The genome sequence of the kissing bug, *Triatoma infestans* (Klug, 1834) (Hemiptera: Reduviidae)

Antonella Bacigalupo[1]
Conceptualization; Investigation; Methodology; Writing – original draft; Writing – review & editing

Esteban San Juan[2, 3]
Investigation; Methodology; Writing – review & editing

Carla Ponce-Revello[2, 3]
Investigation; Writing – review & editing

Martin S. Llewellyn[1]
Resources; Writing – review & editing

José A. Segura[4]
Investigation; Writing – review & editing

Alonso Parra[5]
Investigation; Writing – review & editing

Luis E. Castañeda[6, 3]
Resources; Writing – review & editing

Carezza Botto-Mahan[2, 3]
Conceptualization; Investigation; Resources; Writing – review & editing

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

Mark L. Blaxter[7†]
Supervision

[1] School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, Scotland, UK
[2] Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile
[3] Research Ring in Pest Insects and Climate Change (PIC2), Chile
[4] Office of Zoonosis and Vector Control, Department of Public Health, Regional Ministerial Secretariat of Health of the Metropolitan Region, Santiago, Chile
[5] Office of Zoonosis and Vector Control, Division of Healthy Public Policies and Promotion, Undersecretariat of Public Health, Ministry of Health, Santiago, Chile
[6] Núcleo Interdisciplinario de Biología y Genética, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Santiago, Chile
[7] Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

[†]Correspondence: Mark.Blaxter@sanger.ac.uk

Keywords: Triatoma infestans; kissing bug; genome sequence; chromosomal; Hemiptera

We present a genome assembly from an individual male *Triatoma infestans* (kissing bug; Arthropoda; Insecta; Hemiptera; Reduviidae). The assembly contains two haplotypes with total lengths of 1 411.12 megabases and 1 199.34 megabases. Most of haplotype 1 (84.29%) is scaffolded into 12 chromosomal pseudomolecules, including the X and Y sex chromosomes. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 17.05 kilobases.

Main body

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; Reduvioidea; Reduviidae; Triatominae; *Triatoma*; *Triatoma infestans* (Klug, 1834) (NCBI:txid30076)

Background

Triatoma infestans (Hemiptera: Reduviidae: Triatominae) is a hematophagous hemimetabolous insect that transmits the parasite *Trypanosoma cruzi* (Chacón *et al.*, 2022a), which causes Chagas disease in humans (WHO, 2015). Chagas disease can have acute manifestations, but these are usually nonspecific, and the disease progresses into a chronic phase, with parasites affecting especially the cardiac and digestive muscles (WHO, 2015). The adults of *T. infestans* are macropterous, with head length similar to the pronotum length, entirely black pronotum, black legs with the base of the femora and trochanter yellow, and the abdomen's connexivum with transversal wide dark markings (Lent & Wygodzinsky, 1979). They reproduce by laying eggs, and the five nymphal stages, also hematophagous, are characterised by a pointed tegument, and legs with yellow and dark band markings (Brewer *et al.*, 1981).

This species constitutes the main *T. cruzi* vector in the southern cone of South America. *Triatoma infestans* has a high level of domiciliation, which explains its success as a human vector (Waleckx *et al.*, 2015). The ecotopes described for this species, apart from human dwellings, include bromeliads (Bacigalupo *et al.*, 2006), rock piles (Bacigalupo *et al.*, 2010), rocks, prickly pears, cacti, cliffs, cracks in the ground, tree holes, bird nests, piles of branches (Brenière *et al.*, 2017), chicken coops, and domestic mammal corrals (Gürtler *et al.*, 2014).

Currently, *T. infestans* is distributed in Argentina, Bolivia, Brazil, Chile, Paraguay, and Peru, with recent introduction in Mexico (Belisário *et al.*, 2017; Brenière *et al.*, 2017; Bustamante Gomez *et al.*, 2016; Delgado *et al.*, 2013; Martínez-Hernández *et al.*, 2022; Rolón *et al.*, 2011; Tapia-Garay *et al.*, 2018). Transmission of *T. cruzi* by *T. infestans* was declared interrupted in Uruguay in 1997 (WHO, 2015), and no recent reports of residual foci were found in the literature. Interruption of transmission was also

declared for Chile in 1999 (WHO, 2015), but sylvatic foci of this species have been detected (Ihle-Soto *et al.*, 2019).

Although being described as a nocturnal species, *T. infestans* infected by *T. cruzi* modify their behaviour, moving significantly more than non-infected specimens, and increasing the distance travelled, especially during the photophase (Chacón *et al.*, 2022b). Infection also affects the feeding and defecation behaviours, reducing the time to locate the host, increasing the number of bites, and reducing the time to emit dejections (Chacón *et al.*, 2022a). Environmental conditions also affect feeding, with lower temperatures reducing feeding frequency compared to warmer temperatures, and temperature variability modulating these effects (Álvarez-Duhart *et al.*, 2024). This species can produce one or two generations per year under laboratory conditions, depending on its feeding source (Guarneri *et al.*, 2000). Dispersal of adults by flight occurs when the temperature at sunset is equal to or above 20 °C (Di Iorio & Gürtler, 2017); nymphs and adults can disperse by walking (Abrahan *et al.*, 2011).

