DATA NOTE

The genome sequence of the Common Mummy Wasp, *Aleiodes similis* (Curtis, 1834) (Hymenoptera: Braconidae)

[version 1; peer review: 2 approved]

Gavin R. Broad[1], Natural History Museum Genome Acquisition Lab, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Natural History Museum, London, England, UK

v1 **First published:** 18 Mar 2026, **11**:169
https://doi.org/10.12688/wellcomeopenres.26127.1
Latest published: 18 Mar 2026, **11**:169
https://doi.org/10.12688/wellcomeopenres.26127.1

Abstract

We present a genome assembly from an individual female *Aleiodes similis* (Common Mummy Wasp; Arthropoda; Insecta; Hymenoptera; Braconidae). The assembly contains two haplotypes with total lengths of 280.83 megabases and 284.39 megabases. Most of haplotype 1 (99.93%) is scaffolded into 15 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 32.71 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Aleiodes similis; Common Mummy Wasp; genome sequence; chromosomal; Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 18 Mar 2026	✓ view	✓ view

1. **Lindsey C Perkin**, USDA, Agricultural Research Service, College Station, USA
2. **Marta Coronado-Zamora**, Autonomous University of Barcelona Department of Genetics and Microbiology (Ringgold ID: 196650), Bellaterra, Spain
   Vicente Chillida, Autonomous University of Barcelona Department of Genetics and Microbiology, Bellaterra, Spain

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: Broad GR: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

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.

How to cite this article: Broad GR, Natural History Museum Genome Acquisition Lab, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team *et al.* **The genome sequence of the Common Mummy Wasp, *Aleiodes similis* (Curtis, 1834) (Hymenoptera: Braconidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2026, **11**:169 https://doi.org/10.12688/wellcomeopenres.26127.1

First published: 18 Mar 2026, **11**:169 https://doi.org/10.12688/wellcomeopenres.26127.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Ichneumonoidea; Braconidae; Rogadinae; *Aleiodes*; *Aleiodes incertae sedis*; *Aleiodes similis* (Curtis, 1834) (NCBI: txid144254).

Background

Aleiodes similis is a small (~5.5 mm) braconid wasp, a parasitoid of a range of caterpillars (Lepidoptera) of the family Noctuidae. As with other *Aleiodes* species, the egg is laid in the host with the *Aleiodes* larva developing as a koinobiont endoparasitoid, i.e., the host continues to feed for some time as the wasp larva develops within. The wasp oviposits in an early instar host which is finally mummified as a later instar, allowing the *Aleiodes* to pupate within the hardened cadaver of its host. The host larvae of *A. similis* are mummified in concealment, which is an ecological distinction from the very similar *Aleiodes leptofemur*, hosts of which move to an exposed location before they are mummified (van Achterberg & Shaw, 2016).

Aleiodes similis is plurivoltine, i.e., it has than one generation per year. It attacks a range of noctuid hosts, but the summer generation mostly specialises on *Orthosia* species on trees (Shaw, 1994). The winter is spent in larvae of hosts such as *Xestia* species. This is a widespread species across Europe, on the wing from spring through to autumn, although published records are unreliable owing to many misidentifications and incorrect taxonomy. For example, GBIF incorrectly lists *A. similis* as a junior synonym of *Aleiodes gastritor* (Thunberg). The taxonomy of this and related species will be published by van Achterberg & Shaw in a paper following on from previous instalments in a revision of western Palaearctic *Aleiodes* (van Achterberg *et al.*, 2020; van Achterberg & Shaw, 2016).

The over 100 European species of *Aleiodes* can be tricky to identify, and *A. similis* is not included in any reliable identification keys; *A. similis* is usually a predominantly pale reddish (testaceous), with the mesosternum testaceous, separating it from the morphologically and ecologically similar *A. leptofemur*, which, as the name suggests, also has slenderer femora (van Achterberg & Shaw, 2016). Adults are at least partly nocturnal and frequent visitors to light traps.

