



DATA NOTE

The genome sequence of a cluster fly, *Pollenia pediculata*

Macquart, 1834

[version 1; peer review: 2 approved]

Steven Falk¹, Liam M. Crowley ², Olga Sivell 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

¹Independent researcher, Kenilworth, England, UK

²University of Oxford, Oxford, England, UK

³Natural History Museum, London, England, UK

V1 First published: 11 Apr 2025, 10:190
<https://doi.org/10.12688/wellcomeopenres.23961.1>
Latest published: 11 Apr 2025, 10:190
<https://doi.org/10.12688/wellcomeopenres.23961.1>

Abstract

We present a genome assembly from a female specimen of *Pollenia pediculata* (cluster fly; Arthropoda; Insecta; Diptera; Polleniidae). The assembly contains two haplotypes with total lengths of 1,156.86 megabases and 1,222.48 megabases. Most of haplotype 1 (99.79%) is scaffolded into 6 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 15.82 kilobases.

Keywords

Pollenia pediculata, cluster fly, genome sequence, chromosomal, Diptera



This article is included in the [Tree of Life](#) gateway.

Open Peer Review

Approval Status  

	1	2
version 1 11 Apr 2025	 view	 view
1. Arun Arumugaperumal  , Rajalakshmi Engineering College, Chennai, India		
2. Panagiotis Ioannidis  , Foundation for Research & Technology - Hellas, Crete, Greece		
Any reports and responses or comments on the article can be found at the end of the article.		

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

Author roles: Falk S: Investigation, Resources; Crowley LM: Investigation, Resources; Sivell O: 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: © 2025 Falk S *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: Falk S, Crowley LM, Sivell O *et al.* **The genome sequence of a cluster fly, *Pollenia pediculata* Macquart, 1834 [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:190 <https://doi.org/10.12688/wellcomeopenres.23961.1>

First published: 11 Apr 2025, 10:190 <https://doi.org/10.12688/wellcomeopenres.23961.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Eremoneura; Cyclorrhapha; Schizophora; Calyptratae; Oestroidea; Polleniidae; *Pollenia*; *Pollenia pediculata* Macquart, 1834 (NCBI:txid1266492)

Background

Pollenia pediculata Macquart, 1834 is a species from the family Polleniidae. It was previously included in Calliphoridae as the subfamily Polleniinae (Rognes, 1991a; van Emden, 1954), but was recently raised to family status (Cerretti *et al.*, 2019; Kuty *et al.*, 2019; Yan *et al.*, 2021). *Pollenia* Robineau-Desvoidy 1830 is the only genus from this family occurring in Britain (Chandler, 2024; Gisondi *et al.*, 2020; Sivell, 2021). They are easily distinguished by golden crinkly hairs covering the thorax (Draber-Moňko, 2004; Rognes, 1991a; Sivell, 2021; van Emden, 1954). *P. pediculata* is a fairly dark species, with a yellow basicosta and black palps. It can be easily identified by the tuft of pale hairs on the node at the junction of the subcostal (Sc) and humeral (h) veins on the ventral side of the wing (Draber-Moňko, 2004; Rognes, 1991a; Sivell, 2021).

Pollenia pediculata is widely distributed in the Palaearctic, it also occurs in India and Pakistan and has been introduced into the Nearctic (Canada and USA), Afrotropical (South Africa), Neotropical (Bahamas) and Australasian (New Zealand) Regions (Draber-Moňko, 2004; Gisondi *et al.*, 2020; Rognes, 1987; Rognes, 1991a). It is widespread in Europe and common in Britain (van Emden, 1954; Rognes, 1991a; Draber-Moňko, 2004; Sivell, 2021; Gisondi *et al.*, 2020).

Pollenia pediculata is oviparous and the larvae of this species are parasitoids of the earthworm *Eisenia rosea* (Savigny, 1826) (Rognes, 1991a). According to Rognes (1987), the larvae described by Yahnke and George (1972) were most likely *P. pediculata* and not *P. rudis*. The puparium of *P. pediculata* was described by Rognes (1987). The egg was described by Grzywacz *et al.* (2012) and the first instar larva was described by Szpila (2003). The adults feed on flowers and are important pollinators, they are also attracted to fish bait and can be found feeding on carrion, although they do not breed in it (Jewiss-Gaines *et al.*, 2012; Marshall, 2012; Martin-Vega & Baz, 2013; Rognes, 1991a; Szpila, 2017). *Pollenia* overwinter as adults and can cause a nuisance when aggregating indoors in large numbers. This behaviour has earned them the common name of cluster flies.

The phylogeny of *Pollenia* has been studied by Rognes (1987, 1988, 1991a, 1991b, 1992a, 1992b, 2010) who proposed a number of species groups based on morphology. Rognes (1987) placed *P. pediculata* (as *P. pseudorudis* Rognes, 1985) in the *rudis* species group together with *P. angustigena* Wainwright, 1940, *P. hungarica* Rognes, 1987, *P. longithecata* Rognes, 1987, *P. luteovillosa* Rognes, 1987 and *P. rudis* (Fabricius, 1794).

