DATA NOTE

The genome sequence of a braconid wasp, *Aleiodes alternator* (Nees, 1834) [version 1; peer review: awaiting peer review]

Kevin Radley[1], Mark R. Shaw[2], Gavin R. Broad[3],
Natural History Museum 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

[1]Independent researcher, Wheatfen Reserve, Surlingham, England, UK
[2]National Museums Scotland, Edinburgh, Scotland, UK
[3]Natural History Museum, London, England, UK

v1 **First published:** 13 Aug 2024, **9**:470
https://doi.org/10.12688/wellcomeopenres.22768.1
Latest published: 13 Aug 2024, **9**:470
https://doi.org/10.12688/wellcomeopenres.22768.1

Abstract

We present a genome assembly from an individual female *Aleiodes alternator* (braconid wasp; Arthropoda; Insecta; Hymenoptera; Braconidae). The genome sequence spans 234.90 megabases. Most of the assembly is scaffolded into 19 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 28.92 kilobases in length.

Keywords

Aleiodes alternator, braconid wasp, genome sequence, chromosomal, Hymenoptera

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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: Radley K: Resources; **Shaw MR**: Investigation, Writing – Original Draft Preparation; **Broad GR**: 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 [206194, https://doi.org/10.35802/206194] 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: Radley K, Shaw MR, Broad GR *et al.* **The genome sequence of a braconid wasp, *Aleiodes alternator* (Nees, 1834) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:470
https://doi.org/10.12688/wellcomeopenres.22768.1

First published: 13 Aug 2024, **9**:470 https://doi.org/10.12688/wellcomeopenres.22768.1

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 alternator* (Nees, 1834) (NCBI:txid144253).

Background

Aleiodes alternator is a moderate sized (about 6mm) braconid wasp in the subfamily Rogadinae. It is very variable in colour, but usually black with central metasomal tergites red, orange orbits and largely orange legs with the hind femur apically black. Sometimes the mesosoma is extensively reddish. It is a member of the *Aleiodes bicolor*-group (c.f. van Achterberg & Shaw, 2016) for which a key to species is in preparation (van Achterberg *et al.*, in prep), and within that species-group it is best provisionally recognised by its relatively large size and robust build.

Like other Rogadinae, it is a parasitoid of Lepidoptera larvae, which are mummified by the wasp larva to form a hardened structure in which the parasitoid pupates (Zaldívar-Riverón *et al.*, 2008). Most largeish low-feeding and densely setose larvae in the families Lasiocampidae, and Erebidae (Arctiinae: Arctiini and Lymantriinae) serve as hosts for this plurivoltine parasitoid, and their superficial similarity yet unrelatedness has been interpreted as evidence of historical recruitment to an expanding host repertoire (Shaw, 2002). Overwintering is as an early instar within the overwintering host larva. Just before mummification, the stricken host larva climbs to an exposed position, where increased insolation of the mummy will accelerate adult development. This is a common species over much of Europe, occurring especially in open habitats such as fens, moorland and dune slacks, where the mummies can often be found on prominent vegetation.

There are signs that speciation events may be starting withing this species, with both univoltine races and specialised populations with reduced host capabilities being noted in certain habitats (Shaw, unpublished). This genome is a first step towards exploring such processes.

Genome sequence report

The genome of an adult female *Aleiodes alternator* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 24.20 Gb (gigabases) from 2.37 million reads, providing approximately 99-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 123.09 Gbp from 815.18 million reads, yielding an approximate coverage of 524-fold. Specimen and sequencing information is summarised in Table 1.

Figure 1. Photograph of the *Aleiodes alternator* (iyAleAlte1) specimen used for genome sequencing.

Table 1. Specimen and sequencing data for *Aleiodes alternator*.

Project information			
Study title	Aleiodes alternator		
Umbrella BioProject	PRJEB70745		
Species	*Aleiodes alternator*		
BioSample	SAMEA112964490		
NCBI taxonomy ID	144253		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	iyAleAlte1	SAMEA112975691	Whole organism
Hi-C sequencing	iyAleAlte1	SAMEA112975691	Whole organism
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR12342493	8.15e+08	123.09
PacBio Revio	ERR12340377	2.37e+06	24.2

Manual assembly curation corrected seven missing joins or mis-joins and three haplotypic duplications, reducing the scaffold number by 22.22%. The final assembly has a total length of 234.90 Mb in 20 sequence scaffolds with a scaffold N50 of 12.6 Mb (Table 2). The total count of gaps in the scaffolds is 71. The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.98%) of the assembly sequence was assigned to 19 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 3). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 64.0 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.4.3 completeness of 96.1% (single = 95.7%, duplicated = 0.4%), using the hymenoptera_odb10 reference set (n = 5,991).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/144253.