Aleiodes species have been used as a model group to explore the evolution of host range in parasitoid wasps (Shaw, 1994, 2003); the availability of an increasing number of *Aleiodes* genomes will help in investigating the adaptations which underly host switches. We present a genome sequence for *Aleiodes similis*, generated using the Tree of Life pipeline from a specimen collected in Tonbridge, England, United Kingdom (Figure 1), as part of the Darwin Tree of Life project.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult female *Aleiodes similis* (specimen ID NHMUK010884704, ToLID iyAleSimi1; Figure 1), collected from Tonbridge, England, United Kingdom (latitude 51.19, longitude 0.29) on 2022-08-15. The specimen was collected and identified by Gavin Broad (Natural History Museum). Another specimen was used for RNA sequencing (specimen ID NHMUK010884535, ToLID iyAleSimi3). It was collected from Bure Marshes Nature Reserve, England, United Kingdom (latitude 52.69, longitude 1.46) on 2022-06-29. The specimen was collected by David Lees and identified by Gavin Broad.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The iyAleSimi1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the whole organism was homogenised by powermashing using a PowerMasher II tissue disruptor.

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA

Figure 1. Photograph of the *Aleiodes similis* (iyAleSimi1) specimen used for genome sequencing.

was purified by manual SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 5.31 ng/μL and a yield of 238.95 ng, with a fragment size of 18.3 kb. The 260/280 spectrophotometric ratio was 2.31.

RNA was extracted from whole organism tissue of iyAleSimi3 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the iyAleSimi1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo (dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X to generate 150-bp paired-end reads.

Genome assembly

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

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with 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 MerquryFK (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. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 161 breaks, 230 joins, and removal of 7 haplotypic duplications. This reduced the scaffold count by 69.4%, increased the scaffold N50 by 3.3%, and reduced the total assembly length by 1.6%. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The MerquryFK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for 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 genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Aleiodes similis* specimen generated 19.68 Gb (gigabases) from 1.66 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 254.49 Mb, with a heterozygosity of 1.40% and repeat content of 27.65% (Figure 2). These estimates guided expectations for the assembly.