A molecular study by Szpila *et al.* (2023) showed high phylogenetic support for the *rudis* species group. This study also expressed the need for testing the phylogenetic hypotheses using new-generation sequencing (Szpila *et al.*, 2023). The high-quality genome of *P. pediculata* presented here as well as other *Pollenia* genomes sequenced as part of the Darwin Tree of Life Project (Falk *et al.*, 2023a; Falk *et al.*, 2023b; Falk *et al.*, 2024; Sivell *et al.*, 2025) will aid this research.

The high-quality genome of *Pollenia pediculata* was sequenced from a single specimen from Wytham Woods, Oxford, England. The genome was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland.

Genome sequence report

Sequencing data

The genome of a specimen of *Pollenia pediculata* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 41.14 Gb (gigabases) from 3.99 million reads. GenomeScope analysis of the PacBio HiFi data estimated the haploid genome size at 1,074.95 Mb, with a heterozygosity of 1.75% and repeat content of 50.91%. These values provide an initial assessment of genome complexity and the challenges anticipated during assembly. Based on this estimated genome size, the sequencing data provided approximately 36.0x coverage of the genome. Chromosome conformation Hi-C sequencing produced 121.69 Gb from 805.91 million reads. Table 1 summarises the specimen and sequencing information.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The assembly was improved by manual curation, which corrected 79 misjoins or missing joins. These interventions decreased the scaffold count by 13.79%. The final assembly has a total length of 1,156.86 Mb in 49 scaffolds, with 152 gaps, and a scaffold N50 of 221.33 Mb (Table 2).



Figure 1. Photograph of the *Pollenia pediculata* (idPolPedi2) specimen used for genome sequencing.

Table 1. Specimen and sequencing data for *Pollenia pediculata*.

Project information			
Study title	Pollenia pediculata (tufted clusterfly)		
Umbrella BioProject	PRJEB81200		
Species	<i>Pollenia pediculata</i>		
BioSpecimen	SAMEA113425778		
NCBI taxonomy ID	1266492		
Specimen information			
Technology	ToLID	BioSample accession	Organism part
PacBio long read sequencing	idPolPedi2	SAMEA113425962	whole organism
Hi-C sequencing	idPolPedi1	SAMEA113425504	head and thorax
Sequencing information			
Platform	Run accession	Read count	Base count (Gb)
Hi-C Illumina NovaSeq 6000	ERR13802793	8.06e+08	121.69
PacBio Revio	ERR13800662	3.99e+06	41.14

Table 2. Genome assembly data for *Pollenia pediculata*.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	idPolPedi2.hap1.1	idPolPedi2.hap2.1
Assembly accession	GCA_964300415.1	GCA_964300425.1
Assembly level	chromosome	scaffold
Span (Mb)	1,156.86	1,222.48
Number of contigs	201	1,891
Number of scaffolds	49	1,187
Longest scaffold (Mb)	275.31	-
Assembly metrics* (benchmark)	Haplotype 1	Haplotype 2
Contig N50 length (≥ 1 Mb)	12.4 Mb	3.31 Mb
Scaffold N50 length (= chromosome N50)	221.33 Mb	216.64 Mb
Consensus quality (QV) (≥ 40)	65.3	63.5
k-mer completeness	71.32%	71.49%
Combined k-mer completeness ($\geq 95\%$)	98.64%	
BUSCO** (S > 90%; D < 5%)	C:99.0%[S:98.6%,D:0.4%], F:0.3%,M:0.7%,n:3,285	C:99.1%[S:97.8%,D:1.2%], F:0.3%,M:0.7%,n:3,285
Percentage of assembly mapped to chromosomes ($\geq 90\%$)	99.79%	-
Sex chromosomes (localised homologous pairs)	Not identified	-
Organelles (one complete allele)	Mitochondrial genome: 15.82 kb	-

* BUSCO scores based on the diptera_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (99.79%) was assigned to 6 chromosomal-level scaffolds. These chromosome-level

scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were

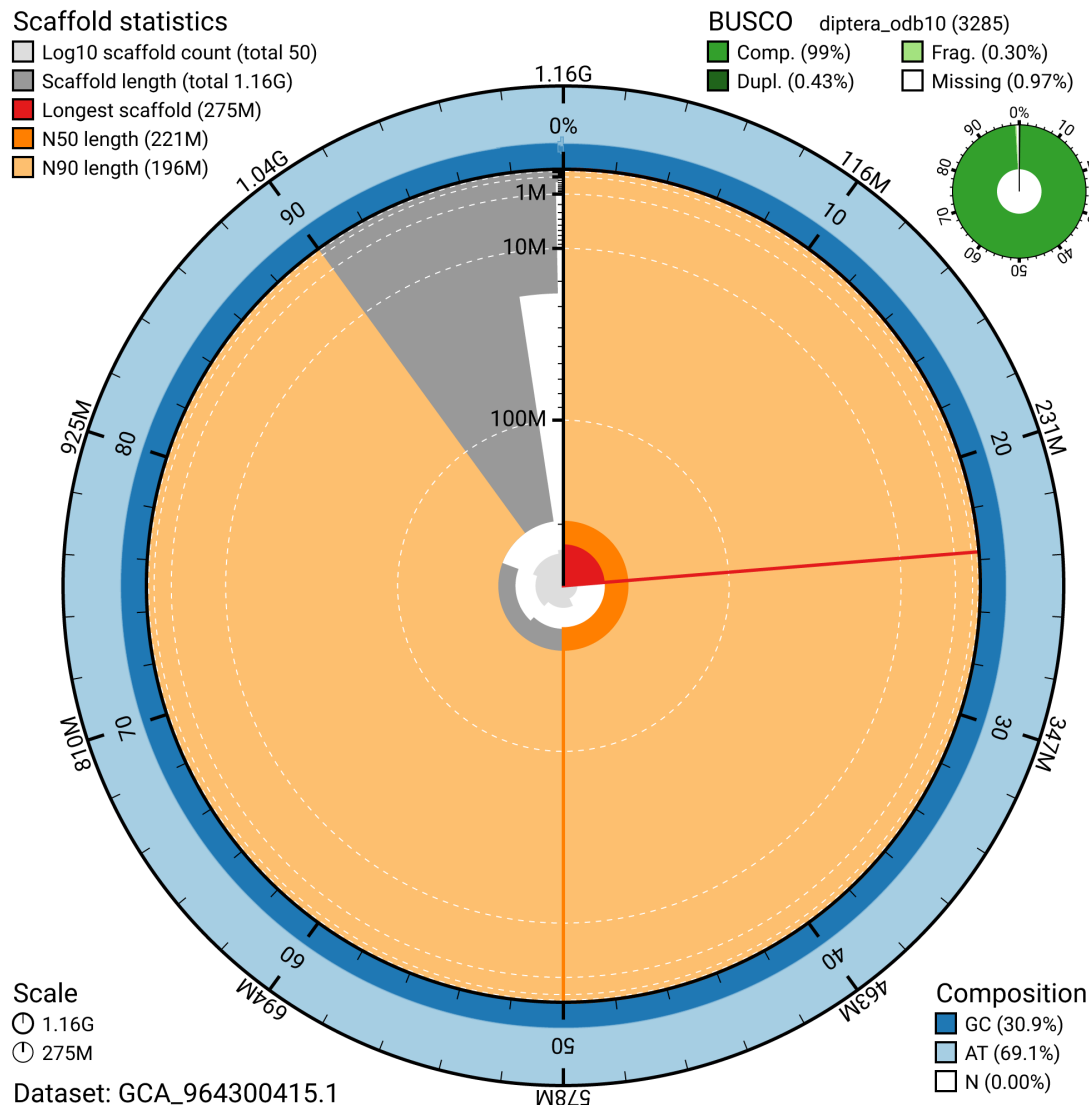


Figure 2. Genome assembly of *Pollenia pediculata*, idPolPedi2.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964300415.1/dataset/GCA_964300415.1/snail.

