Chromosome-level genome assembly of *Triatoma dimidiata* Latreille 1811 (Hemiptera: Reduviidae: Triatominae)

Mateo Andrés Alvarado López

Universidad del Rosario
Maestría en Ciencias Naturales
Facultad de Ciencias Naturales
Bogotá, Colombia
2024

Chromosome-level genome assembly of *Triatoma dimidiata* Latreille, 1811 (Hemiptera: Reduviidae: Triatominae)

Mateo Andrés Alvarado López

Tesis presentada como requisito para obtener el título de:

Magíster en Ciencias Naturales

Director

Juan David Ramírez, Ph.D.
Profesor Titular
Facultad de Ciencias Naturales
Universidad del Rosario

Facultad de Ciencias Naturales
Maestría en Ciencias Naturales
Universidad del Rosario
Bogotá, Colombia
2024

Acknowledgements

I would like to express my deepest gratitude to Dr. Juan David Ramírez for his unwavering support and mentorship over the past four years. Without his guidance, I wouldn't be dreaming as big as I am today. I am also deeply thankful to Dr. Carolina Hernández and Nicolás Luna for their invaluable collaboration and cherished friendship, always offering words of encouragement during challenging times. My heartfelt thanks to Dr. Omar Cantillo-Barraza for providing the insects necessary for this project, as well as for the joy and laughter he brought along the way.

I extend my sincere appreciation to Dr. Martin Llewellyn and Dr. Mark Blaxter, as well as the Wellcome Sanger Institute for facilitating this collaboration, which made this project possible. Special thanks go to Universidad del Rosario for eight incredible years, during which I met some of the most amazing people and had the privilege to learn so much.

Lastly, I want to thank my mom, Claudia, my girlfriend, Silvia, my aunt, Nubia, my grandmother, Matilde, and my two cats, Mono and Popochita. They are the driving force that keeps me going, no matter how difficult things get.

Chromosome-level genome assembly of *Triatoma dimidiata* Latreille, 1811 (Hemiptera: Reduviidae: Triatominae)

Mateo Andrés Alvarado López[1], Omar Cantillo-Barraza[2], Nicolas Luna[1], Carolina Hernández[1,3], Alberto Paniz-Mondolfi[3], Juan David Ramírez[1,3]

1. Centro de Investigaciones en Microbiología y Biotecnología de la Universidad del Rosario (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario. Bogotá, Colombia.
2. Grupo de investigación en Biología y Control de Enfermedades Infecciosas (BCEI), Universidad de Antioquia. Medellín, Colombia.
3. Molecular Microbiology Laboratory, Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai. New York, NY, USA.

Abstract

Triatoma dimidiata is a primary vector of *Trypanosoma cruzi*, the parasite responsible for Chagas disease, which impacts millions worldwide. Its distribution extends from central Mexico to northern Ecuador and Peru, with marked ecological and epidemiological variations across this range, leading to taxonomic challenges and the recognition of *T. dimidiata* as a cryptic species complex. Despite efforts using morphological and molecular markers, phylogenetic analyses remain inconclusive, complicating its classification. Resolving these taxonomic issues is essential for creating effective, targeted vector control strategies. Given that traditional molecular markers have proven insufficient, genome-scale analyses offer a promising alternative. However, the lack of an annotated reference genome for *T. dimidiata* has hindered comprehensive genome-wide studies. To address this, we assembled and annotated a high-quality, chromosome-level haplotype-resolved reference genome of *T. dimidiata*, integrating PacBio High Fidelity (HiFi), Illumina DNA paired-end, Hi-C, and RNAseq reads. The resulting assemblies, with lengths of 1.31 Gb and 1.30 Gb, demonstrate high completeness and contiguity. Approximately, 95% of these sequences were organized into 10 and 12 pseudochromosomes, respectively. Additionally, repetitive content analysis revealed that more than 50% of the genome comprises repetitive elements. Using assembled transcriptomes, *ab initio* methods, and protein homology for gene prediction, we identified 338,033 potential genes. This genome resource fills a crucial gap in *T. dimidiata* genomics, paving the way for functional genomics and population studies. It will facilitate the discovery of genetic markers associated with vital traits like insecticide resistance and ecological adaptation, providing insights essential for developing targeted control measures. Our assembly establishes a foundation for further research into *T. dimidiata*'s biology, population dynamics, and vector competence, advancing global efforts to combat Chagas disease.

Resúmen (Spanish)

Triatoma dimidiata es un vector principal de *Trypanosoma cruzi*, el parásito responsable de la enfermedad de Chagas, que afecta a millones de personas en todo el mundo. Su distribución se extiende desde el centro de México hasta el norte de Ecuador y Perú, con variaciones ecológicas y epidemiológicas marcadas a lo largo de este rango, lo que conlleva a desafíos taxonómicos y al reconocimiento de *T. dimidiata* como un complejo de especies crípticas. A pesar de los esfuerzos

realizados con marcadores morfológicos y moleculares, los análisis filogenéticos siguen siendo inconclusos, dificultando su clasificación. Resolver estos problemas taxonómicos es fundamental para desarrollar estrategias de control de vectores eficaces y específicas. Dado que los marcadores moleculares tradicionales han demostrado ser insuficientes, los análisis a escala genómica ofrecen una alternativa prometedora. Sin embargo, la ausencia de un genoma de referencia anotado para *T. dimidiata* ha limitado estudios genómicos de amplio alcance. Para abordar esta necesidad, ensamblamos y anotamos un genoma de referencia de alta calidad y resolución haplotípica a nivel cromosómico para *T. dimidiata*, integrando lecturas de alta fidelidad de PacBio (HiFi), secuenciación Illumina de extremos emparejados, Hi-C y RNAseq. Los ensamblajes resultantes, con longitudes de 1,31 Gb y 1,30 Gb, demuestran alta completitud y continuidad. Aproximadamente, el 95% de estas secuencias fueron organizadas en 10 y 12 pseudo-cromosomas, respectivamente. Además, el análisis del contenido repetitivo reveló que más del 50% del genoma se compone de elementos repetitivos. Utilizando transcriptomas ensamblados, métodos *ab initio* y homología de proteínas para la predicción de genes, identificamos 338,033 genes potenciales. Este recurso genómico llena un vacío crucial en la genómica de *T. dimidiata*, allanando el camino para estudios de genómica funcional y poblacional. Facilitará el descubrimiento de marcadores genéticos asociados a rasgos importantes como la resistencia a insecticidas y la adaptación ecológica, proporcionando conocimientos esenciales para desarrollar medidas de control específicas. Nuestro ensamblaje establece una base para investigaciones adicionales sobre la biología, la dinámica poblacional y la competencia vectorial de *T. dimidiata*, avanzando en los esfuerzos globales para combatir la enfermedad de Chagas.