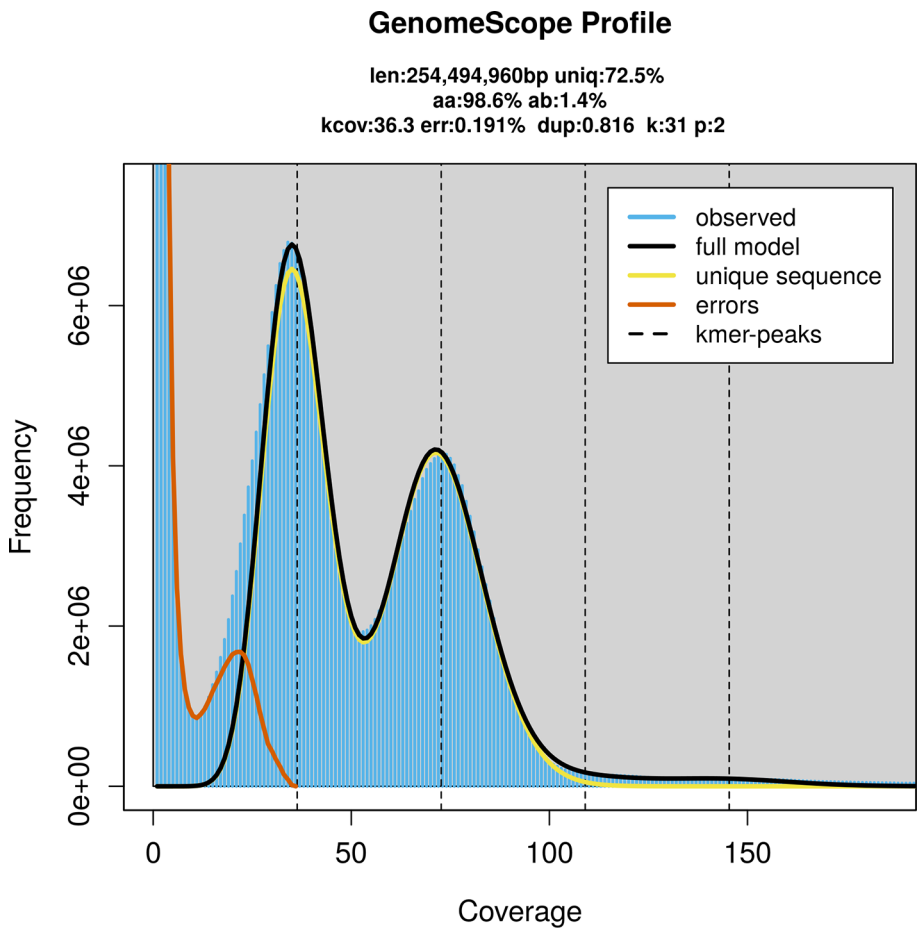

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. **Specimen and sequencing data for BioProject PRJEB73896.**

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	iyAleSimi1	iyAleSimi1	iyAleSimi3
Specimen ID	NHMUK010884704	NHMUK010884704	NHMUK010884535
BioSample (source individual)	SAMEA112964392	SAMEA112964392	SAMEA114805899
BioSample (tissue)	SAMEA112975558	SAMEA112975558	SAMEA114806068
Tissue	whole organism	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq 6000	Illumina NovaSeq X
Run accessions	ERR12760793	ERR12765166	ERR13493924
Read count total	1.66 million	646.54 million	90.41 million
Base count total	19.68 Gb	97.63 Gb	13.65 Gb

Table 2. **Genome assembly statistics.**

Assembly name	iyAleSimi1.hap1.1	iyAleSimi1.hap2.1
Assembly accession	GCA_965249695.1	GCA_965249795.1
Assembly level	chromosome	scaffold
Span (Mb)	280.83	284.39
Number of chromosomes	15	scaffold-level
Number of contigs	359	346
Contig N50	1.49 Mb	1.75 Mb
Number of scaffolds	33	50
Scaffold N50	17.7 Mb	18.0 Mb
Longest scaffold length (Mb)	26.9	-
Organelles	Mitochondrion: 32.71 kb	-

Based on the estimated genome size, the sequencing data provided approximately 73× coverage. Hi-C sequencing produced 97.63 Gb from 646.54 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

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 final assembly has a total length of 280.83 Mb in 33 scaffolds, with 326 gaps, and a scaffold N50 of 17.7 Mb (Table 2).

Most of the assembly sequence (99.93%) was assigned to 15 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). The order and orientation of scaffolds in region ~9.87–10.44 Mb on Chromosome 5, and ~ 8.03–9.52 Mb on Chromosome 15 are uncertain.

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.

For haplotype 1, the estimated QV is 64.6, and for haplotype 2, 64.6. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.6. The *k*-mer completeness is 74.80% for haplotype 1, 75.20% for haplotype 2, and 99.44% for the combined haplotypes (Figure 4).