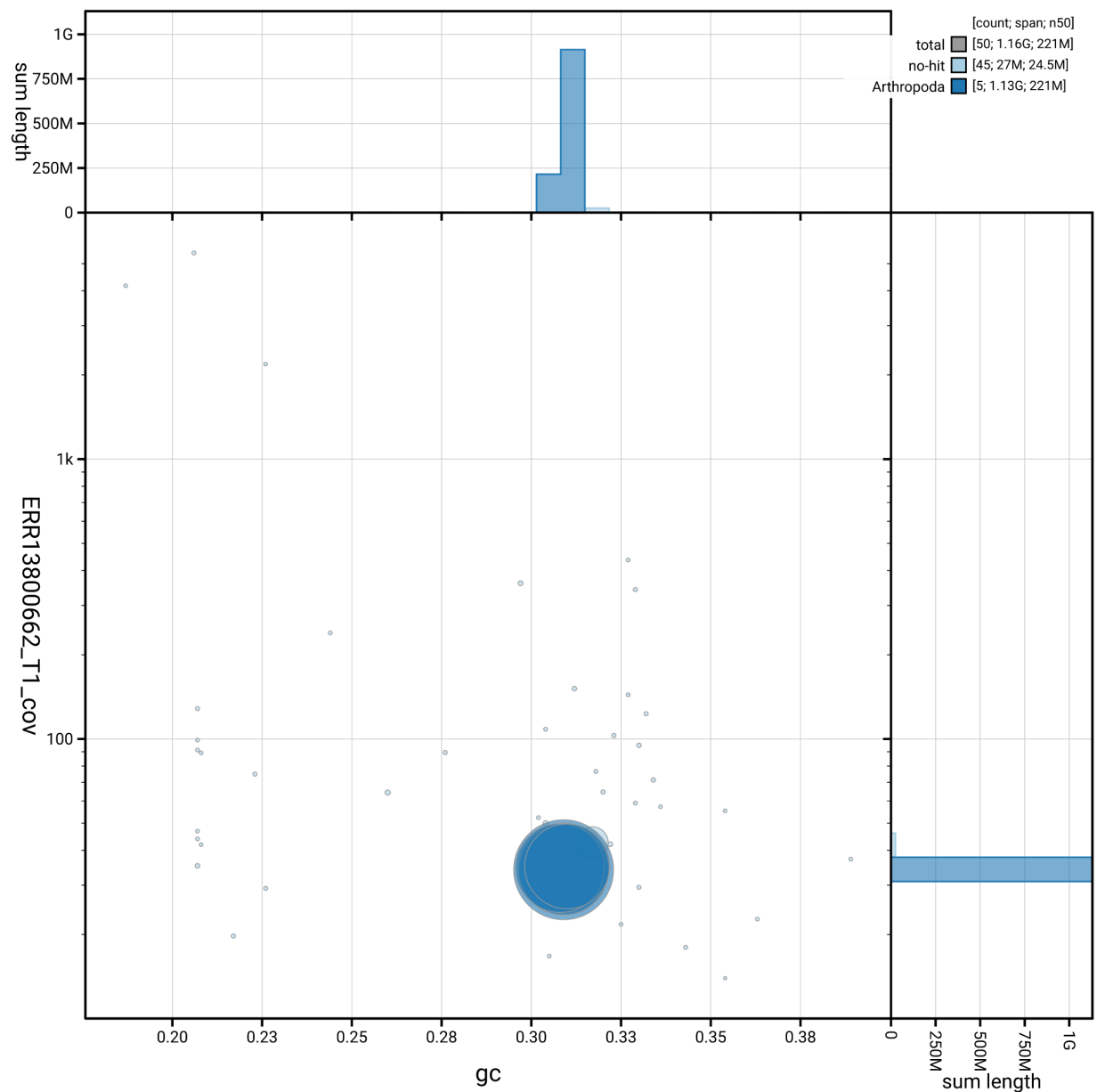


Figure 3. Genome assembly of *Pollenia pediculata*, idPolPedi2.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum 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 at https://blobtoolkit.genomehubs.org/view/GCA_964300415.1/dataset/GCA_964300415.1/blob.

calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while k -mer completeness indicates the proportion of expected k -mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

For haplotype 1, the estimated QV is 65.3, and for haplotype 2, 63.5. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.3. The k -mer

recovery for haplotype 1 is 71.32%, and for haplotype 2 71.49%, while the combined haplotypes have a k -mer recovery of 98.64%. BUSCO v.5.5.0 analysis using the diptera_odb10 reference set ($n = 3,285$) identified 99.0% of the expected gene set (single = 98.6%, duplicated = 0.4%) for haplotype 1.

Table 2 provides assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project (EBP) Report on Assembly Standards September 2024. The assembly achieves the EBP reference standard of 7.C.Q63.