Cytogenetics of *T. infestans* show 20 autosomes and two sexual chromosomes (Panzera *et al.*, 2010). This species shows a remarkable variability in genome size, with flow cytometry measurements of the Andean group showing the highest C-values – 1.98 pg – and the non-Andean group with 1.09-1.52 pg, with differences consisting mainly of the amount of heterochromatin and the location of ribosomal DNA (Panzera *et al.*, 2010, 2014). The Andean group comprises populations from the Andean areas of Bolivia, Peru, and the northernmost areas of Chile, while the Non-Andean group includes Bolivian populations from the Chaco, Argentina, Brazil, north-central Chile, Paraguay, and Uruguay (Panzera *et al.*, 2014; Torres-Pérez *et al.*, 2011). The comparison of the repetitive DNA – the repeatome – of both groups indicates that satellite DNA accounts for most of the genome size variation among them (Pita *et al.*, 2017a).

The previous genomic resources for this species include a partial scaffold-level assembly (GCA_011037195.1, Justi *et al.*, 2020) and complete mitochondrial genomes (Aguilera-Uribe *et al.*, 2020; Fernández & García, 2022; Pita *et al.*, 2017b). Given the relevance of *T. infestans* as the main vector of *T. cruzi*, it is of utmost importance to obtain an annotated chromosomal-level whole genome assembly that would guide future vector research, to explore the genetic diversity of different populations, monitor the genetic basis of insecticide resistance, and identify sources of recent introductions, reducing the neglect of Chagas disease.

This chromosome-level assembly was produced using the Tree of Life pipeline from a specimen collected in Fundo La Batalla, Calera De Tango, Maipo, Región Metropolitana, Chile (Figure 1).

Figure 1: Photograph of an adult male of *Triatoma infestans* (not the specimen used for genome sequencing). Photograph by Carla Ponce-Revello.

Methods

Sample acquisition

The specimen used for genome sequencing was an adult male *Triatoma infestans* (specimen ID SAN20002957, ToLID ihTriInfe1; Figure 1), collected from Fundo La Batalla, Calera De Tango, Maipo, Región Metropolitana, Chile (latitude -33.6553, longitude -70.7882) on 2024-02-23. The specimen was collected and identified by Carezza Botto (Universidad De Chile).

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 ihTriInfe1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the thorax was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Manual MagAttract v3 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. For this sample, the final post-shearing DNA had a Qubit concentration of 8.38 ng/μL and a yield of 3 352.00 ng.

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 25M 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 tissue from the head of the ihTriInfe1 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 X.

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 MERQURY.FK (Rhie *et al.*, 2020). The organelle genomes were assembled using MitoHiFi (Uliano-Silva *et al.*, 2023).

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 3 breaks and 117 joins. This reduced the scaffold count by 10.8% and increased the scaffold N50 by 1.7%. 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 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 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 *Triatoma infestans* specimen generated 82.90 Gb (gigabases) from 6.10 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 243.99 Mb, with a heterozygosity of 0.60% and repeat content of 36.51% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 64× coverage. Hi-C sequencing produced 88.62 Gb from 586.90 million reads, which were used to scaffold the assembly. Table 1 summarises the specimen and sequencing details.

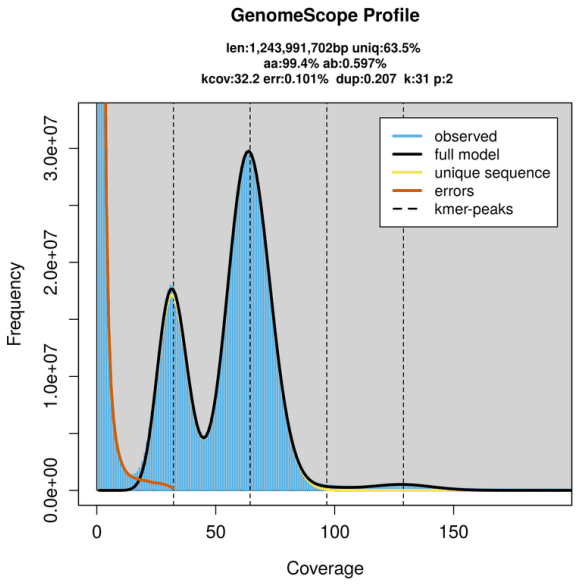

Figure 2: Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1: Specimen and sequencing data for BioProject PRJEB87951

Platform	PacBio HiFi	Hi-C
ToLID	ihTriInfe1	ihTriInfe1
Specimen ID	SAN20002957	SAN20002957
BioSample (source individual)	SAMEA115718423	SAMEA115718423
BioSample (tissue)	SAMEA115718429	SAMEA115718428
Tissue	thorax	head
Instrument	Revio	Illumina NovaSeq X
Run accessions	ERR14835903	ERR14835491
Read count total	6.10 million	586.90 million
Base count total	82.90 Gb	88.62 Gb

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 1 411.12 Mb in 691 scaffolds, with 214 gaps, and a scaffold N50 of 93.57 Mb (Table 2).

