-C. Logically this section would be split into HMW techniques (HiFi and 10X separated too) and then Hi-C. This would aid readability. The section starts with the insect being dissected, please be explicit in which tissue was kept back for Hi-C. I know that it was the head because you mention it a few paragraphs later. But it is unclear at the start where you are explaining about the sample dissection, what tissue is being used for which library preparation. The thorax was disrupted and then DNA was measured from the homogenized sample? To what end? Then it moves on to the HMW DNA extraction, maybe the sentence about measuring DNA in the homogenized sample was supposed to be here? Please also report the DNA extraction outcomes. This could be why your genome is “extremely fragmented”. Could you explain which DNA was used for which library? Was the HMW DNA cleaned up using 0.8 X Ampure beads used for the 10X chromium sequencing? If so, how was the 50ng of DNA input determined, the input for 10X chromium sequencing is usually decided by the genome size of the organism. Was the HMW DNA of 12-20kbp used for the PacBio HiFi reads? Did this work, please show the library size determined by fragment analysis. Why is there no mention of the Hi-C library preparation here, but rather in the sequencing section. How were the cells fixed? Any brief description of the library preparation would be good. The Arima v2.0 kit is just for the Hi-C, which kit was used for the library preparation KAPA, Swift biosciences? These are all important details to include here, so that the research can be properly interpreted and repeated.

Sequencing: Which versions of protocols were used? Again this is a word-for-word copy from previous DTOL genome announcements. But still it is confusing, HiFi and 10x are both listed as using the manufacturers instructions. This is too ambiguous to be repeatable. Split this up and mention which version of the protocol was used. How were the 10X and Hi-C libraries sequenced, together, separate? How much PhiX was used? How was demultiplexing and basecalling performed?

Genome assembly:
The methods used for the assembly are basic and again a cookie cutter one size fits all, but this is not usually the case with genome assembly projects. Why were the 10X reads only used for polishing the genome? Did you assemble the 10X reads on their own? Did you try other assemblers for the long reads, Can u for example? Maybe you would have a less fragmented genome that could be scaffolded better? Why not try to assemble using the 10X reads also? Scaff10X, ARKS or physlr? You say that you assembled the mitochondrial genome and that it was annotated, yet it is not presented anywhere in the genome report. Maybe this could be included as a panel in Fig 2.
It always strikes me as odd reading a DTOL data output that contains no phylogenetic tree. The BUSCOs identified from your species could easily be used to generate a basic phylogeny with any related insect to validate its phylogenetic position and likely with better accuracy than just Athropoda (Fig. 3. and 4.). The Hi-C contact map in Figure 5 clearly shows the 45% of contigs being placed into 15 linked groups. Is this what was expected, have the 15 chromosomes been reported before? What is the other 55%? is this caused by technical issues in the Hi-C library production? Is it due to a huge amount of repetitive regions in the genome? An explanation of this would be useful in the genome. To be truly reproducible, I would like to see the commands used to assemble and analyse the genome in Table 3.

References
1. Branstetter MG, Childers AK, Cox-Foster D, Hopper KR, et al.: Genomes of the Hymenoptera. Curr Opin Insect Sci. 25: 65-75 PubMed Abstract | Publisher Full Text

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

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

Are sufficient details of methods and materials provided to allow replication by others?
No

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, Molecular genomics, Evolution, Ecology, and Insect 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, however I have significant reservations, as outlined above.