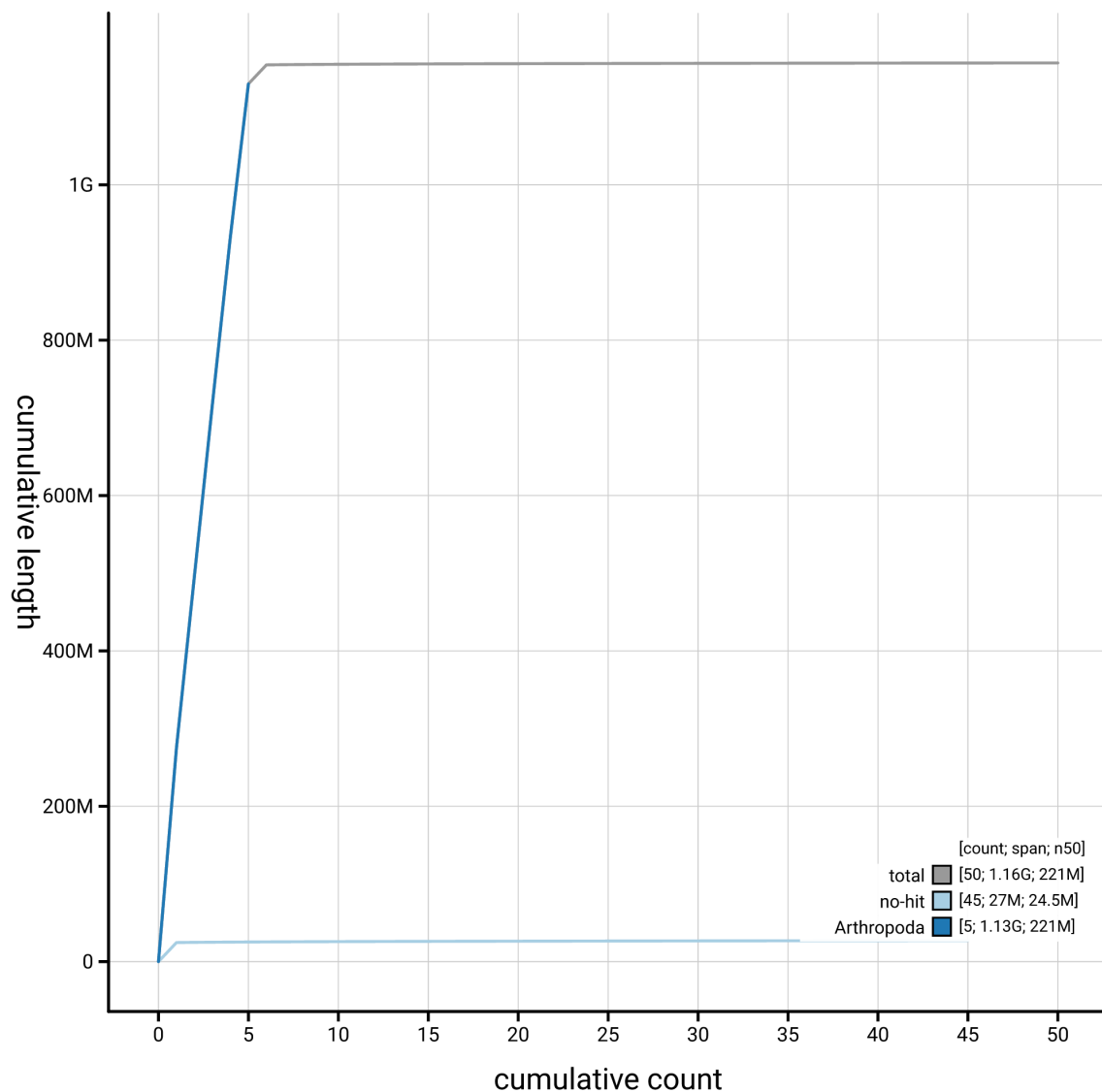


Figure 4. Genome assembly of *Pollenia pediculata*, idPolPedi2.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_964300415.1/dataset/GCA_964300415.1/cumulative.

Methods

Sample acquisition and DNA barcoding

An adult female *Pollenia pediculata* (specimen ID Ox003209, ToLID idPolPedi2) was collected from Wytham Woods, Oxfordshire, United Kingdom (latitude 51.76, longitude -1.34) on 2022-10-03 by netting. The specimen used for Hi-C sequencing (specimen ID Ox002743, ToLID idPolPedi1) was collected from the same location on 2022-06-14 by netting. The specimens were collected by Steven Falk and Liam Crowley, identified by Steven Falk and preserved on dry ice.

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 each specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (Pereira *et al.*, 2022). 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 have been deposited on protocols.io (Beasley *et al.*, 2023).