Keywords

Reference genome, haplotype, Triatominae, Chagas disease, annotation, PacBio, Illumina, RNAseq.

Introduction

Triatoma dimidiata (Latreille, 1811) sensu lato (s.l.) is a key vector of *Trypanosoma cruzi*, the parasite responsible for Chagas disease (CD). Currently, around 8 million people globally are infected with *T. cruzi*, while an additional 75 million are at risk (1,2). The primary route of transmission is through vector-borne means, where insects from the subfamily Triatominae transmit the parasite (3). However, there are other important transmission routes, including oral transmission (via consumption of contaminated food or drink), blood transfusions or organ transplants from infected donors, vertical transmission (from mother to child), and accidental exposure in laboratory settings (4–13).

Triatoma dimidiata is distributed throughout Latin America, from central Mexico to southern Ecuador and northern Peru, with infection rates reaching as high as 70% in some regions (14–16). While it is the primary vector for *T. cruzi* in Central America, in countries like Colombia, Ecuador, and Peru, *T. dimidiata* acts as a secondary vector, reflecting regional differences in vector ecology and transmission dynamics. Molecular studies further support these ecological variations,

suggesting that *T. dimidiata* may actually comprise a cryptic species complex, with genetic differentiation observed across its geographic range (17–21).

Triatoma dimidiata is a nocturnal insect that responds to light, odors, and the movements of potential hosts. Its attraction to artificial light, especially in females, often leads it to invade homes, posing significant challenges for vector control (22–25). During the day, it hides in sheltered areas, but at night it emerges to feed, primarily on human blood. However, *T. dimidiata* also feeds on other mammals, including domestic animals like dogs and cats, as well as cows, birds, marsupials, opossums, small rodents, bats, and other mammals such as buffalo, donkeys, sloths, and pigs (26–28).

At the molecular level, *T. dimidiata* populations show distinct patterns of domestication and gene flow. In Guatemala, highly domesticated populations with limited movement between sylvatic and domiciliary environments have been associated with poor housing conditions, which offer favorable habitats for the insects and result in low levels of gene Flow (29–31). In contrast, populations from the Yucatán Peninsula in Mexico and from Colombia exhibit greater gene flow, indicating higher mobility between sylvatic and domestic habitats (32–34).

Numerous studies have investigated the phylogenetic and taxonomic relationships within *Triatoma dimidiata*. Initially, its morphotypes were considered separate species (35,36), but due to low morphological differentiation, they were later synonymized under *T. dimidiata* (37). However, subsequent research using morphological, cytogenetic, and molecular markers uncovered multiple taxa, including at least one cryptic species (38–41). Molecular analyses with markers such as ITS-2 and mitochondrial DNA have identified distinct clades, leading to subspecies classifications. Yet, some phylogenetic studies have not supported monophyletic groupings (15,17,18,42–45).

Further research, including studies using mitochondrial markers and genotyping-by-sequencing (GBS), has divided *T. dimidiata* into distinct species, confirming the presence of cryptic species (21,46). GBS identified three monophyletic lineages, although challenges remain in achieving taxonomic resolution due to factors like homoplasy and biases in SNP selection (17,47). These findings underscore the complexity of resolving taxonomic conflicts within *T. dimidiata* and highlight the need for a high-quality genome assembly to better explore deep phylogenetic relationships and key genetic insights.

Existing research on triatomine genomics is limited, focusing mostly on cytogenetic methods for determining karyotypes and estimating genome sizes (43,44,48). To date, only two members of the Triatominae subfamily, *Rhodnius prolixus* and *Triatoma rubrofasciata*, have published complete genome assemblies (49–51). However, draft genome assemblies for other Triatominae members, such as *Triatoma infestans*, *Triatoma sanguisuga*, and *Panstrongylus geniculatus*, can be found in databases like NCBI (52–55). Karyotypic studies of *T. dimidiata* reveal that it has 20 autosomes, with an X1X2Y system in males and X1X1X2X2 in females. Flow cytometry estimates its haploid DNA content at 0.980 pg, corresponding to a genome size between 898.56 Mb and 1015.92 Mb (43). Despite this, little is known about other key genomic features, such as gene content, functional elements, repetitive sequences, and structural variants, emphasizing the need for a high-quality genome assembly.

While advancements in triatomine genomics have been limited, significant progress has been made in transcriptomic research. RNAseq studies have provided insights into various biological aspects, such as saliva and salivary gland function, metabolism, and the immune, reproductive, sensory, and detoxification systems of Chagas disease vectors (56–67). RNAseq has also been valuable for taxonomic and systematic research (59,68). A reference genome with comprehensive functional annotation would further enhance these studies by providing a detailed map of gene locations, regulatory elements, and coding sequences, facilitating deeper understanding of triatomine molecular biology (69).

Recent developments in sequencing technologies, such as PacBio HiFi and Oxford Nanopore, combined with chromatin conformation capture techniques (e.g., Hi-C, Pore-C), have significantly improved the quality of genome assemblies, even at the chromosomal level. Although this approach has only been applied to *T. rubrofasciata* among *T. cruzi* vectors, similar efforts for other insect vectors of neglected tropical diseases have resulted in high-quality reference genomes that inform control strategies (70–74). A notable example is *Aedes aegypti*, the vector for various arboviruses. Its genome was sequenced using these technologies, which enabled a genome-wide population genetic screen in the Yucatán Peninsula, identifying SNPs associated with pyrethroid resistance via changes in voltage-gated sodium channels (VGSC), a key pyrethroid target (72,73). These advances highlight the importance of creating a reference genome for *T. dimidiata* to explore vector competence, resistance mechanisms, and resolve taxonomic conflicts, which would inform species-specific and targeted vector control strategies

Generating a high-quality reference genome for *T. dimidiata* will not only help resolve these taxonomic conflicts but also provide a deeper understanding of how its genetic characteristics relate to its biology and ecology. Additionally, this genomic resource will enable the study of the genetic background of *T. dimidiata*'s vector traits, allowing for more informed and effective vector control strategies. In this study, we aim to generate a high-quality, chromosome-level annotated genome assembly for *T. dimidiata*, using a combination of PacBio long reads, Illumina short reads, and RNAseq data from colony individuals in Boyacá, eastern Colombia.

Materials and methods

Sample acquisition

On August 6, 2023, eleven adult *Triatoma dimidiata* specimens were collected from a laboratory colony maintained by the Biología y Control de Enfermedades Infecciosas (BCEI) research group at Universidad de Antioquia in Medellín, Colombia. The colony was originally established using individuals collected in Socotá, Boyacá, Colombia, at an average altitude of 2,413 meters above sea level, where the average temperature is 17°C (Supplementary Fig. 1). These insects were captured by the staff of the Boyacá Health Secretariat during active searches in peridomestic areas between October and December 2021.