Table 2: Genome assembly statistics

Assembly name	ihTriInfe1.hap1.1	ihTriInfe1.hap2.1
Assembly accession	GCA_965641795.1	GCA_965641865.1
Assembly level	chromosome	scaffold
Span (Mb)	1 411.12	1 199.34
Number of chromosomes	12	scaffold-level
Number of contigs	905	1 430
Contig N50	24.93 Mb	30.94 Mb
Number of scaffolds	691	1 331
Scaffold N50	93.57 Mb	95.94 Mb
Longest scaffold length (Mb)	146.63	-
Sex chromosomes	X and Y	-
Organelles	Mitochondrion: 17.05 kb	-

Most of the haplotype 1 assembly sequence (84.29%) was assigned to 12 chromosomal-level scaffolds, representing 10 autosomes and the X and Y sex chromosomes. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). Scaffolds along the length of chromosome Y and on chromosome 2 from ~134.70–138.19Mb have uncertain order and orientation.

The mitochondrial genome was also assembled (length 17.05 kb, OZ282549.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

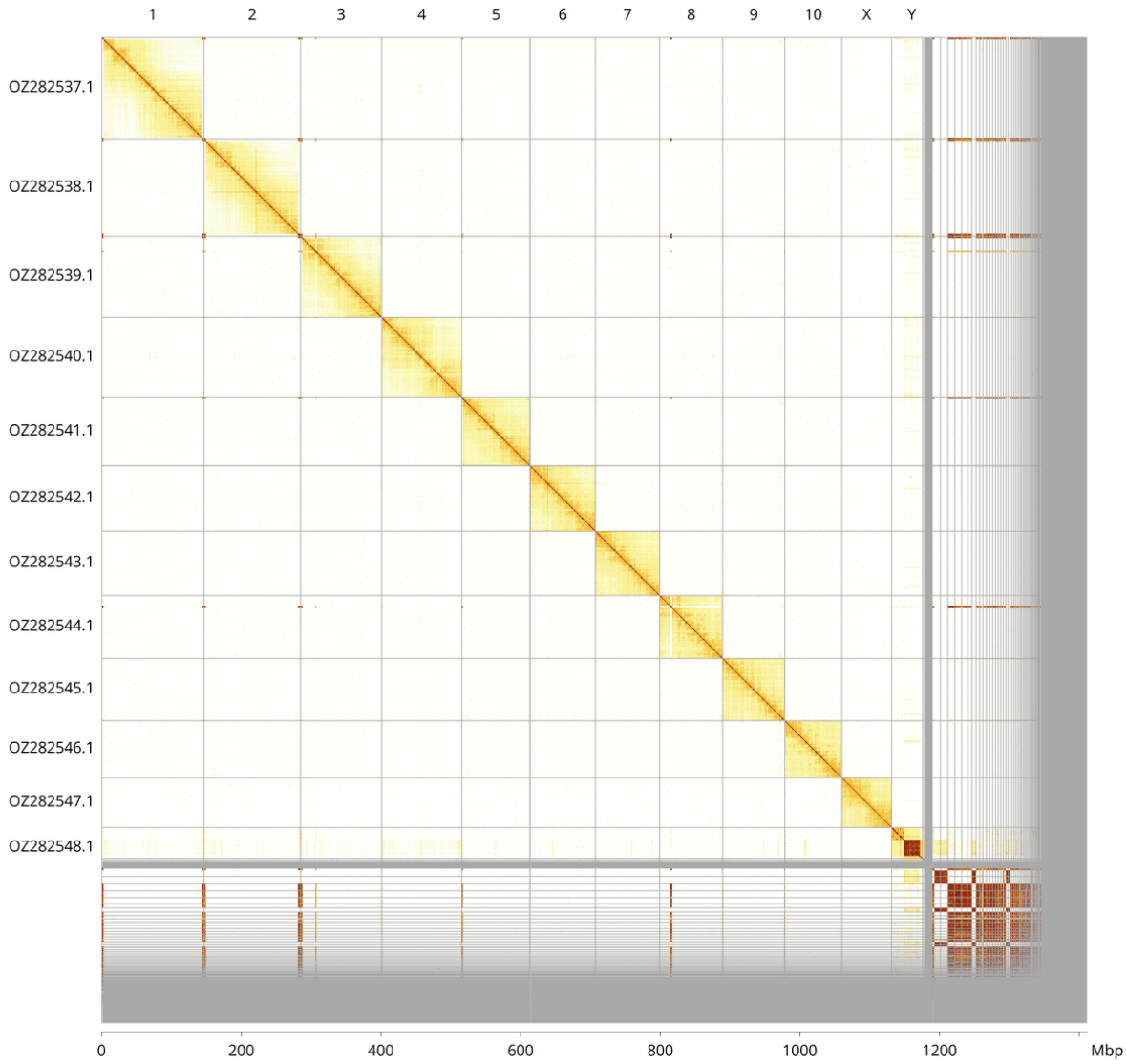

Figure 3: Hi-C contact map of the *Triatoma infestans* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3: Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Triatoma infestans* ihTriInfe1