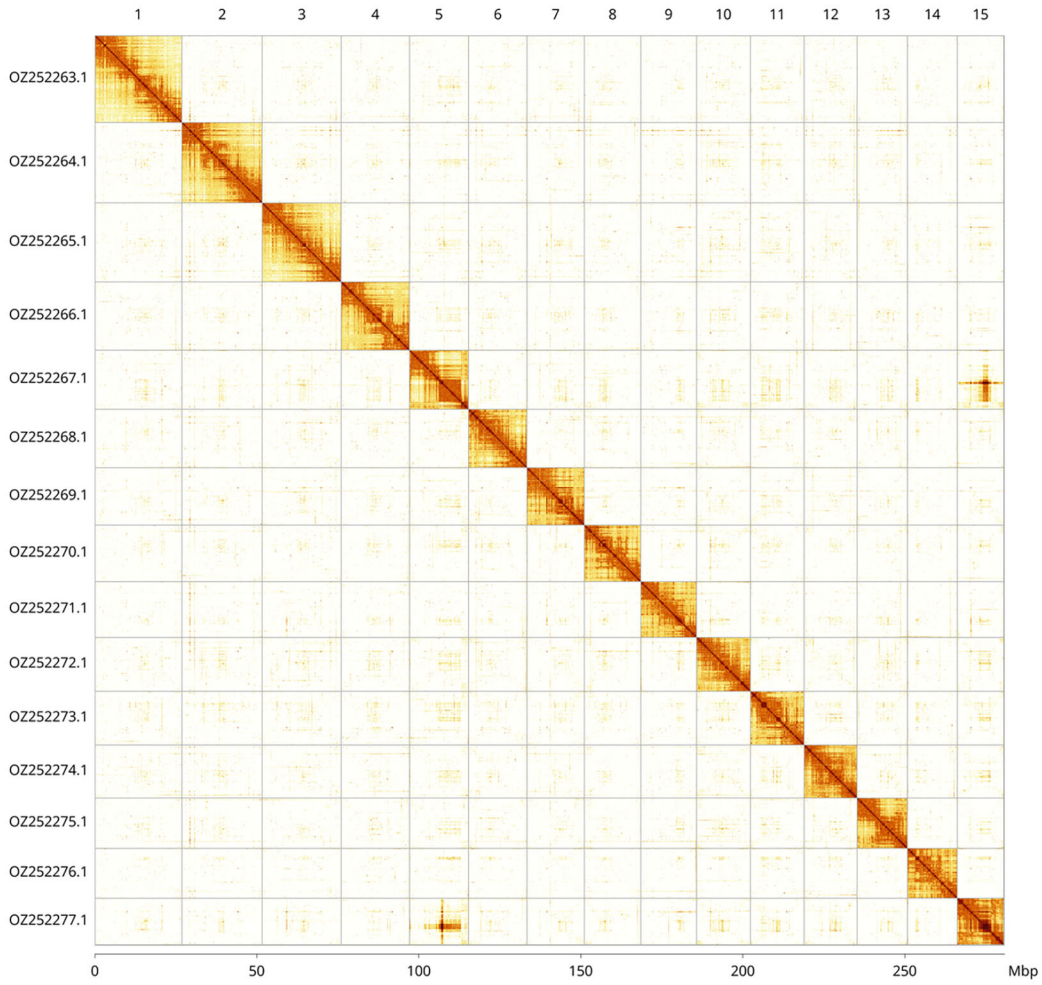

Figure 3. Hi-C contact map of the *Aleiodes similis* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Aleiodes similis* iyAleSimi1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ252263.1	1	26.90	34
OZ252264.1	2	24.77	34.50
OZ252265.1	3	24.39	34
OZ252266.1	4	21.05	34.50
OZ252267.1	5	18.25	35
OZ252268.1	6	18.01	34
OZ252269.1	7	17.70	34
OZ252270.1	8	17.47	34
OZ252271.1	9	17.18	34

Table 3. *Continued*

INSDC accession	Molecule	Length (Mb)	GC%
OZ252272.1	10	16.62	34
OZ252273.1	11	16.60	34.50
OZ252274.1	12	16.35	34.50
OZ252275.1	13	15.55	34
OZ252276.1	14	15.34	34
OZ252277.1	15	14.44	35