Methods

Sample acquisition and nucleic acid extraction

An adult female *Aleiodes alternator* (specimen ID NHMUK010884702, ToLID iyAleAlte1) was collected from Wheatfen Nature Reserve, Surlingham, England, UK (latitude 52.59, longitude –1.43) on 2022-05-26 by rearing from a larva of *Euthrix potatoria* (L.). The specimen was collected by Kevin Radley (Wheatfen Reserve Volunteer Guide) and identified by Mark Shaw (National Museums of Scotland) and preserved by dry freezing at –80 °C in August 2022, after she had parasitized a succession of larvae in the laboratory.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures:

Table 2. Genome assembly data for *Aleiodes alternator*, iyAleAlte1.1.

Genome assembly		
Assembly name	iyAleAlte1.1	
Assembly accession	GCA_963921955.1	
Accession of alternate haplotype	*GCA_963921975.1*	
Span (Mb)	234.90	
Number of contigs	92	
Contig N50 length (Mb)	4.6	
Number of scaffolds	20	
Scaffold N50 length (Mb)	12.6	
Longest scaffold (Mb)	20.06	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	64.0	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:96.1%[S:95.7%,D:0.4%], F:0.9%,M:3.0%,n:5,991	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.98%	*≥ 95%*
Sex chromosomes	None	*localised homologous pairs*
Organelles	Mitochondrial genome: 28.92 kb	*complete single alleles*

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from Rhie *et al.* (2021).

** BUSCO scores based on the hymenoptera_odb10 BUSCO set using version 5.4.3. 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/Aleiodes_alternator/dataset/GCA_963921955.1/busco.