INSDC accession	Molecule	Length (Mb)	GC%
OZ282537.1	1	146.63	34
OZ282538.1	2	138.26	34
OZ282539.1	3	116.12	34

INSDC accession	Molecule	Length (Mb)	GC%
OZ282540.1	4	114.87	34
OZ282541.1	5	97.85	34
OZ282542.1	6	93.57	34
OZ282543.1	7	91.98	34
OZ282544.1	8	90.21	34
OZ282545.1	9	88.92	34
OZ282546.1	10	81.83	34
OZ282547.1	X	71.18	36
OZ282548.1	Y	57.98	31

For haplotype 1, the estimated QV is 63.2, and for haplotype 2, 61.8. When the two haplotypes are combined, the assembly achieves an estimated QV of 62.5. The *k*-mer completeness is 88.23% for haplotype 1, 82.21% for haplotype 2, and 99.42% for the combined haplotypes (Figure 4).

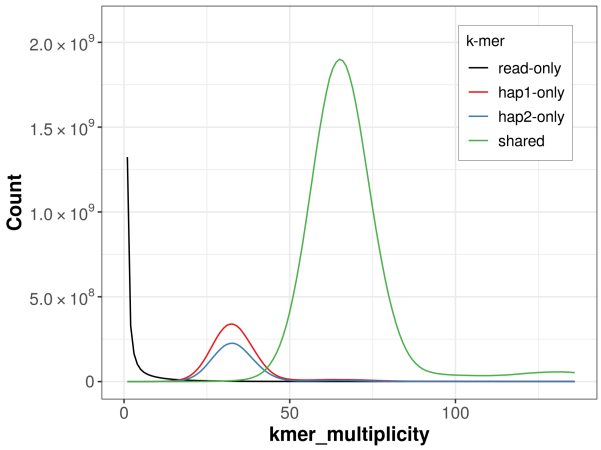

Figure 4: Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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 hemiptera_odb10 reference set (n = 2 510) identified 99.0% of the expected gene set (single = 97.9%, duplicated = 1.1%) 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.

