



DATA NOTE

The genome sequence of the mirid plant bug, *Closterotomus trivialis* (A.Costa, 1853) (Hemiptera: Miridae)

[version 1; peer review: 2 approved]

Laurence Livermore¹, Maxwell V. L. Barclay ¹,
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

¹Natural History Museum, London, England, UK

V1 First published: 15 Sep 2025, 10:514
<https://doi.org/10.12688/wellcomeopenres.24886.1>
Latest published: 15 Sep 2025, 10:514
<https://doi.org/10.12688/wellcomeopenres.24886.1>

Abstract

We present a genome assembly from an individual *Closterotomus trivialis* (mirid plant bug; Arthropoda; Insecta; Hemiptera; Miridae). The genome sequence has a total length of 1 163.70 megabases. Most of the assembly (94.03%) is scaffolded into 16 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 19.42 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

Closterotomus trivialis; mirid plant bug; genome sequence; chromosomal; Hemiptera



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

Open Peer Review

Approval Status

	1	2
version 1		
15 Sep 2025	view	view

1. **Roberto Biello** , John Innes Centre
Department of Crop Genetics, Norwich, UK
2. **Zheng Li** , Miami University Department
of Biology, Oxford, USA

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: Livermore L: Writing – Original Draft Preparation; Barclay MVL: Investigation, Resources, 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.

Copyright: © 2025 Livermore L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Livermore L, Barclay MVL, Natural History Museum Genome Acquisition Lab *et al.* **The genome sequence of the mirid plant bug, *Closterotomus trivialis* (A.Costa, 1853) (Hemiptera: Miridae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, 10:514 <https://doi.org/10.12688/wellcomeopenres.24886.1>

First published: 15 Sep 2025, 10:514 <https://doi.org/10.12688/wellcomeopenres.24886.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Paraneoptera; Hemiptera; Prosorrhyncha; Heteroptera; Euheteroptera; Neoheteroptera; Panheteroptera; Cimicomorpha; Cimicoidea; Miridae; Mirinae; Mirini; *Closterotomus*; *Closterotomus trivialis* (Costa, 1853) (NCBI:txid2026310)

Background

Closterotomus trivialis (Costa, 1853) is a plant bug (Heteroptera: Miridae) with a length of 6.5–8 mm (including wings) and around 2–3 mm wide. Females are slightly larger than the males and differ in their colouration – both are macropterous. Both sexes undergo a progressive transition from lighter greenish and brown colouration to darker colouration (e.g., dark brown to blackish) as they mature, with males exhibiting more uniform darker shades over their entire dorsal surface while the females retain green highlights on their head, pronotum, scutellum and edges of the hemelytron. In both sexes, although more frequently in males, mature individuals have a mostly red cuneus. The wing membrane is a dark, patchy brown, with venation that darkens with age from yellowish brown to reddish. The dorsal surface of the body is covered with a fine light pubescence with some sparse black hairs. The head (including eyes) is as wide as it is long, the eyes are large and globular, and the antennae are approximately 5/6 of the body length with the first segment shorter and thicker than the subsequent segments (Bantock & Botting, 2018; Barbagallo, 1970). In the UK this species is distinctive, especially with its colouration when mature, but may be confused with *Pantilius tunicatus* (Fabricius, 1781). The latter has similar bright colouration but with dark dorsal spots/punctuation, and a striking light-to-dark gradient on antennal segments III and IV. Wagner and Weber (1964) provide descriptions and keys to all *Closterotomus* species occurring in the UK.

Originally described as *Phytocoris trivialis* by Costa in 1853 (Costa, 1853: 259), and subsequently transferred to the genus *Calocoris* Fieber, 1858, it was included in the resurrected genus *Closterotomus* Fieber, 1858 as part of Rosenzweig's revision of the *Calocoris* complex (Kerzhner & Josifov, 1999: 89; Rosenzweig, 1997).

Closterotomus trivialis has one generation a year and lays its eggs in the wood of its host plants, often in dry crevices (Barbagallo, 1970). The eggs overwinter and in the UK the adults have been recorded from May to July, mostly frequently observed in June (NBN Atlas Partnership, 2024).