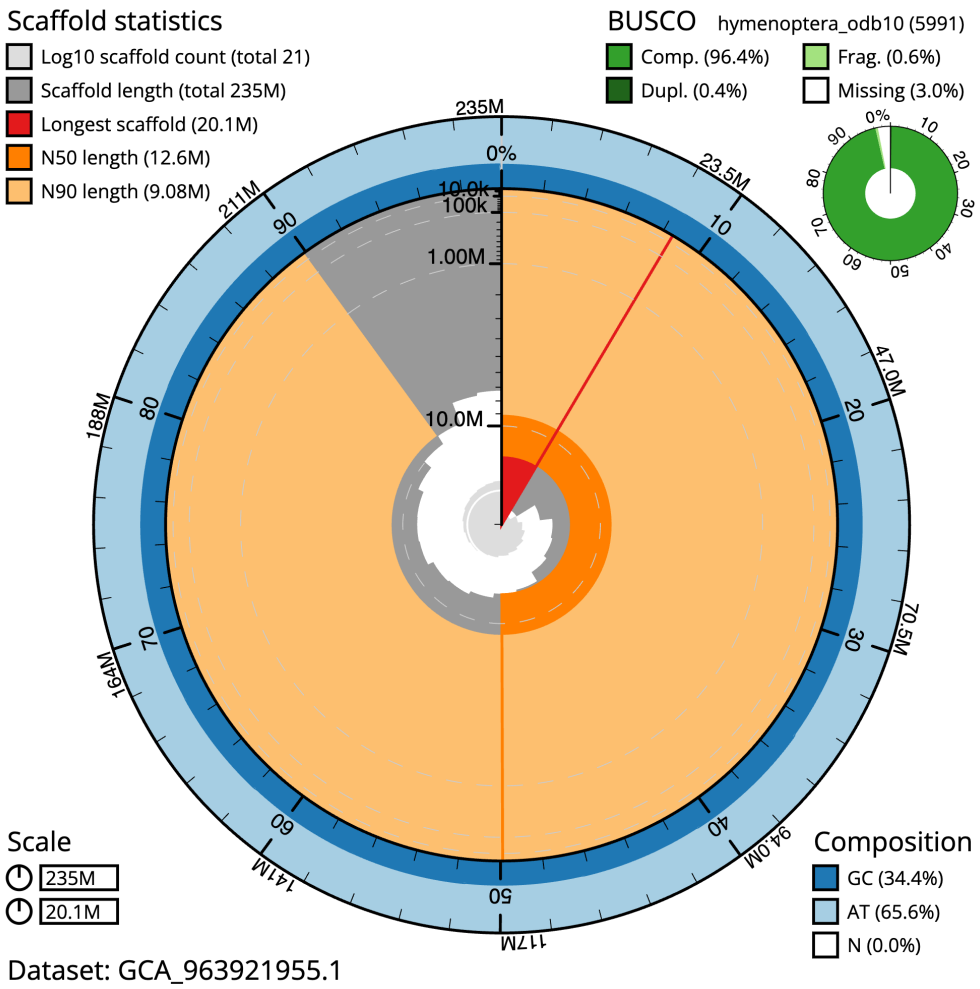

Figure 2. Genome assembly of *Aleiodes alternator*, iyAleAlte1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 234,912,219 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (20,060,207 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (12,584,262 and 9,075,812 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the hymenoptera_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Aleiodes_alternator/dataset/GCA_963921955.1/snail.

sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the iyAleAlte1 sample was weighed and dissected 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 at 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 (Strickland *et al.*, 2023): in brief, the method employs AMPure PB beads to eliminate shorter fragments and concentrate the DNA. 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.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific

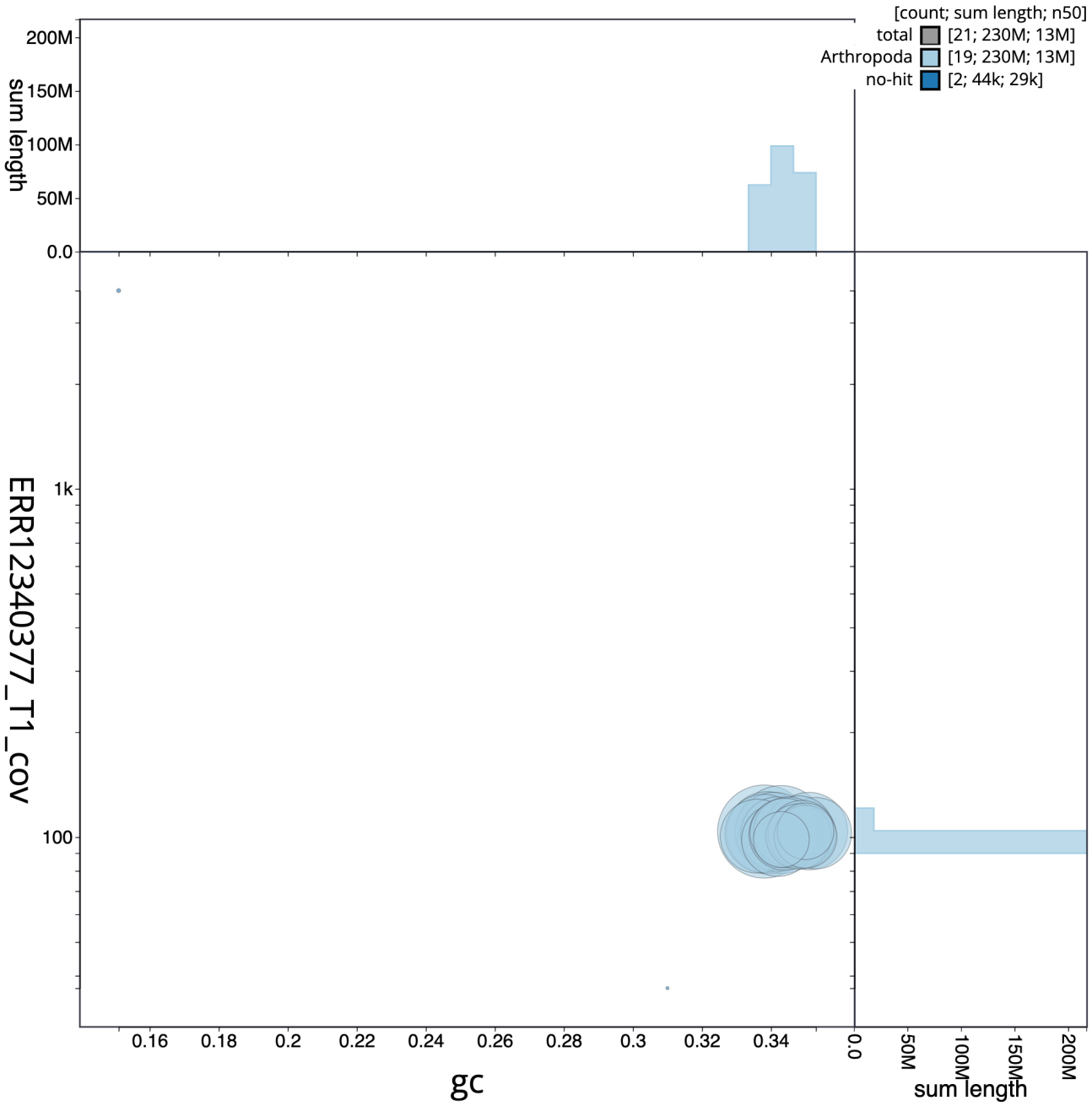

Figure 3. Genome assembly of *Aleiodes alternator*, iyAleAlte1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Aleiodes_alternator/dataset/GCA_963921955.1/blob.

Operations core at the WSI on a Pacific Biosciences Revio instrument. Hi-C data were also generated from remaining whole organism tissue of iyAleAlte1 using the Arima-HiC v2 kit. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly

Original assembly of HiFi reads is performed using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed with purge_dups (Guan *et al.*, 2020). Hi-C reads are further mapped with