After collection, the insects were sent to the BCEI laboratory, where they were identified using the taxonomic key for Triatominae (Hemiptera: Reduviidae) by Lent & Wygodzinsky (1979). They were then reared under controlled conditions in the insectary, with temperatures between 24-26°C, relative humidity of 60-70%, and a 12:12 hour photoperiod. The insects were fed on chickens (*Gallus*

gallus). The selected eleven specimens were further identified and subsequently transferred to the Centro de Investigaciones en Microbiología y Biotecnología de la Universidad del Rosario (CIMBIUR) in Bogotá, Colombia, for further processing.

These specimens were then prepared for export as part of the Wellcome Sanger Institute's Tree of Life (ToL) project. The ToL project aims to sequence, assemble, and annotate the genomes of thousands of species, contributing to a comprehensive genomic reference library to support research into biodiversity, species adaptation, and evolutionary relationships. Given *T. dimidiata*'s role as a vector of *T. cruzi*, CIMBIUR was invited to participate in the ToL project's focus on species of public health importance, aligning with its broader goals of advancing scientific knowledge for both conservation and health-related research.

To facilitate the export to the Wellcome Sanger Institute, the required permits were obtained, including an export permit from Colombia's National Environmental Licensing Authority (ANLA), granted under authorization number 3258 on May 16, 2023. Additionally, the UK Animal and Plant Health Agency issued the corresponding import permit (ITIMP23.0761). All export and import procedures adhered to the requirements of the Darwin Tree of Life project, under agreement number 255420.

Nucleic acid extraction

High molecular weight (HMW) DNA extraction, essential for generating PacBio High Fidelity (HiFi) and Hi-C reads, was conducted at the Wellcome Sanger Institute's Tree of Life Core Laboratory. The extraction followed a standard workflow, including sample preparation, homogenization, DNA extraction, fragmentation, and purification (75). The objective was to obtain long DNA fragments suitable for PacBio HiFi sequencing, known for producing highly accurate long reads. Using the Automated MagAttract v2 protocol (76), DNA was extracted and then fragmented into 12–20 kb pieces using the Megaruptor 3 system (77). The DNA was further purified with AMPure PB beads to remove shorter fragments and concentrate the high-quality, long DNA necessary for sequencing (78). This process yielded 58 Gb of HiFi reads, crucial for achieving chromosomal-level resolution in the genome assembly. Additionally, 90 Gb of Hi-C reads were produced to capture chromatin conformation, ensuring accurate scaffolding of the genome.

At Universidad del Rosario, DNA and RNA were extracted using the DNeasy Blood & Tissue Kit (Qiagen) and TRIzol (ThermoFisher Scientific), respectively, following the manufacturers' protocols. These extracted samples were prepared for Illumina sequencing, which generates high-quality reads that enhance overall accuracy. Illumina sequencing not only helps correct errors that may occur during the initial genome assembly from PacBio long reads but also ensures a more reliable and comprehensive analysis. The Illumina sequencing produced 90 Gb of data, which was used to polish and refine the PacBio-based assembly. Additionally, 12 Gb of RNAseq data were generated, playing a crucial role in functional genome annotation by linking genetic sequences to their biological functions. The quality and integrity of the nucleic acids were assessed using a NanoDrop spectrophotometer and a Qubit fluorometer before being sent to Novogene for sequencing.

Library Preparation and Sequencing

Pacific Biosciences HiFi circular consensus sequencing libraries were constructed following the manufacturer's protocols, and DNA sequencing was performed by the Wellcome Sanger Institute's Scientific Operations Core using a PacBio Revio instrument.

Hi-C data were generated using the Arima-HiC v2 kit from whole *Triatoma dimidiata* specimens. Tissue fixation in a formaldehyde-containing TC buffer crosslinked the DNA, which was then digested with a restriction enzyme master mix. The resulting 5'-overhangs were filled and labeled with biotinylated nucleotides. The biotinylated DNA was subsequently fragmented, enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit. Hi-C sequencing was carried out on an Illumina NovaSeq 6000, generating paired-end reads of 150 bp.

For the DNA and RNA samples prepared at Universidad del Rosario, sequencing was also conducted at Novogene using an Illumina NovaSeq 6000 instrument. The sequencing service handled sample preparation, library construction, and quality control, producing paired-end reads of approximately 150 bp.

Genome assembly

Before proceeding with downstream analysis, an additional round of adapter filtering was performed on the HiFi reads using HiFiAdapterFilt (79). Canonical k-mers were then calculated using 31-mers from the PacBio HiFi reads with KMC v3.2.1 (80). Next, genome profiling of *Triatoma dimidiata* was carried out using HiFi and Illumina DNA reads with GenomeScope v2.0 (81).

The initial genome assembly was conducted with Hifiasm (82) utilizing the HiFi reads along with Hi-C reads. The Hi-C Integrated Assembly option, under default parameters, was used to produce two haplotype assemblies. Haplotype duplications in each assembly were removed using purge_dups (83) (Fig. 1).