Closterotomus trivialis is recently established in the UK and was first reported from a discrete London population in 2009, originally thought to be feeding on ornamental *Hypericum* in gardens but subsequently observed to be polyphagous and feeding on native flora (Bantock, 2010; Flanagan, 2013). Since 2009 it has expanded its UK range to other cities, and has been

recorded as far north as Windermere (Flanagan, 2013; NBN Atlas Partnership, 2024) and in County Cork, Ireland (van der Noll & Nelson, 2017). It has a broad distribution in Europe, North Africa and Asia, and in recent years has expanded its northerly range in Europe (GBIF Secretariat, 2024; Kerzhner & Josifov, 1999: 89). It has been recorded as being introduced with imported plant material (Aukema, 2003: 46).

Closterotomus trivialis is a pest species on both citrus and olive trees (Barbagallo, 1970). On olive trees it can cause severe damage from feeding on the flowers, especially the young flowers, causing bud abortion and the dropping of fruiting organs (Perdikis *et al.*, 2009).

We present a chromosomally complete genome sequence for *Closterotomus trivialis*, as part of the Darwin Tree of Life project. The assembly was produced using the Tree of Life pipeline from a specimen collected in Tynemouth Street, England, United Kingdom (Figure 1).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was a *Closterotomus trivialis* (specimen ID NHMUK014400242, ToLID ihCloTriv1; Figure 1), collected from Tynemouth Street, England, United Kingdom (latitude 51.4725, longitude -0.1858) on 2021-05-12. Another specimen was used for RNA sequencing (specimen ID NHMUK014440604, ToLID ihCloTriv2). It was collected from Wetherby Gardens, Kensington, London,



Figure 1. Photographs of the *Closterotomus trivialis* (ihCloTriv1) specimen used for genome sequencing.

England, United Kingdom (latitude 51.49, longitude -0.185) on 2022-05-25. Both specimens were collected and identified by Maxwell Barclay (Natural History Museum). Sample metadata were collected in line with the Darwin Tree of Life project standards described by [Lawniczak et al. \(2022\)](#).

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 ihCloTriv1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the abdomen was homogenised by [powermashing](#) using a PowerMasher II tissue disruptor.

HMW DNA was extracted 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 was purified by [automated 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.

RNA was extracted from whole organism tissue of ihCloTriv2 in the Tree of Life Laboratory at the WSI using the [RNA Extraction: Automated MagMax™ mirVana 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 using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen head and thorax tissue of the ihCloTriv1 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 and 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 (Cheng *et al.*, 2021) with the `--primary` option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the `--break` option for handling potential misassemblies. The scaffolded assemblies were 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) 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 116 breaks, 321 joins, and removal of 18 haplotypic duplications. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextView was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate k -mer completeness and assembly quality for the primary and alternate 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 *Closterotomus trivialis* specimen generated 55.98 Gb (gigabases) from 5.01 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 030.86 Mb, with a heterozygosity of 1.48% and repeat content of 41.11% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 32× coverage. Hi-C sequencing produced 136.08 Gb from 901.21 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 primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC