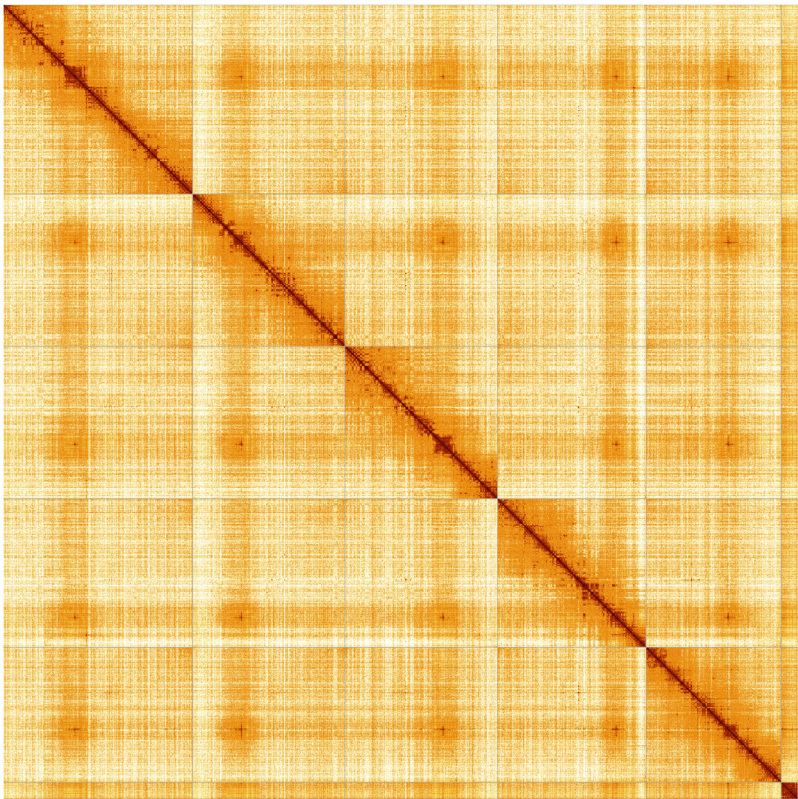


Figure 5. Genome assembly of *Pollenia pediculata*, idPolPedi2.hap1.1: Hi-C contact map of the idPolPedi2.hap1.1 assembly, produced in PretextView. Chromosomes are shown in order of size from left to right and top to bottom.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Pollenia pediculata*, idPolPedi2.

INSDC accession	Name	Length (Mb)	GC%
OZ199114.1	1	275.31	31
OZ199115.1	2	221.39	31
OZ199116.1	3	221.33	31
OZ199117.1	4	215.71	31
OZ199118.1	5	196.12	31
OZ199119.1	6	24.52	31.5
OZ199120.1	MT	0.02	22.5

Metadata collection for samples adhered to the Darwin Tree of Life project standards described by [Lawniczak et al. \(2022\)](#).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation

and purification. Detailed protocols are available on protocols.io ([Denton et al., 2023b](#)). The idPolPedi2 sample was prepared for DNA extraction by weighing and dissecting it on dry ice ([Jay et al., 2023](#)). Tissue from the whole organism was homogenised using a PowerMasher II tissue disruptor ([Denton et al., 2023a](#)). HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol ([Oatley et al., 2023](#)). The DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system ([Bates et al., 2023](#)). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA ([Strickland et al., 2023](#)). 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.

Hi-C sample preparation

Tissue from the head and thorax of the idPolPedi1 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit. In brief, 20–50 mg of frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde concentration. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the Arima-HiC v2 kit manufacturer’s

instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation. Additionally, the biotinylation percentage was estimated using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) and Qubit HS Assay Kit and Arima-HiC v2 QC beads.

Library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core.

PacBio HiFi

At a minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA) as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and gDNA 55kb BAC analysis kit.

Samples were sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries were normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases were hybridised to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration (in the range 200–300 pM), and spiked with a Revio sequencing internal control. Samples were sequenced on Revio 25M SMRT cells (Pacific Biosciences, California, USA). The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using SPRIselect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, A-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required,

determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly

Prior to assembly of the PacBio HiFi reads, a database of k -mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the k -mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), resulting in a pair of haplotype-resolved assemblies. The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQUERY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>.

Assembly quality assessment

The Merquery.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate k -mer completeness and assembly quality for both haplotypes using the k -mer databases ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The blobtoolkit pipeline is a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoAT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipeline

aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these 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), relying on the Conda package manager, the Bioconda

initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

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

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	666652151335353eef2fcd58880bcef5bc2928e1	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhyllp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
MercuryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

Life website [here](#). 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: *Pollenia pediculata* (tufted clusterfly). Accession number PRJEB81200; <https://identifiers.org/ena.embl/PRJEB81200>. The genome sequence is released openly for reuse. The *Pollenia pediculata* 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 available 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](#) and [Table 2](#).

Author information

Members of the University of Oxford and Wytham Woods Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.12157525>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.12158331>.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: <https://doi.org/10.5281/zenodo.12162482>.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: <https://doi.org/10.5281/zenodo.12165051>.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: <https://doi.org/10.5281/zenodo.12160324>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.12205391>.

Members of the Darwin Tree of Life Consortium are listed here: <https://doi.org/10.5281/zenodo.4783558>.

Acknowledgements

Thank you to Duncan Sivell for comments on the draft of the introduction.

References

- Allio R, Schomaker-Bastos A, Romiguier J, et al.: **MitoFinder: efficient automated large-scale extraction of mitochondrial data in target enrichment phylogenomics**. *Mol Ecol Resour.* 2020; **20**(4): 892–905. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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 MJ, 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](#)
- Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor[®]3 for LI PacBio**. *protocols.io.* 2023. [Publisher Full Text](#)
- Beasley J, Uhl R, Forrest LL, et al.: **DNA barcoding SOPs for the Darwin Tree of Life project**. *protocols.io.* 2023; [Accessed 25 June 2024]. [Publisher Full Text](#)
- Buchfink B, Reuter K, Drost HG: **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](#)
- Cerretti P, Stireman JO, Badano D, et al.: **Reclustering the cluster flies (Diptera: Oestroidea, Polleniidae)**. *Syst Entomol.* 2019; **44**(4): 957–972. [Publisher Full Text](#)
- Challis R, Kumar S, Sotero-Caio C, et al.: **Genomes on a Tree (GoAT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved]**. *Wellcome Open Res.* 2023; **8**: 24. [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](#)
- Chandler PJ: **An update of the 1998 checklist of Diptera of the British Isles [updated 18 April 2024]**. *Dipterists Forum*, 2024. [Reference Source](#)
- Cheng H, Concepcion GT, Feng X, et al.: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm**. *Nat Methods.* 2021; **18**(2): 170–175. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Jarvis ED, Fedrigo O, et al.: **Haplotype-resolved assembly of diploid genomes without parental data**. *Nat Biotechnol.* 2022; **40**(9): 1332–1335. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna** [version 1; peer review: 2 approved]. *Wellcome Open Res.* 2023; **8**: 123.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.