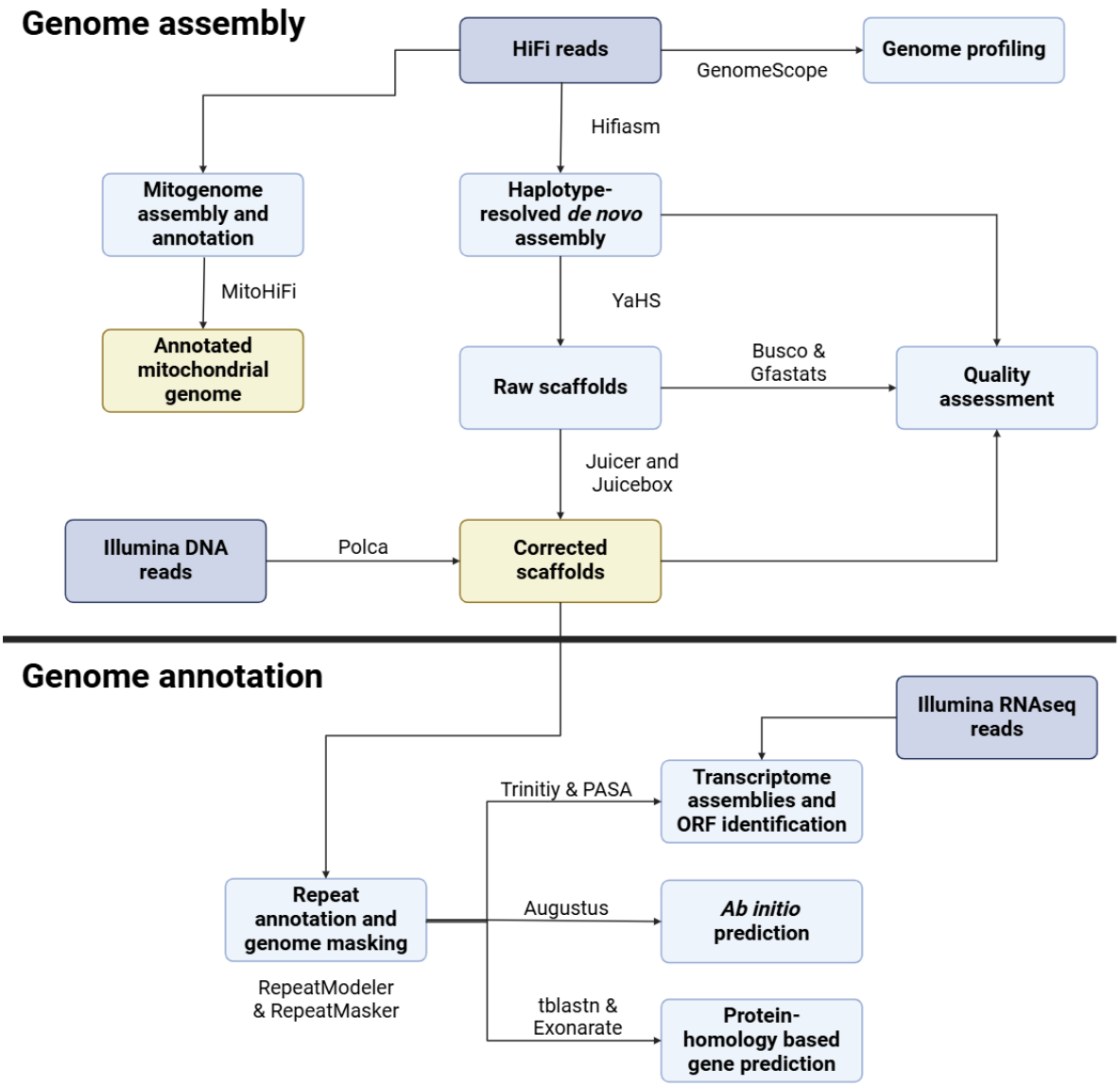


Figure 1. This graphic represents the complete workflow for genome assembly and annotation. Blue squares highlight the raw data and its specific use within the workflow. White squares denote intermediate steps, while yellow squares represent the final outputs. Arrows between the squares indicate the tools used at each stage to generate new results.

Three additional steps were completed with the Hi-C data and resulting assemblies before scaffolding. First, genome assemblies were indexed using Samtools (84) and BWA (85). Next, Hi-C reads were aligned to both assemblies using BWA-MEM (85), and Pairtools (86) was used to record valid ligation events. The resulting file was then sorted using the "sort" function from Pairtools. Finally, PCR duplicates were removed with the "dedup" function in Pairtools, and the final BAM file for scaffolding was generated using the "split" function from Pairtools, followed by sorting the file with Samtools.

Finally, each haplotype was scaffolded using YaHS (87) with the mapped Hi-C reads and polished with the paired-end DNA reads using Polca (88). The assemblies were decontaminated using FCS adaptor (89) and FCS-GX (90), and haplotype duplications were removed with purge_dups (Fig. 1).

Assembly evaluation was conducted using Gfastats (91) and BUSCO (92). The mitochondrial genome was assembled and annotated using MitoHiFi with MitoFinder (93), which utilizes previously reported mitochondrial genomes and their annotations to select the final contig and ensure sequence quality. The "findMitoReference.py" script from MitoHiFi was used to retrieve a suitable mitochondrial reference genome (Fig. 1). The final mitochondrial assembly and its annotations were visualized using Geneious Prime v.2025.0.1

To compare our mitochondrial assembly with previously reported *T. dimidiata* mitochondrial genomes, a phylogenetic reconstruction was performed. First, a BLAST search was conducted using the term "Triatoma dimidiata mitochondrial complete genome," and all available complete mitochondrial genomes were downloaded. A mitochondrial genome sequence of *Rhodnius prolixus* was included as an outgroup. The retrieved sequences were aligned using the Geneious alignment tool, and the phylogenetic tree was reconstructed using a Neighbor-Joining approach in Geneious Prime v.2025.0.1 (94).

Assembly Curation

Hi-C contact maps were generated for each assembly using Juicer (included within YaHS) and Juicer Tools, with the scaffolding output files as inputs. The contact maps were then loaded into Juicebox v2.17.00 (95), where the assemblies were manually inspected and corrected. Afterward, the corrected Hi-C maps were processed using Juicer's post-processing function, and final Hi-C contact maps were generated for both assemblies (Fig. 1).

Repeat annotation and genome masking

Haplotype two was selected for annotation, as it was less fragmented and exhibited a higher degree of completeness compared to haplotype one (see Results section). The first step in annotating the *T. dimidiata* genome assembly was to identify repetitive elements using RepeatModeler v.2.0.5 (96) and the Dfam database (Release 3.8) (97). For this analysis, partitions 0 and 8 of the Dfam database were used to configure both RepeatModeler and RepeatMasker (Fig. 1). RepeatModeler annotated the genome sequences by matching them to known genes in the Dfam database or labeling them as unknown if no matches were found. This information was then fed into RepeatMasker.

To complete the repeat annotation and masking, three rounds of RepeatMasker were run on the genome assembly, using the RepeatModeler dataset of known and unknown genes, as well as the selected Dfam partitions. The resulting masked genome assembly was subsequently used for downstream analyses.

Genome annotation

To annotate the *Triatoma dimidiata* genome, we employed a combined strategy that included gene prediction based on homology, transcriptomic evidence from pooled RNAseq data, and ab initio methods (Fig. 1). First, the clean RNAseq data was assembled using Trinity (98), applying both de novo and genome-guided approaches. PASA was then used to identify open reading frames (ORFs) from the resulting transcriptome assemblies (99).

The *ab initio* gene prediction was carried out using Augustus (100), utilizing the two assembled transcriptomes and known genes from *Triatoma rubrofasciata* (50). For protein homology-based gene prediction, tblastn (101) was used to align available protein sequences from the NCBI protein database for *Aedes aegypti*, *Drosophila melanogaster*, *Triatoma infestans*, *Rhodnius prolixus*, *Cimex lectularius*, *Halyomorpha halys*, and *Oncopeltus fasciatus*. Splice sites and exons from the aligned sequences were located using Exonerate, and genes with coding regions smaller than 150 bp were discarded (Fig. 1).

Evaluation of the Final Assembly

The genome assembly statistics were analyzed using Gfastats (91), which provided key metrics such as total assembly size, N50, and the number of contigs, offering insights into the quality and contiguity of the assembled genome. N50 is particularly important, as they reflect the length of the shortest contig required to cover 50% of the total assembly and the number of contigs contributing to this 50%, respectively. A higher N50 indicates a more contiguous assembly. Additionally, BUSCO scores were computed using the "hemiptera_odb10" and "insecta_odb10" datasets (92), evaluating the completeness of the assembly by assessing the presence of highly conserved single-copy orthologs. These scores provide a measure of how well the assembly captures the expected gene content for hemipteran insects and more broadly for the insect class. High BUSCO scores indicate a high-quality, complete genome assembly, while lower scores suggest missing or fragmented regions in the assembly.