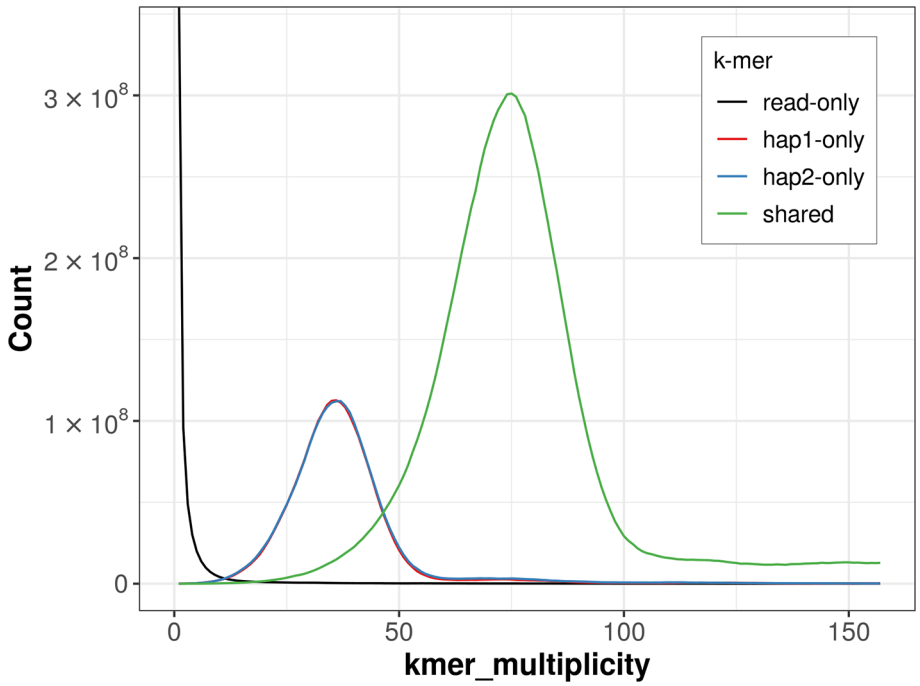

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

BUSCO analysis using the hymenoptera_odb10 reference set (n = 5 991) identified 96.4% of the expected gene set (single = 96.0%, duplicated = 0.4%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q64**, meeting the recommended reference standard.