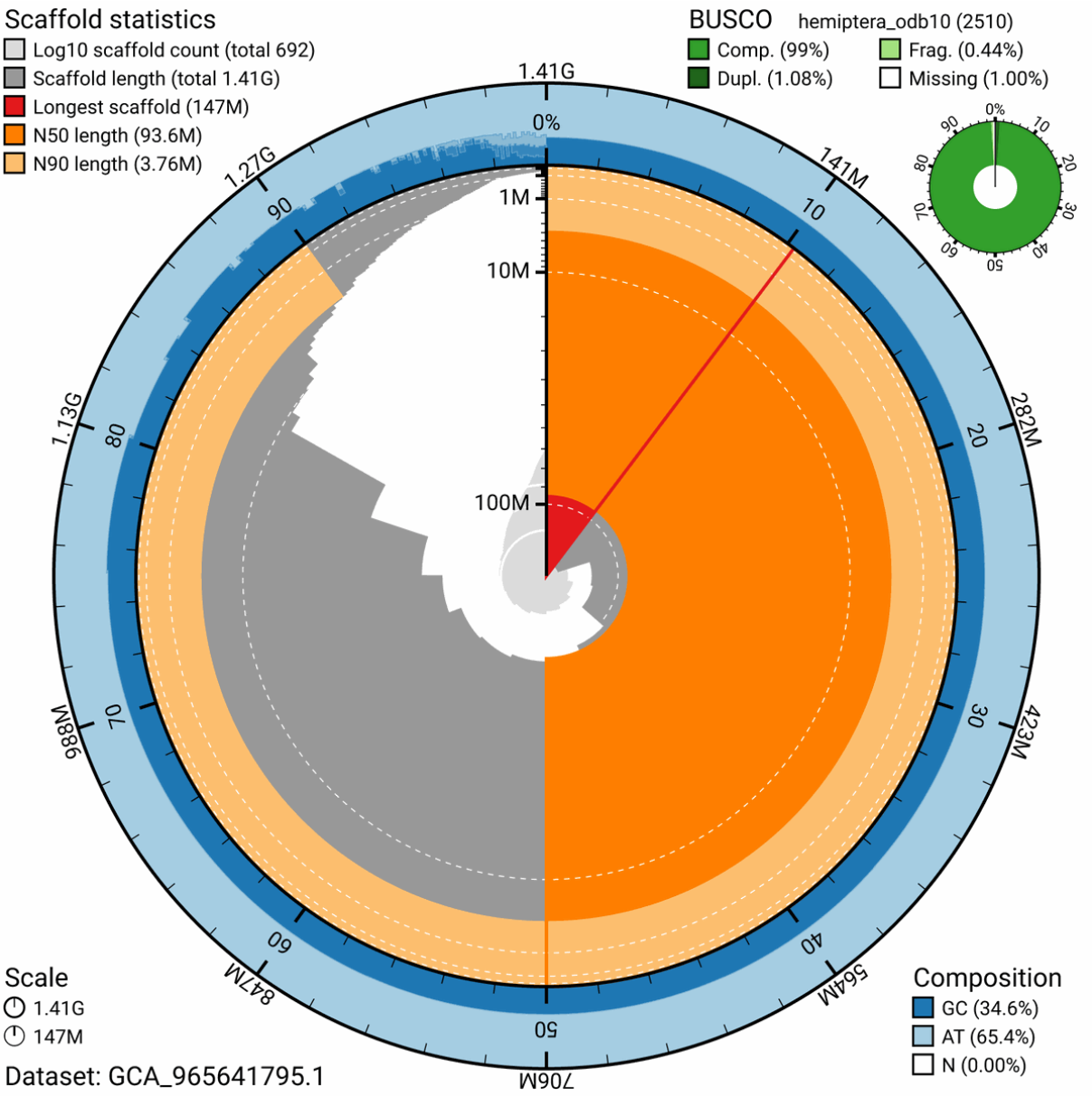

Figure 5: Assembly metrics for ihTriInfe1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost 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.

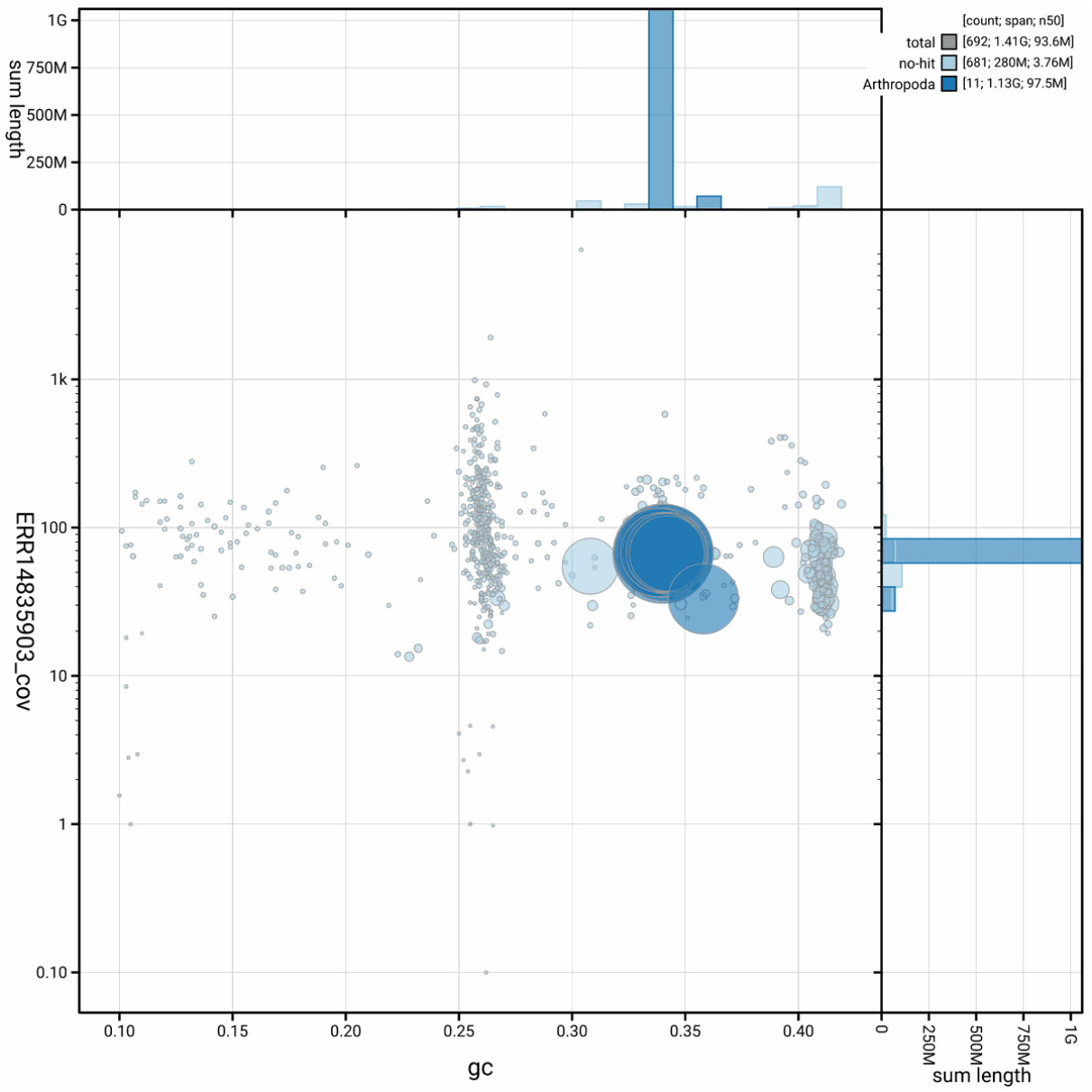

Figure 6: BlobToolKit GC-coverage plot for ihTriInfe1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes 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 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 **7.7.Q63**.