Results

Genome size estimation

Results from GenomeScope 2.0 revealed a haploid genome size of 1,089,739,802 bp (Fig. 2). The percentage of unique 31-mers is 74.5%, suggesting that around 25% of the genome is repetitive content. Heterozygosity configurations are represented by “aa” (99.5%) and “ab” (0.516%). Mean k-mer coverage for heterozygous bases is 23.9, with a read error (“err”) of 0.409% and an average read duplication rate (“dup”) of 1.01. The k-mer length used was 31 and the inferred ploidy is 2.

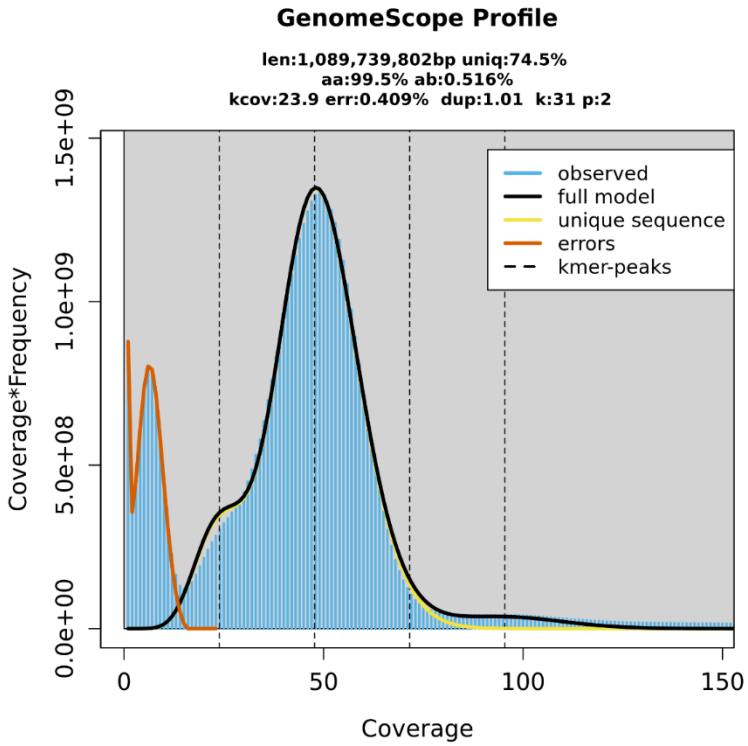


Figure 2. The genome profile generated by GenomeScope 2.0 provides a detailed insight of the genomic properties of *T. dimidiata*. On top: calculated haploid genome length ("len"), proportion of the genome comprised by non-repetitive sequences ("uniq"), heterozygosity configurations are represented by "aa" and "ab", mean k-mer coverage for heterozygous bases ("kcov"), read error ("err") and average read duplication rate ("dup"). The graph shows the observed coverage frequency and the modeled distribution of unique sequences, errors, and k-mer peaks.

Genome assembly

The initial assembly generated with Hifiasm, integrating Hi-C reads, produced two assemblies—one for each haplotype. Haplotype 1 consisted of 27,575 contigs, with an N50 of 1.30 Mb, a maximum contig length of 44 Mb, and a total genome size of 1.95 Gb. Haplotype 2 consisted of 5,623 contigs, with an N50 of 5.56 Mb, a maximum length of 48 Mb, and a total size of 1.48 Gb (Table 1). Following the removal of haplotypic duplications, haplotype 1 underwent a 12.43% reduction in size, decreasing from 1.95 Gb to 1.72 Gb, while the N50 dropped to 2.98 Mb, with the maximum contig length remaining similar to the initial assembly. In contrast, haplotype 2 remained unchanged, as no haplotypic duplications were identified for removal (Table 2).

Both assemblies were evaluated using BUSCO, revealing a high degree of completeness. Haplotype 1 exhibited 93.6% complete BUSCOs, while haplotype 2 showed 99% completeness, with both assemblies containing less than 5% missing BUSCOs (Table 3). These metrics indicate the strong quality and robustness of both assemblies.

Table 1. Assembly statistics for each haplotype after initial assembly. All values shown correspond to base pairs.

Value	Haplotype 1	Haplotype 2
N50	1,309,957	5,567,295
N60	93,425	3,429,777
N70	55,760	1,942,275
N80	36,772	240,038
N90	23,712	72,903
N100	3,581	4,659
Gaps	0	0
N count	0	0
Largest	44,042,886	48,133,479
N	27,575	5,623
Average	70,907.57	263,834.00
Sum	**1,955,276,355**	**1,483,538,600**

Table 2. Assembly statistics for each haplotype after removal of haplotypic duplications. All values shown correspond to base pairs.

Value	Haplotype 1	Haplotype 2
N50	2,980,894	5,567,295
N60	483,939	3,429,777
N70	79,223	1,942,275
N80	49,361	240,038
N90	30,245	72,903
N100	3,581	4,659
Gaps	0	0
N count	0	0
Largest	44,042,886	48,133,479
N	18,172	5,623
Average	95,059.62	263,834.00
Sum	**1,727,423,351**	**1,483,538,600**

Table 3. Results of quality assessment of haplotype assemblies with BUSCO. Each haplotype has the percentage, and the corresponding number of genes found for each database.

Value	Hemiptera				Insecta			
	Haplotype 1		Haplotype 2		Haplotype 1		Haplotype 2	
Complete	93.1%	2337	99.0%	2485	94.3%	1289	98.8%	1350
Single copy	89.5%	2246	96.8%	2430	81.6%	1115	95.0%	1298
Duplicated	3.6%	91	2.2%	55	12.7%	174	3.8%	52
Fragmented	1.1%	27	0.6%	14	1.5%	21	0.8%	11
Missing	5.8%	146	0.4%	11	4.2%	57	0.4%	6
Total	100%	2510	100%	2510	100%	1367	100%	1367

After scaffolding the contigs of each haplotype assembly into pseudochromosomes, we obtained two assemblies with significantly improved contiguity. Over 90% of both assemblies were successfully anchored to 10 pseudochromosomes. However, for haplotype 2, two additional pseudochromosomal molecules were identified, potentially corresponding to sex chromosomes, which contributed to the observed difference in genome size between haplotypes.