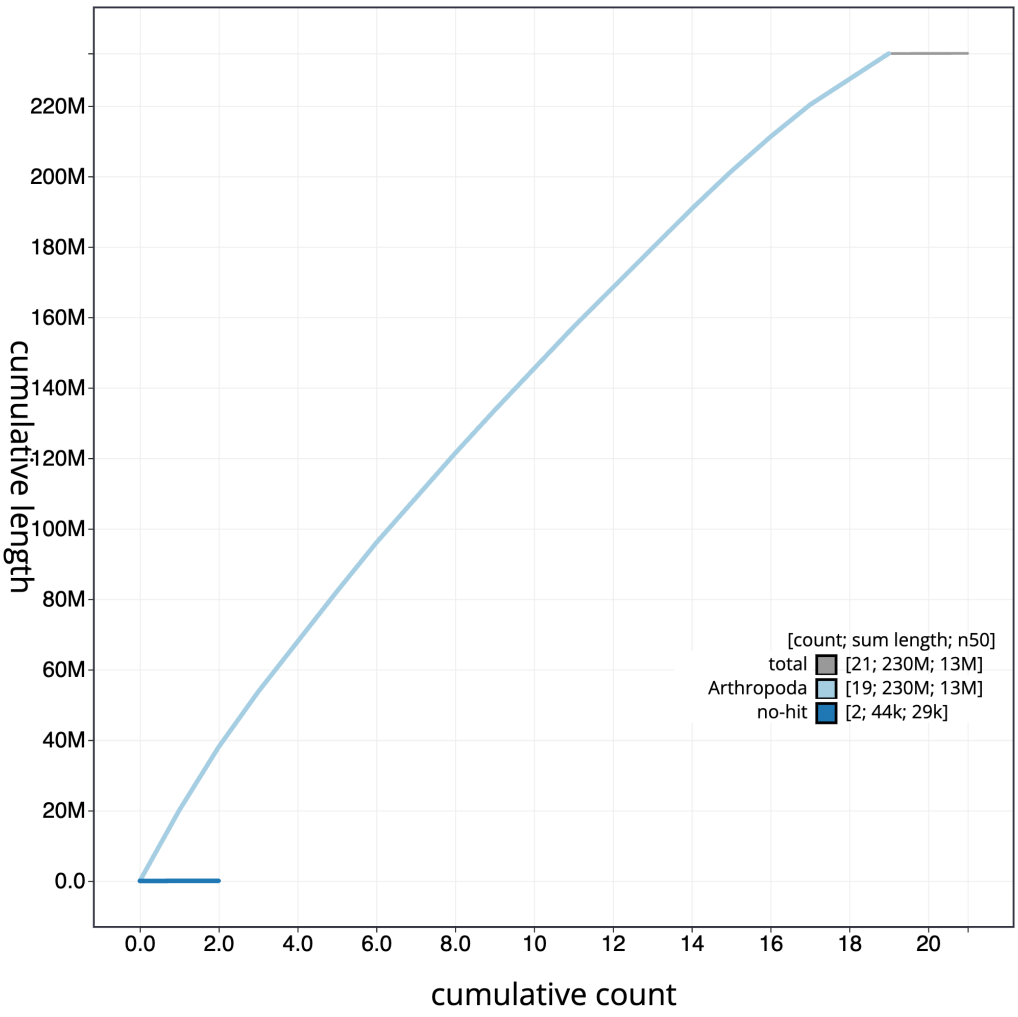

Figure 4. Genome assembly of *Aleiodes alternator* iyAleAlte1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Aleiodes_alternator/dataset/GCA_963921955.1/cumulative.

bwamem2 (Vasimuddin *et al.*, 2019) to the primary contigs, which are further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option. Scaffolded assemblies are evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) 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 (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The entire process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

Evaluation of the final assembly

The final assembly was post-processed and evaluated with the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a), "sanger-tol/genomenote" (Surana *et al.*, 2023b), and "sanger-tol/blobtoolkit"

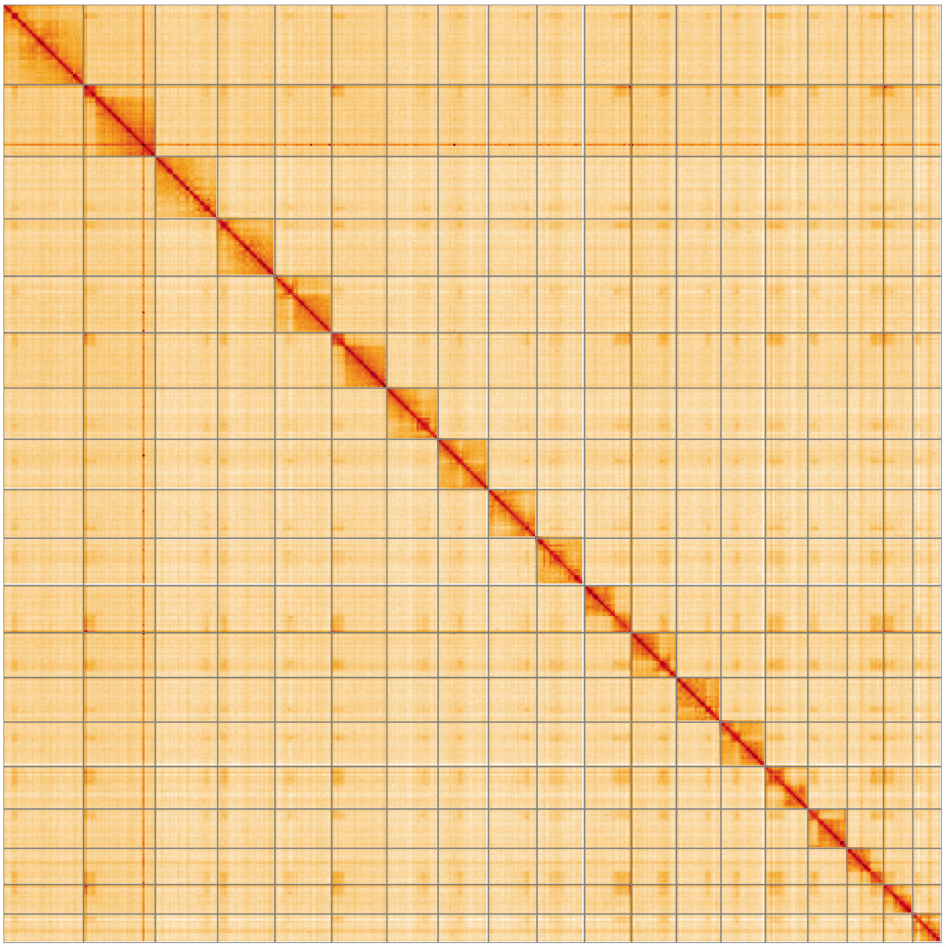

Figure 5. Genome assembly of *Aleiodes alternator* iyAleAlte1.1: Hi-C contact map of the iyAleAlte1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Dlw_kHh0SHS9pgQWfhPegQ.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Aleiodes alternator*, iyAleAlte1.