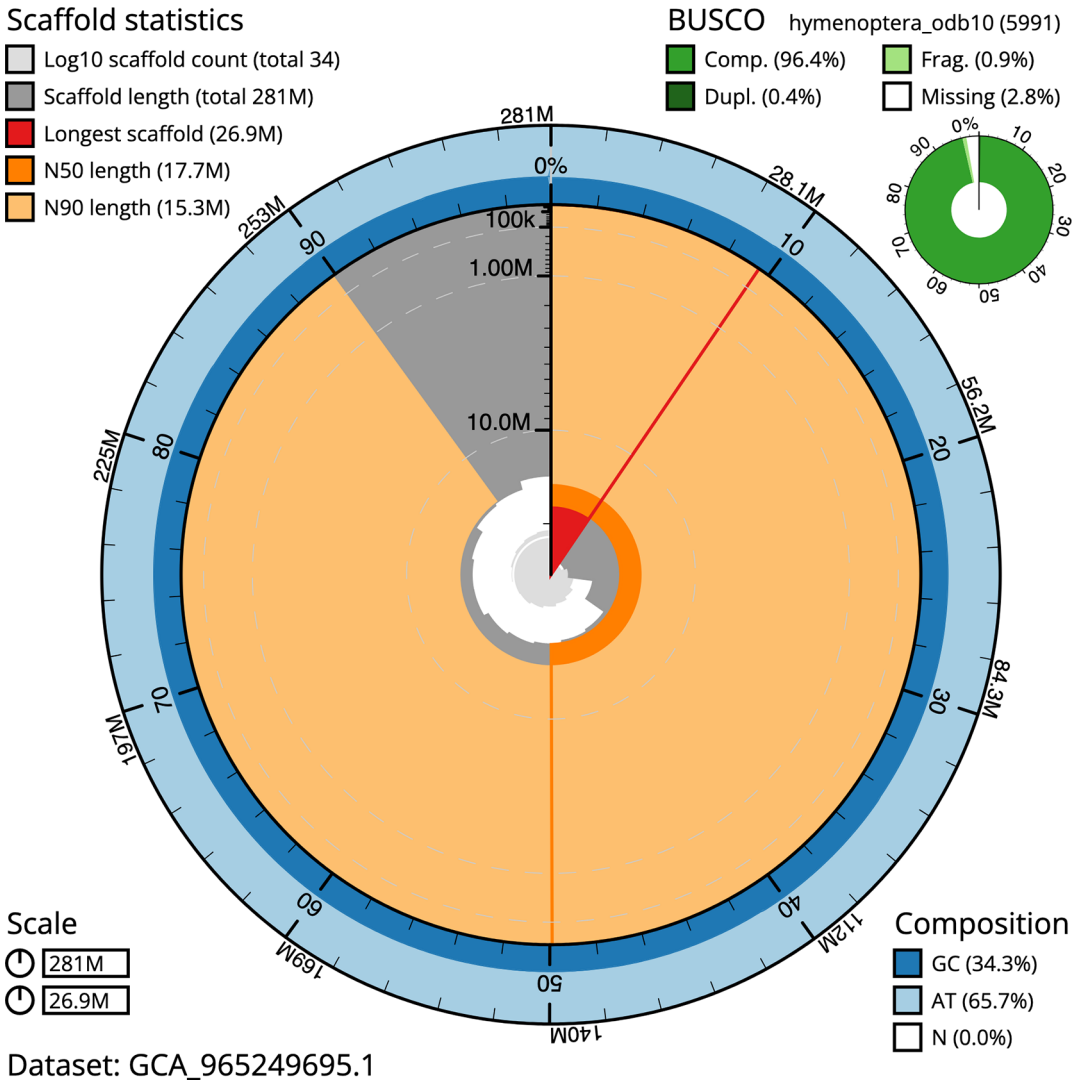

Figure 5. Assembly metrics for iyAleSimi1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