For haplotype 1, the final assembly consisted of 17,727 contigs, with an N50 of 88 Mb, a maximum contig length of 148 Mb, and a total genome size of 1.72 Gb. In contrast, haplotype 2 had 5,259 contigs, with an N50 of 92 Mb, a maximum length of 149 Mb, and a total genome size of 1.48 Gb. After manual curation, genome sizes were adjusted to 1.08 Gb for haplotype 1 and 1.17 Gb for haplotype 2, addressing previously undetected haplotypic duplications, resulting in 1,892 and 1,348 scaffolds, respectively. The final BUSCO scores remained consistent with initial evaluations, confirming the completeness of both assemblies.

A Hi-C contact map was generated for both haplotypes (Fig. 3), and the length of each chromosome was calculated (Table 4). Final assembly statistics for both haplotypes are provided in Table 5, and comparisons to other available assemblies from closely related Triatominae species are provided in Table 6, showcasing significant improvements in contiguity, completeness, and resolution.

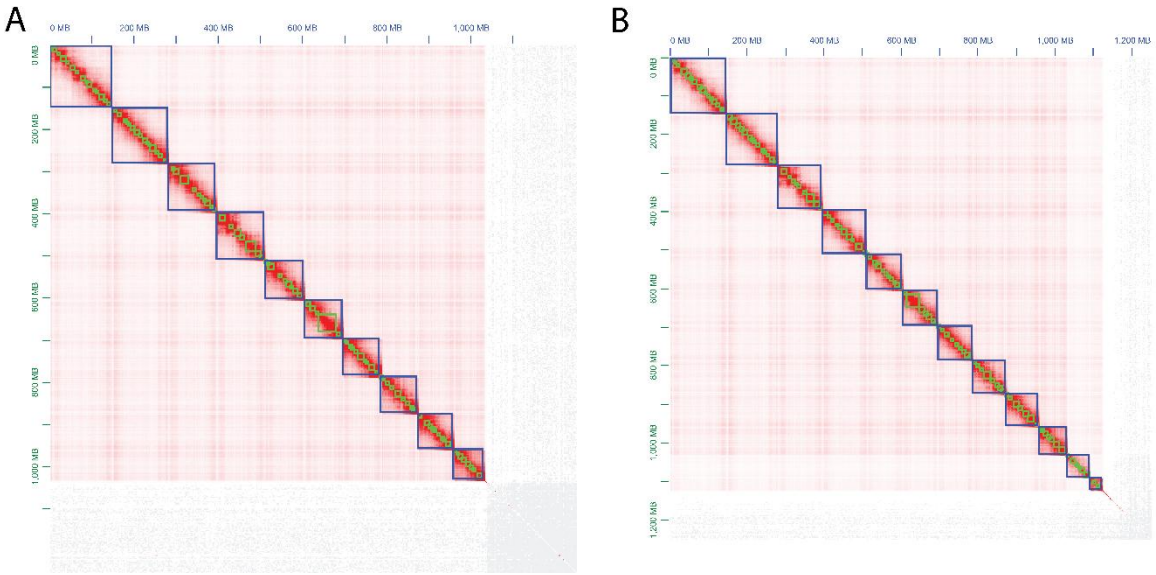


Figure 3. Genome-wide Hi-C heatmap of the haplotype assemblies of *Triatoma dimidiata*. A) Genome-wide Hi-C heatmap of haplotype 1 of *T. dimidiata.* B) Genome-wide Hi-C heatmap of haplotype 2 of *T. dimidiata.* The blue squares represent chromosomes and the small green squares inside the blue squares represent contigs that make up the chromosome.

Table 4. Chromosome lengths for *Triatoma dimidiata* across haplotype 1 and haplotype 2. The values represent the length of each chromosome in base pairs (bp). Chromosomes are listed from 1 to 12, with values provided for both haplotypes where available. Missing data is indicated as "NA" where haplotype information is not available.

Chromosome	Haplotype 1	Haplotype 2
1	147,752,925	147,380,600
2	134,095,334	134,689,094
3	115,015,212	114,207,885
4	113,614,872	113,811,786
5	93,656,895	93,830,915
6	92,806,257	92,212,520
7	90,314,915	89,893,439
8	87,088,030	87,548,738
9	85,284,227	85,019,853
10	72,722,340	73,894,870
11	NA	59,359,128
12	NA	31,304,854
TOTAL	**1,032,351,007**	**1,123,153,682**

Table 5. final assembly statistics for both haplotypes of *T. dimidiata*, including metrics such as N50, N60, N70, N80, N90, N100 (which represent the contig or scaffold lengths containing respective percentages of the genome), gaps, N count (unknown bases), largest contig or scaffold, total number of contigs, average contig length, and the sum of all contig lengths

Value	Haplotype 1	Haplotype 2
N50	92,806,257	92,212,520
N60	90,314,915	87,548,738
N70	85,284,227	85,019,853
N80	242,257	59,359,128
N90	88,554	211,282
N100	59,827	57,237
Gaps	49,900	46,100
N count	499	461
Largest	147,752,925	147,380,600
N	2,961	1,348
Average	441,393.61	977,651.81
Sum	**1,306,966,488**	**1,317,874,641**

Table 6. Comparison of genome assembly statistics for *Triatoma dimidiata* (haplotype 2) with other key vector species, including *Rhodnius prolixus* (versions RproC1 and 3.0.3), *Triatoma rubrofasciata*, *Triatoma sanguisuga, Triatoma infestans,* and *Panstrongylus geniculatus*. The table presents the number of scaffolds, N50 values (in megabases), and total genome lengths (in gigabases or megabases). N50 is a metric used to assess genome assembly contiguity, with higher values indicating better assembly quality. The comparison highlights the relatively low fragmentation and greater contiguity of the *T. dimidiata* haplotype 2 assembly compared to other insect vectors.

Statistic	*T. dimidiata (haplotype 2)*	*R. prolixus* RproC1 (51)	*R. prolixus 3.0.3* (49)	*T. rubrofasciata* (50)	*T. sanguisuga* (52)	*T. infestans*	*P. geniculatus*
# Scaffolds	1,348	27,872	17,755	1,303	Not scaffolded	14,951	1,308
N50	92 Mb	1.1 Mb	47 Mb	50.7 Mb	95 Mb	108,8 Kb	112.1 Mb
Total length	1,17 Gb	702.6 Mb	706.8 Mb	680 Mb	1.2 Gb	1.3 Gb	1.4 Gb

MitoHiFi's *findMitoReference.py* script identified mitochondrial genome accession number AF301594.1, assembled from a *T. dimidiata* specimen collected in Guatemala, as the best available reference (102). Using this reference, MitoHiFi produced a final circular mitochondrial assembly of 17,103 bp, containing 36 annotated genes, including 13 protein-coding genes, two rRNA genes, and 22 tRNA genes (Fig. 4A).

Our phylogenetic reconstruction, which included previously assembled *T. dimidiata* mitogenomes, showed no clear differentiation between our sequence and other *T. dimidiata* sequences from Central America. This suggests a high degree of mitochondrial genome conservation with minimal differentiation across individuals from different geographical regions (Figure 4B).