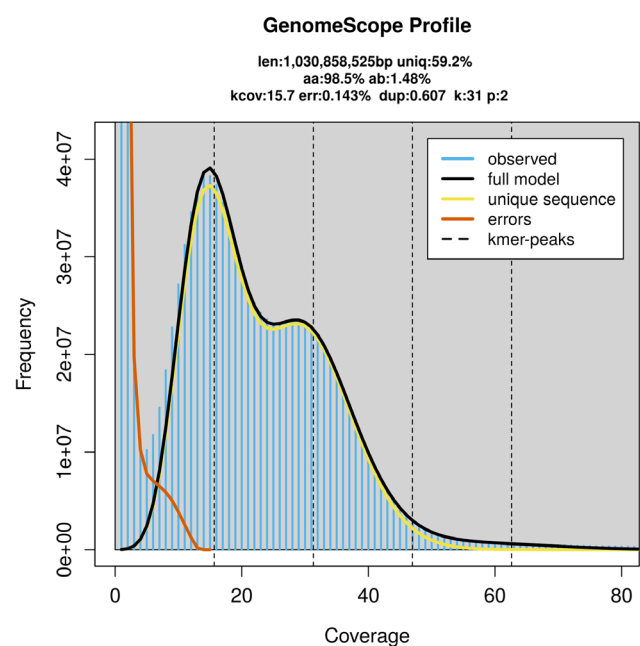


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB53562.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	ihCloTriv1	ihCloTriv1	ihCloTriv2
Specimen ID	NHMUK014400242	NHMUK014400242	NHMUK014440604
BioSample (source individual)	SAMEA9359418	SAMEA9359418	SAMEA114805944
BioSample (tissue)	SAMEA9359482	SAMEA9359516	SAMEA114806143
Tissue	abdomen	head thorax	whole organism
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq X
Run accessions	ERR9871433; ERR9863248	ERR9866449	ERR13962503
Read count total	5.01 million	901.21 million	71.72 million
Base count total	55.98 Gb	136.08 Gb	10.83 Gb

databases. The final assembly has a total length of 1 163.70 Mb in 752 scaffolds, with 985 gaps, and a scaffold N50 of 65.62 Mb (Table 2).

Most of the assembly sequence (94.03%) was assigned to 16 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; 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.

The combined primary and alternate assemblies achieve an estimated QV of 60.1. The *k*-mer completeness is 80.53% for the primary assembly, 68.34% for the alternate haplotype, and 98.72% for the combined assemblies (Figure 4).

BUSCO v.5.7.1 analysis using the hemiptera_odb10 reference set (*n* = 2 510) identified 96.7% of the expected gene set (single = 94.4%, duplicated = 2.3%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **6.C.Q59**, meeting the recommended reference standard.

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

Table 2. Genome assembly statistics.

Assembly name	ihCloTriv1.1
Assembly accession	GCA_965211935.1
Alternate haplotype accession	GCA_965211925.1
Assembly level	chromosome
Span (Mb)	1 163.70
Number of chromosomes	16
Number of contigs	1 737
Contig N50	3.34 Mb
Number of scaffolds	752
Scaffold N50	65.62 Mb
Sex chromosomes	Not identified
Organelles	Mitochondrion: 19.42 kb

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)