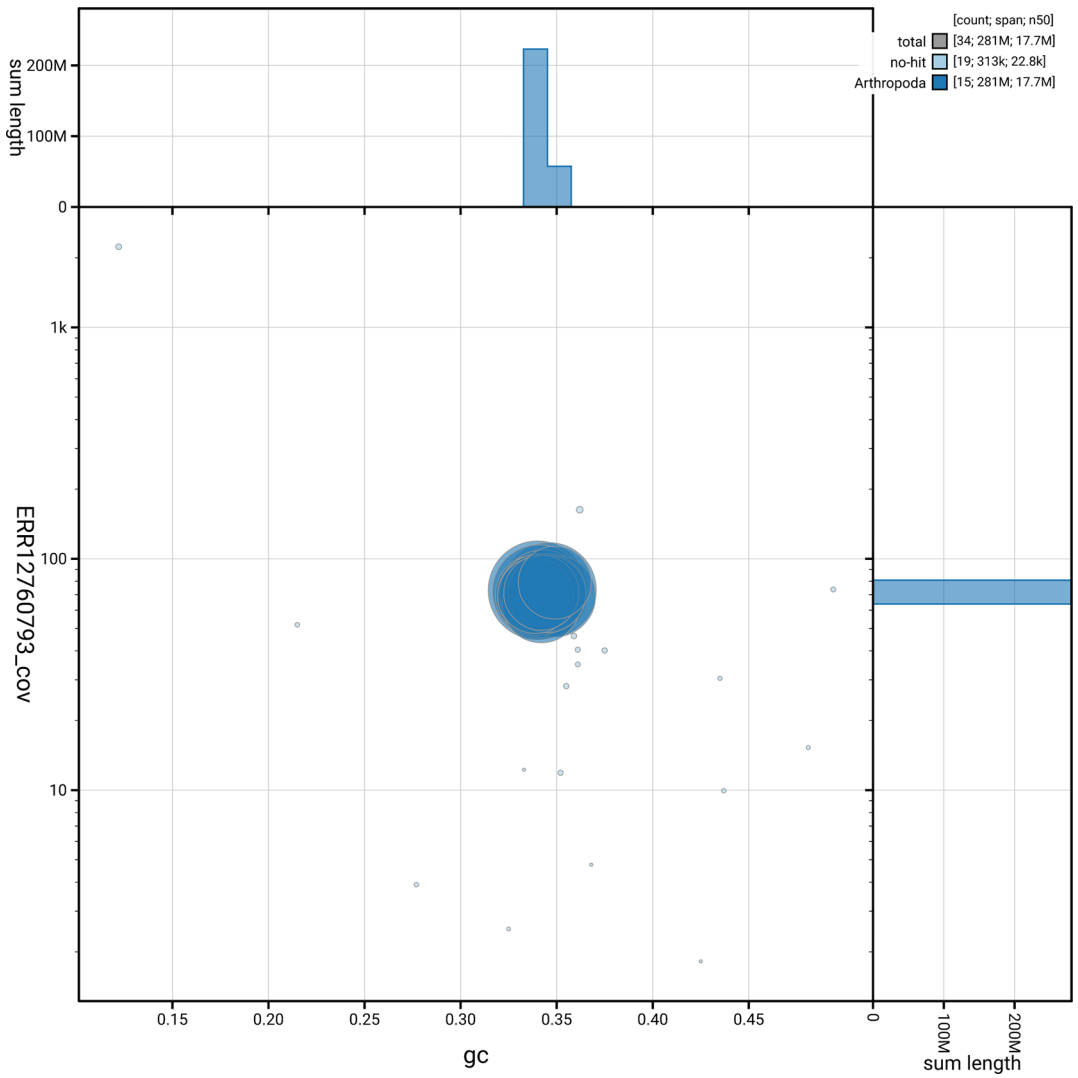

Figure 6. BlobToolKit GC-coverage plot for iyAleSimi1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *Aleiodes similis* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q64	6.C.Q40
Contig N50 length	1.49 Mb	≥ 1 Mb
Scaffold N50 length	17.70 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 64.6; haplotype 2: 64.6; combined: 64.6	≥ 40
k-mer completeness	Haplotype 1: 74.80%; Haplotype 2: 75.20%; combined: 99.44%	≥ 95%

Table 4. *Continued*

Measure	Value	Benchmark
BUSCO	C:98.7% [S:98.1%, D:0.7%], F:0.3%, M:0.9%, n:2 124	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.93%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologues.

Author information

Contributors are listed at the following links:

- Members of the Natural History Museum Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Wellcome Sanger Institute – Legal and governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: Aleiodes similis. Accession number PRJEB73896. The genome sequence is released openly for reuse. The *Aleiodes similis* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Tables 1 and 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Table 5. Software versions and sources.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	https://github.com/sanger-tol/blobtoolkit
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

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.
Publisher 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 M-J, Orchard S, *et al.*: **UniProt: The Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND.** *Nat. Methods.* 2021; **18**(4): 366–368.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 Genes Genomes Genet.* 2020; **10**(4): 1361–1374.
Publisher Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved de novo assembly using phased assembly graphs with Hifiasm.** *Nat. Methods.* 2021; **18**(2): 170–175.
Publisher 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.
Publisher Full Text

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial Arthropod fauna.** *Wellcome Open Res.* 2023; **8**: 123.
Publisher Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2).
PubMed Abstract | Publisher Full Text | Free Full Text

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

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat. Biotechnol.*

2020; **38**(3): 276–278.
PubMed Abstract | **Publisher Full Text**

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

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: Web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
Publisher Full Text

Kurtzer GM, Sochat V, Bauer MW: **Singularity: Scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
Publisher Full Text

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

Merkel D: **Docker: Lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239).
Publisher Full Text

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432.
Publisher 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.
Publisher 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.
Publisher Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools.** *Database.* 2020; **2020**: baaa062.
Publisher Full Text

Shaw MR: **Parasitoid host ranges.**Hawkins BA, Sheehan W, editors. *Parasitoid Community Ecology.* Oxford; New York; Tokyo: Oxford University Press; 1994; 111–44.
Publisher Full Text

Shaw MR: **Host ranges of Aleiodes species (Hymenoptera: Braconidae), and an evolutionary hypothesis.**Melika G, Thuróczy C, editors. *Parasitic Wasps: Evolution, Systematics, Biodiversity and Biological Control [2002].* Budapest, Hungary: Agroinform; 2003; 321–27.