Table 4: Earth Biogenome Project summary metrics for the *Triatoma infestans* assembly

Measure	Value	Benchmark
EBP summary (haplotype 1)	7.7.Q63	6.C.Q40
Contig N50 length	24.93 Mb	≥ 1 Mb
Scaffold N50 length	93.57 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 63.2; haplotype 2: 61.8; combined: 62.5	≥ 40
k-mer completeness	Haplotype 1: 88.23%; Haplotype 2: 82.21%; combined: 99.42%	≥ 95%
BUSCO	C:99.0% [S:97.9%; D:1.1%]; F:0.4%; M:0.6%; n:2 510	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	84.29%	≥ 90%

Data availability

European Nucleotide Archive: Triatoma infestans (kissing bug). Accession number PRJEB87951. The genome sequence is released openly for reuse. The *Triatoma infestans* genome sequencing initiative is part of 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
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.8.3	https://gitlab.com/ezlab/busco

Software	Version	Source
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
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	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	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	v0.8.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2.2	https://github.com/c-zhou/yahs

Grant information

This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540). *Triatoma infestans* capture was supported by the ANID ATE 230025 Project.

Author information

Contributors are listed at the following links:

- 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

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.

References

Abrahan LB, Gorla DE, Catalá SS: Dispersal of *Triatoma infestans* and other triatominae species in the arid Chaco of Argentina: Flying, walking or passive carriage? The importance of walking females. *Mem Inst Oswaldo Cruz*. 2011; **106**(2): 232–39. https://doi.org/10.1590/s0074-02762011000200019

Aguilera-Uribe M, Meza-Lázaro RN, Kieran TJ, *et al.*: Phylogeny of the North-Central American clade of blood-sucking reduviid bugs of the tribe Triatomini (Hemiptera: Triatominae) based on the mitochondrial genome. *Infect Genet Evol*. 2020; **84**: 104373. https://doi.org/10.1016/j.meegid.2020.104373

Altschul SF, Gish W, Miller W, *et al.*: Basic Local Alignment Search Tool. *Journal of Molecular Biology*. 1990; **215**(3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

Álvarez-Duhart B, Cavieres G, González A, *et al.*: Influence of temperature variability on the feeding behavior and blood consumption of *Triatoma infestans* (Hemiptera: Reduviidae). *J Med Entomol*. 2024; **61**(5): 1222–31. https://doi.org/10.1093/jme/tjae034

Bacigalupo A, Segura JA, García A, *et al.*: First finding of Chagas disease vectors associated with wild bushes in the Metropolitan Region of Chile. *Rev Med Chil*. 2006; **134**(10): 1230–36. https://doi.org/10.4067/s0034-98872006001000003

Bacigalupo A, Torres-Pérez F, Segovia V, *et al.*: Sylvatic foci of the Chagas disease vector *Triatoma infestans* in Chile: Description of a new focus and challenges for control programs. *Mem Inst Oswaldo Cruz*. 2010; **105**(5): 633–41. https://doi.org/10.1590/s0074-02762010000500006

Bateman A, Martin M-J, Orchard S, *et al.*: UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*. 2023; **51**(D1): D523–31. https://doi.org/10.1093/nar/gkac1052 PMID: 36408920 PMCID: PMC9825514

Belisário CJ, Pessoa GC, Silva EM, *et al.*: Genetic characterization of residual *Triatoma infestans* populations from Brazil by microsatellite. *Genetica*. 2017; **145**(1): 105–14. https://doi.org/10.1007/s10709-017-9949-y

Brenière SF, Buitrago R, Waleckx E, *et al.*: Wild populations of *Triatoma infestans*: Compilation of positive sites and comparison of their ecological niche with domestic population niche. *Acta Trop*. 2017; **176**: 228–35. https://doi.org/10.1016/j.actatropica.2017.08.009

Brewer, Garay M, Gorla D, *et al.*: Caracterización estadios ninfales *Triatoma infestans* Klug, 1834 (Hemiptera: Reduviidae). *Rev Soc Entomol Argent*. 1981; **40**: 91–102.

Buchfink B, Reuter K, Drost H-G: Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*. 2021; **18**(4): 366–68. https://doi.org/10.1038/s41592-021-01101-x PMID: 33828273 PMCID: PMC8026399

Bustamante Gomez M, Gonçalves Diotaiuti L, Gorla DE: Distribution of pyrethroid resistant populations of *Triatoma infestans* in the Southern Cone of South America. *PLoS Negl Trop Dis*. 2016; **10**(3): e0004561. https://doi.org/10.1371/journal.pntd.0004561

Chacón F, Bacigalupo A, Álvarez-Duhart B, *et al.*: The parasite load of *Trypanosoma cruzi* modulates feeding and defecation patterns of the Chagas disease vector *Triatoma infestans*. *Microorganisms*. 2022a; **10**(5): 1003.