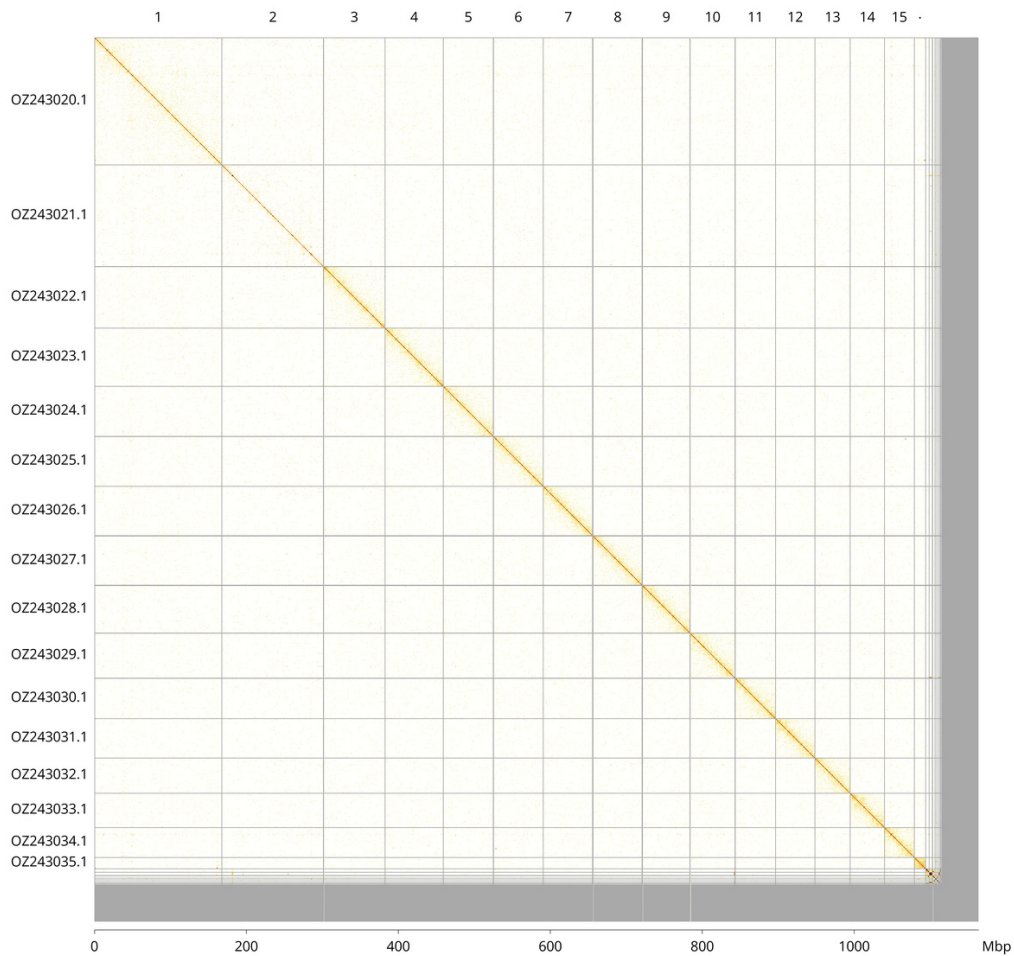


Figure 3. Hi-C contact map of the *Closterotomus trivialis* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Closterotomus trivialis* ihCloTriv1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ243020.1	1	167.81	41
OZ243021.1	2	133.95	40.50
OZ243022.1	3	80.96	40.50
OZ243023.1	4	76.44	40
OZ243024.1	5	66.06	40
OZ243025.1	6	65.62	40
OZ243026.1	7	65.36	40
OZ243027.1	8	65.32	40
OZ243028.1	9	62.73	40

INSDC accession	Molecule	Length (Mb)	GC%
OZ243029.1	10	59.16	40
OZ243030.1	11	53.26	40.50
OZ243031.1	12	51.88	40
OZ243032.1	13	46.08	40
OZ243033.1	14	45.12	40.50
OZ243034.1	15	39.61	40
OZ243035.1	16	14.90	39.50

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.