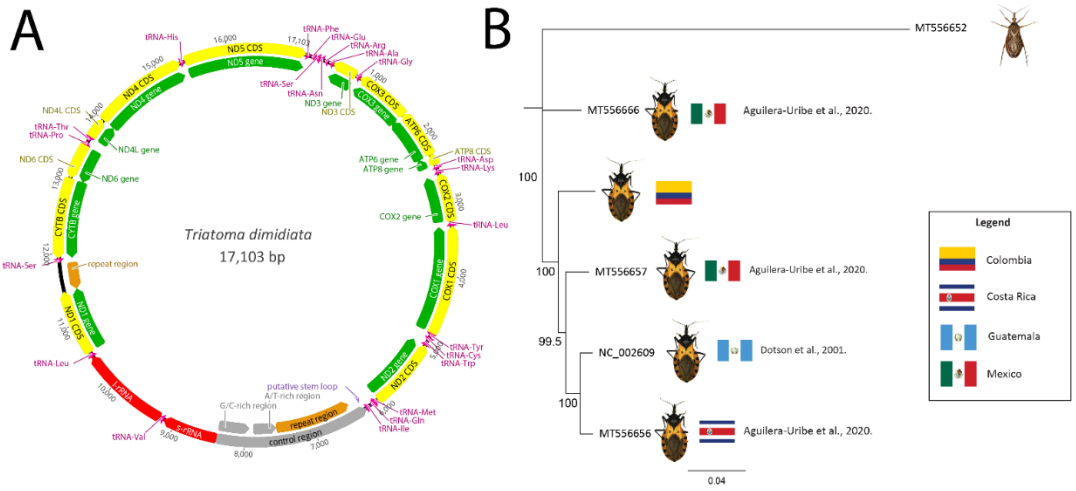


Figure 4. A) Annotation of the mitochondrial genome of *Triatoma dimidiata* produced by MitoHiFi. Final assembly reported a genome size of 17,388 bp and 36 genes annotated, arrows indicate the direction in which the gene codes. Size of the purple squares represent the size of the gene in the mitogenome. **B)** Phylogenetic reconstruction of the mitochondrial genomes of *Triatoma dimidiata* using a Neighbor-Joining approach in Geneious Prime. The tree includes samples from Mexico, Colombia, Costa Rica, and Guatemala, with *Rhodnius prolixus* as the outgroup. Branches are labeled

with accession numbers and bootstrap values indicating support for each grouping. Flags denote the geographic origin of the samples, and the data is sourced from Aguilera-Uribe et al. (2020) and Dotson et al. (2001).

Repeat-content identification and annotation

The repeat analysis of the *Triatoma dimidiata* genome assembly revealed that approximately 50.46% of the 1.17 Gb genome consists of repetitive elements. Retroelements occupy 17.55% of the genome, with long interspersed nuclear elements (LINEs) accounting for the largest portion (12.90%). Within LINEs, the RTE/Bov-B family is the most abundant (2.80%), followed by R1/LOA/Jockey (1.90%) and L2/CR1/Rex (1.59%). Short interspersed nuclear elements (SINEs) make up a smaller fraction (1.73%), with Penelope elements contributing 0.02%. LTR retroelements, including BEL/Pao (1.20%) and Gypsy/DIRS1 (1.34%), constitute 2.93% of the genome (Supplementary Table 1).

DNA transposons represent 5.57% of the genome, with the most abundant families being Tc1-IS630-Pogo (2.98%) and hobo-Activator (1.43%). Other families, such as PiggyBac (0.34%) and Tourist/Harbinger (0.12%), make smaller contributions. Rolling-circle elements comprise 13.51%, and unclassified elements account for 10.76% of the genome, reflecting a large portion of repetitive sequences that remain poorly characterized. Small RNA sequences and simple repeats add 1.36% and 2.30%, respectively. Overall, interspersed repeats account for 33.90% of the genome, underscoring the diversity and complexity of repetitive sequences in *T. dimidiata* (Supplementary Table 1).

The gene prediction analysis for *T. dimidiata* yielded the following results: For ORF identification, the genome-guided (GG) transcriptome identified 734,140 ORFs, corresponding to 34,694 genes and mRNA sequences, 85,079 five prime UTR regions, 287,523 exons, 183,755 coding sequences (CDS), and 73,701 three prime UTR regions. The de novo (DN) transcriptome identified 621,564 ORFs, corresponding to 28,242 genes and mRNA sequences, 75,381 five prime UTR regions, 244,927 exons, 150,596 CDS, and 65,934 three prime UTR regions. Ab initio predictions from Augustus identified 87,719 genes, including 20,344 genes and mRNA sequences, 23,050 CDS, 9,920 start codons, 8,480 stop codons, and 5,581 introns. Additionally, protein-homology methods resulted in 575,777 protein hits. These combined results provide a comprehensive overview of gene prediction for *T. dimidiata* using multiple approaches (Table 6).

Table 7. This table compares the results from three gene prediction approaches: ORF identification (genome-guided and de novo), ab initio prediction using Augustus, and protein-homology methods. The Genome-Guided, ORF Identification and De Novo ORF Identification columns show the number of ORFs, genes, mRNA sequences, exons, coding sequences (CDS), and untranslated regions (five prime UTR and three prime UTR) predicted from transcriptome assemblies. The Augustus column presents ab initio predictions, including the number of genes, mRNA sequences, CDS, start codons, stop codons, and introns identified. The Protein-Homology column shows the total number of protein hits identified via homology searches. This comparative table highlights the outputs of each method in terms of genome feature prediction.

Feature	Genome-Guided ORF Identification	De Novo ORF identification	Augustus	Protein-Homology
ORFs	734,140	621,564	NA	NA
Genes	34,694	28,242	20,344	NA
mRNA	34,694	28,242	20,344	NA
Exons	287,523	244,927	NA	NA
CDS	183,755	150,596	23,050	NA
Five prime UTR	85,079	75,381	NA	NA
Three prime UTR	73,701	65,934	NA	NA
Start codon	NA	NA	9,920	NA
Stop codon	NA	NA	8,480	NA
Introns	NA	NA	5,581	NA
Protein hits	NA	NA	NA	575,777

Discussion

A reliable chromosome-level genome assembly is essential for advancing genomic research across species. In this study, we produced a high-quality, haplotype-resolved de novo genome assembly for *Triatoma dimidiata* using individuals from a laboratory colony originating from Boyacá, Colombia (Supplementary Fig. 1). This assembly was generated through a combination of PacBio High Fidelity (HiFi) reads, Hi-C technology, and Illumina DNA sequencing, enabling a comprehensive and accurate representation of the genome (Fig. 1). The total assembly lengths assigned to chromosomes—1.02 Gb for haplotype 1 and 1.12 Gb for haplotype 2 (Table 4)—align with our estimated genome size of 1.08 Gb (Fig. 2), consistent with previously reported flow cytometry estimates (43,44) (Table 5). The assembly demonstrates high completeness, confirmed by BUSCO analysis, and strong contiguity, with N50 values of 92 Mb for both haplotypes (Table 3). Additionally, the number of chromosomes identified in each assembly aligns with previously reported values for *T. dimidiata* (Table 4; Fig.4) (43,44,48). This assembly fills a crucial gap in genomic resources for *T. dimidiata*, which, to the best of our knowledge, lacked a plausible reference genome until now.