Chacón F, Muñoz-San Martín C, Bacigalupo A, *et al.*: *Trypanosoma cruzi* parasite load modulates the circadian activity pattern of *Triatoma infestans*. *Insects*. 2022b; **13**(1): 76.

Challis R, Richards E, Rajan J, *et al.*: BlobToolKit – interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*. 2020; **10**(4): 1361–74. https://doi.org/10.1534/g3.119.400908 PMID: 32071071 PMCID: PMC7144090

Cheng H, Concepcion GT, Feng X, *et al.*: Haplotype-resolved *de novo* assembly using phased assembly

graphs with Hifiasm. *Nature Methods*. 2021; **18**(2): 170–75. https://doi.org/10.1038/s41592-020-01056-5 PMID: 33526886 PMCID: PMC7961889

Cheng H, Jarvis ED, Fedrigo O, *et al.*: Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology*. 2022; **40**(9): 1332–35. https://doi.org/10.1038/s41587-022-01261-x PMID: 35332338 PMCID: PMC9464699

Danecek P, Bonfield JK, Liddle J, *et al.*: Twelve years of SAMtools and BCFtools. *GigaScience*. 2021; **10**(2): https://doi.org/10.1093/gigascience/giab008 PMID: 33590861 PMCID: PMC7931819

Delgado S, Ernst KC, Pumahuanca ML, *et al.*: A country bug in the city: Urban infestation by the Chagas disease vector *Triatoma infestans* in Arequipa, Peru. *Int J Health Geogr*. 2013; **12**: 48. https://doi.org/10.1186/1476-072X-12-48

Di Iorio O, Gürtler RE: Seasonality and temperature-dependent flight dispersal of *Triatoma infestans* (Hemiptera: Reduviidae) and other vectors of Chagas disease in Western Argentina. *J Med Entomol*. 2017; **54**(5): 1285–92. https://doi.org/10.1093/jme/tjx109

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): 3047–48. https://doi.org/10.1093/bioinformatics/btw354 PMID: 27312411 PMCID: PMC5039924

Ewels PA, Peltzer A, Fillinger S, *et al.*: The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*. 2020; **38**(3): 276–78. https://doi.org/10.1038/s41587-020-0439-x PMID: 32055031

Fernández CJ, García BA: Variation in the mitochondrial genome of the Chagas disease vector *Triatoma infestans* (Hemiptera: Reduviidae). *Neotrop Entomol*. 2022; **51**(3): 483–92. https://doi.org/10.1007/s13744-022-00953-y

Formenti G, Abueg L, Brajuka A, *et al.*: Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*. 2022; **38**(17): 4214–16. https://doi.org/10.1093/bioinformatics/btac460 PMID: 35799367 PMCID: PMC9438950

Guarneri AA, Pereira MH, Diotaiuti L: Influence of the blood meal source on the development of *Triatoma infestans*, *Triatoma brasiliensis*, *Triatoma sordida*, and *Triatoma pseudomaculata* (Heteroptera, Reduviidae). *J Med Entomol*. 2000; **37**(3): 373–79. https://doi.org/10.1093/jmedent/37.3.373

Gürtler RE, Cecere MC, Fernández M del P, *et al.*: Key source habitats and potential dispersal of *Triatoma infestans* populations in Northwestern Argentina: Implications for vector control. *PLoS Negl Trop Dis*. 2014; **8**(10): e3238. https://doi.org/10.1371/journal.pntd.0003238

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; : https://doi.org/10.1101/2025.04.11.648334

Howe K, Chow W, Collins J, *et al.*: Significantly improving the quality of genome assemblies through curation. *GigaScience*. 2021; **10**(1): https://doi.org/10.1093/gigascience/giaa153 PMID: 33420778 PMCID: PMC7794651

Ihle-Soto C, Costoya E, Correa JP, *et al.*: Spatio-temporal characterization of *Trypanosoma cruzi* infection and discrete typing units infecting hosts and vectors from non-domestic foci of Chile. *PLoS Negl Trop Dis*. 2019; **13**(2): e0007170. https://doi.org/10.1371/journal.pntd.0007170

Justi SA, Dorn PL, Cahan SH: The genome of *Triatoma infestans* (Hemiptera: Reduviidae): The most important Chagas disease vector. 2020; : https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/011/037/195/GCA_011037195.1_UVM_Tinf_1.0/GCA_011037195.1_UVM_Tinf_1.0_genomic.fna.gz

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: HiGlass: Web-based visual exploration and analysis of genome interaction maps. *Genome Biology*. 2018; **19**(1): 125. https://doi.org/10.1186/s13059-018-1486-1 PMID: 6109259 PMCID: PMC6109259

Kurtzer GM, Sochat V, Bauer MW: Singularity: Scientific containers for mobility of compute. *PLOS ONE*. 2017; **12**(5): e0177459. https://doi.org/10.1371/journal.pone.0177459 PMID: 28494014 PMCID: PMC5426675

Lent H, Wygodzinsky PW: Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bull Am Mus Nat Hist*. 1979; **163**: 3.