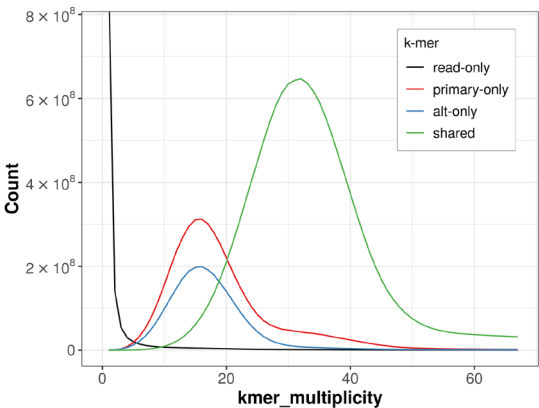


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

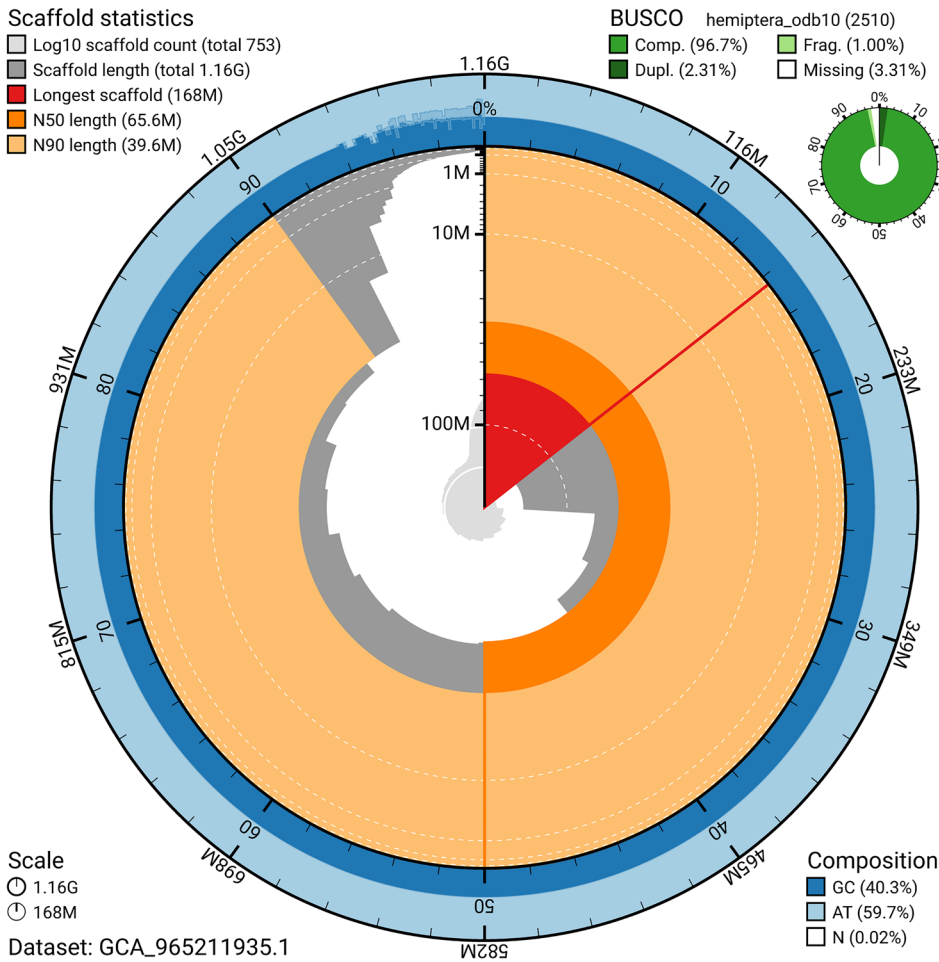


Figure 5. Assembly metrics for ihCloTriv1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 hemiptera_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the [BlobToolKit viewer](#).