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project.** *Wellcome Open Res.* 2024; **9**: 339.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: A Python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
Publisher Full Text

van Achterberg C, Shaw MR: **Revision of the western Palaearctic species of *Aleiodes* Wesmael (Hymenoptera, Braconidae, Rogadinae). Part 1: Introduction, key to species groups, outlying distinctive species, and revisionary notes on some further species.** *ZooKeys.* 2016; **639**: 1–164.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

van Achterberg C, Shaw MR, Quicke DLJ: **Revision of the western Palaearctic species of *Aleiodes* Wesmael (Hymenoptera, Braconidae, Rogadinae). Part 2: Revision of the *A. apicalis* group.** *ZooKeys.* 2020; **919**: 1–259.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE; 2019; 314–24.
Publisher Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1).
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 03 April 2026

https://doi.org/10.21956/wellcomeopenres.28774.r150995

✔ **Marta Coronado-Zamora**
Autonomous University of Barcelona Department of Genetics and Microbiology (Ringgold ID: 196650), Bellaterra, Catalonia, Spain

Vicente Chillida
Autonomous University of Barcelona Department of Genetics and Microbiology, Bellaterra, Catalonia, Spain

This data note describes the first high quality chromosomal assembly of the Common Mummy Wasp, *Aleiodes similis*. The genome data are available at NCBI BioProject, and data is accessible as checked in March 2026. The genome assembly is highly contiguous, with overall very good assembly metrics. The mitochondrial genome is also included. The genome currently lacks gene annotations, although RNA-seq was also produced.

Comments:

- Regarding the RNA extraction, although the text states the use of Nanodrop, Qubit, and Agilent RNA 6000 Pico kit, neither the RNA concentration nor any RNA integrity metrics are reported. Nevertheless, for DNA extraction, quality metrics are provided.
- Regarding haplotype 2, why it was assembled to the scaffold level while the haplotype 1 was able to be assembled to the chromosome level? Also, haplotype 2 has better statistics for N50 contig and scaffold values.
- The other specimen collected for RNA-sequencing was also a female?
- The sentence “For haplotype 1, the estimated QV is 64.6, and for haplotype 2, 64.6. When the two haplotypes are combined, the assembly achieves an estimated QV of 64.6.” is highly repetitive, it could be rephrased.
- The EBP metric is not explained, it would be useful to explain how to interpret this value.
- There are some software cited in Table 5 but not mentioned in the main text: Cooler, fasta_windows and GoaT CLI.

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: genomics, bioinformatics, population genetics, adaptation

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

Reviewer Report 28 March 2026

https://doi.org/10.21956/wellcomeopenres.28774.r150991

Lindsey C Perkin
USDA, Agricultural Research Service, College Station, Texas, USA

Summary: The paper describes the genome assembly of the common mummy wasp, *Aleiodes similis*. The author chose this insect because it is a parasitoid of a range of Noctuidae caterpillars and hard to distinguish morphologically with *A. gastritor*. The author clearly describes and link the protocols and materials and methods for the assembly to repeated.
The datasets are useable and put into depositories. I only have a couple edits and suggestions.
1) In the Background section.
The first sentence in paragraph two should read "..i.e., it has MORE than one generation per year."
2) RNA is extracted and sequenced but doesn't seem to be used. Could you do gene annotations?

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?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: gene expression, genomics, molecular biology

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.