[Publisher Full Text](#)

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.

[Publisher Full Text](#)

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.

[PubMed Abstract](#) | [Publisher Full Text](#)

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Draber-Mońko A: **Calliphoridae. Plujki (Insecta: Diptera).** Fauna, Warszawa: Muzeum I Instytut Zoologii, 2004; 23.

[Reference Source](#)

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](#)

Falk S, Mitchell R, Sivell O, *et al.*: **The genome sequence of the black-bellied cluster fly, *Pollenia amentaria* (Scopoli, 1763)** [version 1; peer review: 1 approved, 1 approved with reservations]. *Wellcome Open Res.* 2023a; **8**: 466.

[Publisher Full Text](#)

Falk S, Sivell O, University of Oxford and Wytham Woods Genome Acquisition Lab, *et al.*: **The genome sequence of the narrow-cheeked clusterfly, *Pollenia angustigena* (Wainwright, 1940)** [version 1; peer review: 2 approved]. *Wellcome Open Res.* 2023b; **8**: 270.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Falk S, Sivell O, University of Oxford and Wytham Woods Genome Acquisition Lab, *et al.*: **The genome sequence of the dark-based cluster fly, *Pollenia labialis* Robineau-Desvoidy, 1863** [version 1; peer review: 1 approved]. *Wellcome Open Res.* 2024; **9**: 110.

[Publisher Full Text](#)

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfstats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Gisondi S, Rognes K, Badano D, *et al.*: **The world Polleniidae (Diptera, Oestroidea): key to genera and checklist of species.** *Zookeys.* 2020; **971**: 105–155.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Grzywacz A, Szpila K, Pape T: **Egg morphology of nine species of *Pollenia* Robineau-Desvoidy, 1830 (Diptera: Calliphoridae).** *Microsc Res Tech.* 2012; **75**(7): 955–967.

[PubMed Abstract](#) | [Publisher Full Text](#)

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps.** 2022.

[Reference Source](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): g1aa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.

[Publisher Full Text](#)

Jewiss-Gaines A, Marshall SA, Whitworth TL: **Cluster flies (Calliphoridae: Polleniinae: *Pollenia*) of North America.** *Can J Arthropod Identif.* 2012; **19**(19): 1–22.

[Publisher 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](#)

Kutty SN, Meusemann K, Bayless KM, *et al.*: **Phylogenomic analysis of**

calyptratae: resolving the phylogenetic relationships within a major radiation of diptera. *Cladistics.* 2019; **35**(6): 605–622.

[PubMed Abstract](#) | [Publisher Full Text](#)

Lawńczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project** [version 1; peer review: 2 approved with reservations]. *Wellcome Open Res.* 2022; **7**: 187.

[Publisher Full Text](#)

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Marshall SA: **Flies, the natural history and diversity of diptera.** Firefly Books, Ltd, 2012.

[Reference Source](#)

Martín-Vega D, Baz A: **Sex-biased captures of sarcosaprophagous diptera in carrion-baited traps.** *J Insect Sci.* 2013; **13**: 14.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].

[Reference Source](#)

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.

[Publisher Full Text](#)

Pereira L, Sivell O, Sivess L, *et al.*: **DTOL Taxon-specific Standard Operating Procedure for the terrestrial and freshwater arthropods working group.** 2022.

[Publisher Full Text](#)

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.

[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.*: **Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rognes K: **The taxonomy of the *Pollenia rudis* species-group in the Holarctic Region (Diptera: Calliphoridae).** *Syst Entomol.* 1987; **12**(4): 475–502.

[Publisher Full Text](#)

Rognes K: **The taxonomy and phylogenetic relationships of the *Pollenia semicineria* species-group (Diptera: Calliphoridae).** *Syst Entomol.* 1988; **13**(3): 315–345.

[Publisher Full Text](#)

Rognes K: **Blowflies (Diptera, Calliphoridae) of Fennoscandia and Denmark.** Fauna Entomologica Scandinavica 24. E.J. Brill/Scandinavian Science Press Ltd., Leiden, 1991a.

[Reference Source](#)

Rognes K: **Revision of the cluster-flies of the *Pollenia viatica* species-group (Diptera: Calliphoridae).** *Syst Entomol.* 1991b; **16**(4): 439–498.