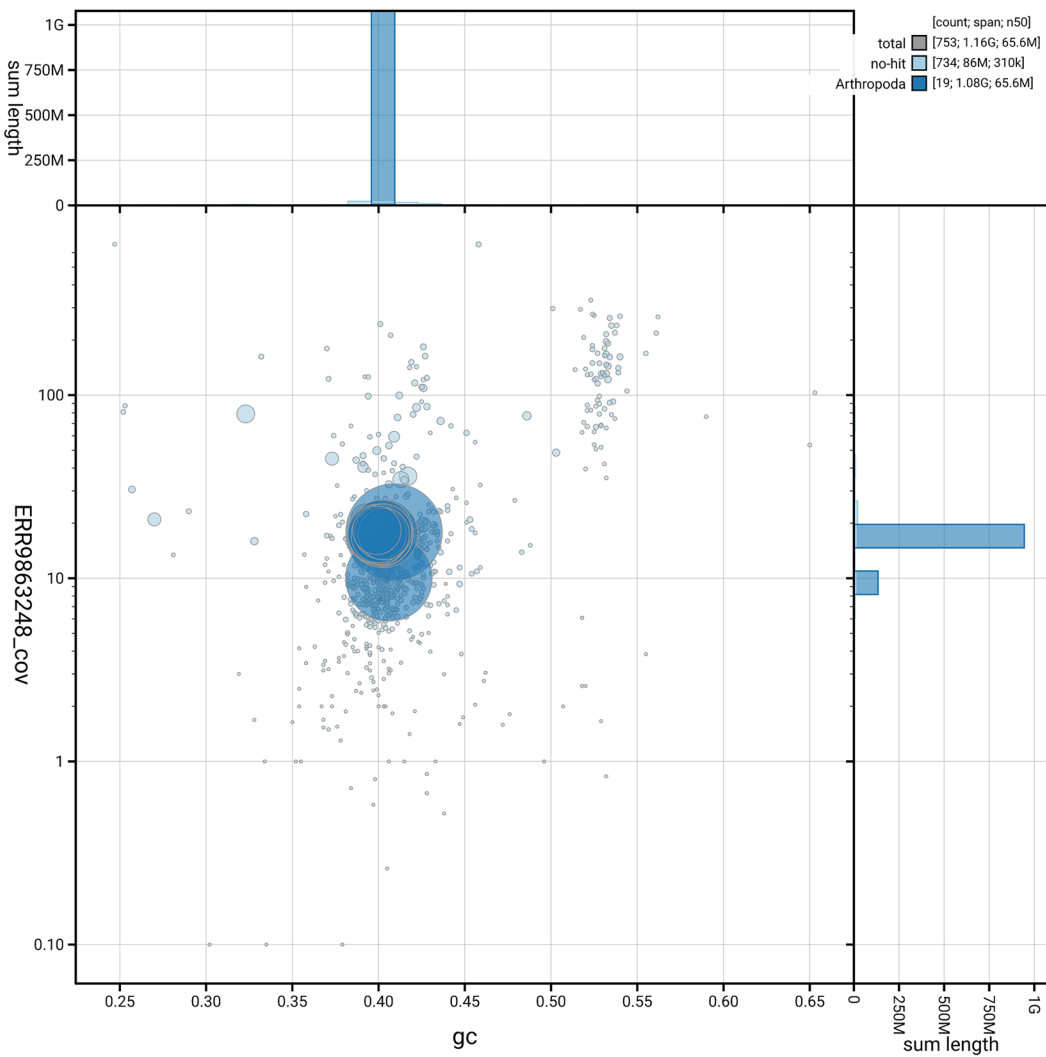


Figure 6. BlobToolKit GC-coverage plot for ihCloTriv1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Closterotomus trivialis* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q59	6.C.Q40
Contig N50 length	3.34 Mb	≥ 1 Mb
Scaffold N50 length	65.62 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.4; alternate: 61.0; combined: 60.1	≥ 40
<i>k</i> -mer completeness	Primary: 80.53%; alternate: 68.34%; combined: 98.72%	≥ 95%
BUSCO	C:96.7% [S:94.4%; D:2.3%]; F:1.0%; M:2.3%; n:2 510	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	94.03%	≥ 90%

Data availability

European Nucleotide Archive: *Closterotomus trivialis*. Accession number [PRJEB53562](#). The genome sequence is released openly for reuse. The *Closterotomus trivialis* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and 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 [Table 1](#) and [Table 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
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
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	0.16.1-r375	https://github.com/chhyllp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MercuryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	2	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	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
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	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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](#)

References

- Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Aukema B: **Recent changes in the Dutch Heteroptera fauna (Insecta: Hemiptera).** In: *Changes in Ranges: Invertebrates on the Move. Proceedings of the 13th International Colloquium of the European Invertebrate Survey, Leiden, 2–5 September 2001.* 2003; 39–52.
[Reference Source](#)
- Bantock T: ***Closterotomus trivialis* (Costa) in North London.** *Het News, 2nd Series.* 2010; (15): 3.
[Reference Source](#)
- Bantock T, Botting J: **(Miridae) *Closterotomus trivialis*.** 2018.
[Reference Source](#)
- Barbagallo S: **Contributo alla conoscenza del *Calocoris* (*Closterotomus*) *trivialis* (Costa) (Rhynchota-Heteroptera, Miridae).** *Morfologia dell'adulto e biologia.* *Entomologica.* 1970; **6**: 1–104.
[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](#)
- 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, 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](#)
- Costa A: ***Cimicum regni neapolitani centuria tertia et quarta fragmentum.*** Napoli, 1853.
[Reference Source](#)
- 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](#)
- 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](#)
- 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](#)
- Flanagan J: ***Closterotomus trivialis* in N of England.** *Het News, 2nd Series.* 2013; (19–20): 9.
[Reference Source](#)
- 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](#)
- GBIF Secretariat: **GBIF occurrence download.** 2024.
[Publisher Full Text](#)
- Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–98.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kerzhner IM, Josifov M: **Miridae Hahn, 1833.** In: B. Aukema and C. Rieger (eds), *Catalogue of the Heteroptera of the Palaearctic Region, Volume 3. Cimicomorpha II.* Netherlands Entomological Society, 1999; i–xiv; 1–577.
- 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](#)
- Lawniczak 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, Seppely 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](#)
- NBN Atlas Partnership: ***Closterotomus trivialis* (A. Costa, 1853) map on the NBN Atlas.** 2024.
[Reference Source](#)
- Perdikis D, Garantonakis N, Kitsis P, *et al.*: **Studies on the damage potential of *Closterotomus trivialis* and *Aphanosoma italicum* on olive fruit setting.** *Bulletin of Insectology.* 2009; **62**(2): 215–19.
[Reference Source](#)
- 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](#)