INSDC accession	Name	Length (Mb)	GC%
OY998097.1	1	20.06	34.0
OY998098.1	2	17.96	34.5
OY998099.1	3	15.63	34.0
OY998100.1	4	14.35	34.0
OY998101.1	5	14.18	34.0
OY998102.1	6	13.84	35.0
OY998103.1	7	12.81	34.0
OY998104.1	8	12.58	33.5
OY998105.1	9	12.18	34.0
OY998106.1	10	11.87	34.5
OY998107.1	11	11.76	35.5
OY998108.1	12	11.24	35.0
OY998109.1	13	11.17	34.5
OY998110.1	14	11.14	34.5
OY998111.1	15	10.62	35.0
OY998112.1	16	9.84	35.0
OY998113.1	17	9.08	35.0
OY998114.1	18	7.38	35.0
OY998115.1	19	7.18	34.5
OY998116.1	MT	0.03	15.0

(Muffato *et al.*, 2024). The pipeline sanger-tol/readmapping aligns the Hi-C reads with bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The sanger-tol/genomenote pipeline transforms the Hi-C alignments into a contact map with BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020), which is then visualised with HiGlass (Kerpedjiev *et al.*, 2018). It also provides statistics about the assembly with the NCBI datasets (Sayers *et al.*, 2024) report, computes *k*-mer completeness and QV consensus quality values with FastK and MERQURY.FK, and a completeness assessment with BUSCO (Manni *et al.*, 2021).

The sanger-tol/blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads with 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 lineage, the pipeline aligns the BUSCO genes to the Uniprot Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND (Buchfink *et al.*, 2021) blastp. The genome is also split into chunks according to the density of the BUSCO genes from the closest taxonomically lineage, and each chunk is aligned to the Uniprot Reference Proteomes database with DIAMOND blastx. Genome sequences that have no hit are then chunked with seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). All those outputs are combined with the blobtools suite into a blobdir for visualisation.

The genome assembly and evaluation pipelines were developed using the nf-core tooling (Ewels *et al.*, 2020), use MultiQC (Ewels *et al.*, 2016), and make extensive use of the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), and 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 Life website here. By agreeing with and signing up

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 5.5.0	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	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	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	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84a a44357826c0b6753eb28de	https://github.com/higlass/higlass
Merqury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets

Software tool	Version	Source
Nextflow	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/genomenote	1.1.1	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.2.1	https://github.com/sanger-tol/readmapping
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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 alternator*. Accession number PRJEB70745; https://identifiers.org/ena.embl/PRJEB70745 (Wellcome Sanger Institute, 2024). The genome sequence is released openly for reuse. The *Aleiodes alternator* 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 Natural History Museum Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.12159242.

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.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
PubMed Abstract | 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

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

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

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

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *Gigascience.* 2021; **10**(2): giab008.
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**. *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

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

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.
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

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022; [Accessed 19 October 2022].
Reference Source

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

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

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

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

Muffato M, Butt Z, Challis R, *et al.*: **sanger-tol/blobtoolkit: v0.3.0 – Poliwag**. 2024.
Publisher Full Text

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

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

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
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.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | Publisher Full Text | Free Full Text

Sayers EW, Cavanaugh M, Clark K, *et al.*: **GenBank 2024 Update.** *Nucleic Acids Res.* 2024; **52**(D1): D134–D137.
PubMed Abstract | Publisher Full Text | Free Full Text

Shaw MR: **Host ranges of *Aleiodes* species (Hymenoptera: Braconidae), and an evolutionary hypothesis.** In: Melika, G. and Thuróczy, C. (eds.) *Parasitic wasps: Evolution, Systematics, Biodiversity and Biological Control.* Budapest, Hungary: Agroinform, 2002; 321–327.
Reference Source

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

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
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.
PubMed Abstract | Publisher Full Text | Free 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

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

Wellcome Sanger Institute: **The genome sequence of a braconid wasp, *Aleiodes alternator* (Nees, 1834).** European Nucleotide Archive. [dataset], accession number PRJEB70745, 2024.

Zaldívar-Riverón A, Shaw MR, Sáez AG, *et al.*: **Evolution of the parasitic wasp subfamily Rogadinae (Braconidae): phylogeny and evolution of lepidopteran host ranges and mummy characteristics.** *BMC Evol Biol.* 2008; **8**(1): 329.
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