[Publisher Full Text](#)

Rognes K: **Revision of the cluster-flies of the *Pollenia vagabunda* species-group (Diptera: Calliphoridae).** *Insect Syst Evol.* 1992a; **23**(1): 95–114.

[Publisher Full Text](#)

Rognes K: **Revision of the cluster-flies of the *Pollenia venturii* species-group, with a cladistic analysis of Palaearctic species of *Pollenia* Robineau-Desvoidy (Diptera: Calliphoridae).** *Insect Syst Evol.* 1992b; **23**(3): 233–248.

[Publisher Full Text](#)

Rognes K: **Revision of the cluster flies of the *Pollenia haeretica* species-group (Diptera, Calliphoridae).** *Zootaxa.* 2010; **2499**(1): 39–56.

[Publisher Full Text](#)

Sivell O: **Blow flies (Diptera: Calliphoridae, Polleniidae, Rhiniidae).** Handbooks for the Identification of British Insects. St Albans: Royal Entomological Society, 2021; **10**(16).

[Reference Source](#)

Sivell O, Mitchell R, Falk S, *et al.*: **The genome sequence of a cluster fly, *Pollenia griseotomentosa* (Jacentkovský, 1944)** [version 1; peer review: awaiting peer review]. *Wellcome Open Res.* 2025; **10**: 42.

[Publisher Full Text](#)

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI**. *protocols.io*. 2023.
[Publisher Full Text](#)

Szpila K: **First instar larvae of nine West-Palaeartic species of *Pollenia*; Robineau-Desvoidy, 1830 (Diptera: Calliphoridae)**. *Entomol Fenn*. 2003; **14**(4): 193–210.

[Publisher Full Text](#)

Szpila K: **Annotated list of blowflies (Diptera: Calliphoridae) recorded during studies of insect succession on large carrion in Poland**. *Dipteron, Bulletin of the Dipterological Section of the Polish Entomological Society*. 2017; **33**: 85–93.

[Publisher Full Text](#)

Szpila K, Piwczynski M, Glinkowski W, *et al.*: **First molecular phylogeny and species delimitation of West Palaeartic *Pollenia* (Diptera: Polleniidae)**. *Zool J Linn Soc*. 2023; **197**(1): 267–282.

[Publisher Full Text](#)

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2024; **9**: 339.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashennikova K, *et al.*: **MitoHiFi: a python**

pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*. 2023; **24**(1): 288.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

van Emden FI: **Diptera Cyclorrhapha, Calyptrata (I) Section (a). Tachinidae and Calliphoridae**. In: *Handbooks for the identification of British insects*. London: The Royal Entomological Society of London, 1954; **10**(4a): 1–133.

[Reference Source](#)

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems**. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.

[Publisher Full Text](#)

Yahnke W, George JA: **Rearing and immature stages of the cluster fly (*Pollenia rudis*) (Diptera: Calliphoridae) in Ontario**. *Can Entomol*. 1972; **104**(4): 567–576.

[Publisher Full Text](#)

Yan L, Pape T, Meusemann K, *et al.*: **Monophyletic blowflies revealed by phylogenomics**. *BMC Biol*. 2021; **19**(1): 230.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool**. *Bioinformatics*. 2023; **39**(1): btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 30 May 2025

<https://doi.org/10.21956/wellcomeopenres.26437.r123022>

© 2025 Ioannidis P. 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.



Panagiotis Ioannidis 

Foundation for Research & Technology - Hellas, Crete, Greece

This manuscript describes the genome sequencing and assembly of *Pollenia pediculata*, a dipteran insect. The results show that the genome is of very good quality.

Moreover, all sections are well-written but I particularly liked the introduction because the authors present a particularly challenging topic (i.e. the phylogenetic placement of *Pollenia pediculata*) that can be resolved with whole genome sequencing. Even though I know that the format for these papers is pretty well-defined, it would be nice if authors could optionally add a figure such as a species tree (pretty generic without a focus on any particular gene/gene family).

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: Insect genomics, bioinformatics

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

Reviewer Report 27 May 2025

<https://doi.org/10.21956/wellcomeopenres.26437.r123019>

© 2025 Arumugaperumal A. 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.



Arun Arumugaperumal 

Rajalakshmi Engineering College, Chennai, India

This is the first report of the whole genome sequence of *Pollenia pediculata* also known as cluster fly. The genome assembly is of size 1157 Mb. The mitogenome assembly is of size 15.82 kb. The genome is distributed in 6 chromosomes.

The data accession numbers mentioned in the data note is working fine. High-end sequencing technology was used to assemble the genome sequence. It is evident by the N50 value of 12.4 Mb. A BUSCO score of 99% indicates near completeness of the genome sequenced. The photograph showing the specimen organism is very clear.

In page 3 change *P. pendiculata* to *P. pediculata*.

The data note can be accepted for Indexing.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, 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.