Rosenzweig VY: Revised classification of the *Calocoris* complex and related genera (Heteroptera: Miridae). *Zoosystematica Rossica*. 1997; **6**(1–2): 139–69.

[Reference Source](#)

Schoch CL, Ciufo S, Domrachev M, et al.: NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*. 2020; **2020**: baaa062.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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 der Noll L, Nelson B: *Deraeocoris ruber* (L.) And *Closterotomus trivialis* (Costa) two plant bugs (Hemiptera, Miridae) new to Ireland from Co. Cork. *The Irish Naturalists' Journal*. 2017; **35**(2): 133–34.

[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](#)

Wagner E, Weber HH: *Heteropteres Miridae*. *Faune de France* **67**. 1964.

[Reference Source](#)

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 14 October 2025

<https://doi.org/10.21956/wellcomeopenres.27413.r133366>

© 2025 Li Z. 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.



Zheng Li 

Miami University Department of Biology, Oxford, Ohio, USA

The authors present a 1,163.70 megabase genome assembly of *Closterotomus trivialis* (mirid plant bug) using PacBio HiFi long reads and HIC sequencing data with total sequencing coverage of 32x. The genome is of high quality, with 94.03% percent of it localized on 16 chromosomal pseudomolecules. They also assembled the mitochondrial genome, which is approximately 19.42 kilobases in length. This genome represents a valuable resource for understanding the biology of an emerging agricultural pest species affecting citrus and olive production. It can also help in performing comparative genomic analysis to understand the evolutionary history of Miridae. The authors provided a solid background on sexual dimorphism, anatomical and life cycle distribution, and pest impact. The method and metadata are clear, and the results are well documented.

Here are some points that can be improved in the manuscript:

- In Figure 5, the assembly metrics for ihCloTriv1.1 are presented, including BUSCO statistics. However, the figure shows 3.31% of BUSCOs as missing, while Table 4 reports M: 2.3%. This discrepancy should be checked and clarified.
- While authors performed a BlobToolKit GC-coverage plot for ihCloTriv1.1, the authors did not include a cumulative sequence length plot, which is commonly presented in genome papers.

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: Genome evolution

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 07 October 2025

<https://doi.org/10.21956/wellcomeopenres.27413.r133362>

© 2025 Biello R. 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.



Roberto Biello 

John Innes Centre Department of Crop Genetics, Norwich, England, UK

The authors present a chromosome-level genome assembly for *Closterotomus trivialis*. This represents an important resource for studying the evolution and invasive potential of this pest species, which affects both citrus and olive trees. Sequencing was performed using PacBio HiFi and Hi-C technologies, and both primary and alternate assemblies were generated. The assembly meets high-quality standards, as demonstrated by high BUSCO scores and k-mer completeness. The methods are clearly described, and all associated data are publicly available.

Minor comments:

- Given that both HiFi and Hi-C data were available from the same individual, it would be interesting to understand why the authors chose to assemble the genome using Hifiasm with the -primary option, rather than performing a haplotype-resolved (diploid) assembly, which might have provided additional insights into genomic features such as structural variation.
- The authors should clarify the sex of the specimen sequenced and indicate the expected sex chromosomes in this species, as this information is relevant for interpreting the assembly and chromosome composition.

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: Genome assembly and annotation, Evolution, 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.