Li H: Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018; **34**(18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191 PMID: 29750242 PMCID: PMC6137996

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. *Molecular Biology and Evolution*. 2021; **38**(10): 4647–54. https://doi.org/10.1093/molbev/msab199 PMID: 34320186 PMCID: PMC8476166

Martínez-Hernández F, Villalobos G, Montañez-Valdez OD, *et al.*: A new record of the introduced species *Triatoma infestans* (Hemiptera: Reduviidae) in Mexico. *J Med Entomol*. 2022; **59**(6): 2150–57. https://doi.org/10.1093/jme/tjac078

Merkel D: Docker: Lightweight Linux containers for consistent development and deployment. *Linux J.* 2014; **2014**(239): https://dl.acm.org/doi/10.5555/2600239.2600241

Panzera F, Ferreiro MJ, Pita S, *et al.*: Evolutionary and dispersal history of *Triatoma infestans*, main vector of Chagas disease, by chromosomal markers. *Infect Genet Evol*. 2014; **27**: 105–13. https://doi.org/10.1016/j.meegid.2014.07.006

Panzera F, Pérez R, Panzera Y, *et al.*: Cytogenetics and genome evolution in the subfamily Triatominae (Hemiptera, Reduviidae). *Cytogenet Genome Res*. 2010; **128**(1-3): 77–87. https://doi.org/10.1159/000298824

Pita S, Panzera F, Mora P, *et al.*: Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. *PLoS One*. 2017a; **12**(7): e0181635. https://doi.org/10.1371/journal.pone.0181635

Pita S, Panzera F, Vela J, *et al.*: Complete mitochondrial genome of *Triatoma infestans* (Hemiptera, Reduviidae, Triatominae), main vector of Chagas disease. *Infect Genet Evol*. 2017b; **54**: 158–63. https://doi.org/10.1016/j.meegid.2017.06.003

Ranallo-Benavidez TR, Jaron KS, Schatz MC: GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*. 2020; **11**(1): 1432. https://doi.org/10.1038/s41467-020-14998-3 PMID: 32188846 PMCID: PMC7080791

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): 1665–80. https://doi.org/10.1016/j.cell.2014.11.021 PMID: 25497547 PMCID: PMC5635824

Rhie A, McCarthy SA, Fedrigo O, *et al.*: Towards complete and error-free genome assemblies of all

vertebrate species. *Nature*. 2021; **592**(7856): 737–46. https://doi.org/10.1038/s41586-021-03451-0 PMID: 33911273 PMCID: PMC8081667

Rhie A, Walenz BP, Koren S, *et al.*: Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*. 2020; **21**(1): https://doi.org/10.1186/s13059-020-02134-9 PMID: 32928274 PMCID: PMC7488777

Rolón M, Vega MC, Román F, *et al.*: First report of colonies of sylvatic *Triatoma infestans* (Hemiptera: Reduviidae) in the Paraguayan Chaco, using a trained dog. *PLoS Negl Trop Dis*. 2011; **5**(5): e1026. https://doi.org/10.1371/journal.pntd.0001026

Schoch CL, Ciufo S, Domrachev M, *et al.*: NCBI taxonomy: A comprehensive update on curation, resources and tools. *Database*. 2020; **2020**: baaa062. https://doi.org/10.1093/database/baaa062 PMID: 32761142 PMCID: PMC7408187

Tapia-Garay V, Figueroa DP, Maldonado A, *et al.*: Assessing the risk zones of Chagas' disease in Chile, in a world marked by global climatic change. *Mem Inst Oswaldo Cruz*. 2018; **113**(1): 24–29. https://doi.org/10.1590/0074-02760170172

Torres-Pérez F, Acuña-Retamar M, Cook JA, *et al.*: Statistical phylogeography of Chagas disease vector *Triatoma infestans*: Testing biogeographic hypotheses of dispersal. *Infect Genet Evol*. 2011; **11**(1): 167–74. https://doi.org/10.1016/j.meegid.2010.09.005

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. https://doi.org/10.1186/s12859-023-05385-y PMID: 37464285 PMCID: PMC10354987

Vasimuddin Md, 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; 314–24. https://doi.org/10.1109/IPDPS.2019.00041

Waleckx E, Gourbière S, Dumonteil E: Intrusive versus domiciliated triatomines and the challenge of adapting vector control practices against Chagas disease. *Mem Inst Oswaldo Cruz*. 2015; **110**(3): 324–38. https://doi.org/10.1590/0074-02760140409

WHO: Chagas disease in Latin America: An epidemiological update based on 2010 estimates. *WER*. 2015; **6**: 33–43. https://iris.who.int/bitstream/handle/10665/242316/WER9006_33-44.PDF?sequence=1&isAllowed=y

Zhou C, McCarthy SA, Durbin R: YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics*. 2023; **39**(1): https://doi.org/10.1093/bioinformatics/btac808 PMID: 36525368 PMCID: PMC9848053