Before this study, genome assemblies for Chagas disease vectors were available only for *Rhodnius prolixus* and *Triatoma rubrofasciata* (49–51), with unpublished assemblies for *Triatoma infestans*, *Triatoma sanguisuga*, and *Panstrongylus geniculatus* listed in the NCBI database (53–55). While these assemblies provided valuable insights into *T. cruzi* vectors, the genome of *T. dimidiata* remained unresolved. Our assembly shows significantly less fragmentation and greater contiguity compared to those of *R. prolixus* and *T. infestans*, improving upon the contiguity of the *T.*

rubrofasciata assembly (Table 6). It also exhibits similar statistics to the assemblies of *P. geniculatus* and *T. sanguisuga*, which were generated using similar sequencing technologies (Table 6). This highlights the advantages of combining long-read sequencing with chromatin capture data to produce more contiguous genome assemblies. Generating a highly contiguous reference genome with minimal fragmentation is critical for downstream applications such as population genetic variation studies, molecular marker development, and comparative genomics (103,104). The consistency of our assembly with previously reported genome sizes and chromosome numbers supports its accuracy and completeness, minimizing the risk of structural errors (105). This is especially significant for non-model organisms like *T. dimidiata*, where genomic resources are limited. Our assembly will enable more reliable research on evolutionary biology, species adaptation, and public health (106–110).

To date, five mitochondrial genome assemblies have been reported for *T. dimidiata*, all describing similar mitochondrial structures (Fig. 4). These assemblies consistently identify 13 protein-coding genes, 22 tRNAs, and ribosomal subunits, but the reported lengths vary. Dotson & Beard (2001) reported a length of 17,019 bp, while Aguilera-Uribe et al. (2020) and Gunter et al. (2024) reported lengths between 15,440 bp and 16,077 bp (102,111,112). Our assembly presents the longest mitochondrial genome to date at 17,103 bp (Fig. 4A), likely due to the use of long-read sequencing, which can more accurately resolve repetitive regions (113–115). This more complete mitochondrial genome will provide valuable insights into evolutionary relationships and population genetics, as well as support functional studies of mitochondrial genes (116).

Our phylogenetic reconstruction shows that the mitochondrial sequences of *T. dimidiata* individuals collected from various geographic locations lack the level of mitochondrial differentiation needed to make population-level inferences (Fig. 4B). This finding aligns with previous studies that used different mitochondrial markers to address taxonomic uncertainties in *T. dimidiata*, where consistent differentiation patterns to define subspecies could not be identified (21,117). However, as more mitochondrial genomes become available and high-throughput sequencing becomes increasingly accessible, future population genetic analyses at the whole-mitochondrial scale could offer a more robust approach to resolving these taxonomic conflicts.

Our repeat content analysis reveals that 50.46% of the *T. dimidiata* genome consists of repetitive sequences, a pattern consistent with other triatomine species. For example, *T. infestans* has about 40% repetitive DNA, while *T. delpontei* shows 59.64% (118,119). The repeatome of *T. dimidiata* is dominated by retroelements (17.55%), with LINEs being the most prevalent (12.90%). DNA transposons, including the hobo-Activator and Tc1-IS630-Pogo elements, account for 5.57%, which are often associated with developmental processes in other insects (Supplementary Table 1) (120). Notably, *T. dimidiata* has significantly less satellite DNA (only 1,465 bp; Supplementary Table 1) compared to *T. delpontei*, where satellite DNA comprises over 48% of the genome. This divergence may be attributed to species-specific amplification, deletion, and variation in repetitive sequences, as observed in other species like *Drosophila* (121–123). Understanding repetitive elements is vital for deciphering genome size, structure, and stability. These elements influence chromosomal architecture, gene expression, and evolutionary dynamics. In *T. dimidiata*, repetitive sequences may also play a role in traits like insecticide resistance and ecological adaptation, both critical for effective vector control strategies (124–127). Such information paves the way for new research focused on exploring the functional roles of repetitive elements across various biological aspects of *T. dimidiata*, potentially uncovering their influence on vectorial capacity, adaptation to environmental pressures, and interactions with host and pathogen.

Preliminary annotation results revealed a high number of coding sequences (CDS) in the transcriptomes, although *ab initio* predictions indicated fewer CDS (Table 7). Both methods identified similar numbers of genes and mRNAs, suggesting that the discrepancy in CDS counts is due to differences in approach. ORF identification focuses on coding regions, while *ab initio* prediction incorporates annotated genes from closely related species like *T. rubrofasciata* (Table 7) (99,100). Protein homology results further reflect the genome's completeness, as many sequences shared orthologs with other epidemiologically relevant species (Table 7). We estimate the final annotation will include between 12,000 and 16,000 protein-coding genes, consistent with related species (50,51). This comprehensive annotation will advance understanding of *T. dimidiata* biology, providing essential insights into gene functions, regulatory networks, and metabolic pathways. By identifying key coding sequences, it opens new avenues for research into specific biological mechanisms, such as pathways linked to insecticide resistance, which are crucial for devising effective vector control strategies and mitigating resistance development (128).

While this genome assembly establishes a strong foundation for future research on *T. dimidiata*, annotating complex genomes like this one remains challenging. Successful annotation—accurately identifying genes, regulatory elements, and other functional regions—requires significant computational power and access to high-quality, tissue-specific RNA sequencing data. For research groups with limited funding and computational resources, these demands are especially daunting, as advanced sequencing platforms and the computational capabilities needed to process large datasets are often cost-prohibitive, creating barriers to comprehensive, high-quality annotations (129–132). Additionally, one limitation of this study is that DNA and RNA were extracted from different individuals, which can introduce variability and potentially impact the coherence of the genome assembly, complicating the interpretation of sequence data across samples. To address this, we used individuals from a laboratory colony of *T. dimidiata* originating from Boyacá, Colombia, to provide a more genetically consistent sample pool, thereby reducing genetic variability and supporting a more coherent assembly. Ultimately, despite these limitations, this genome assembly represents an essential step forward, offering a foundational resource that can drive new discoveries and promote more effective strategies for addressing the challenges associated with *T. dimidiata* as a vector of *T. cruzi*.

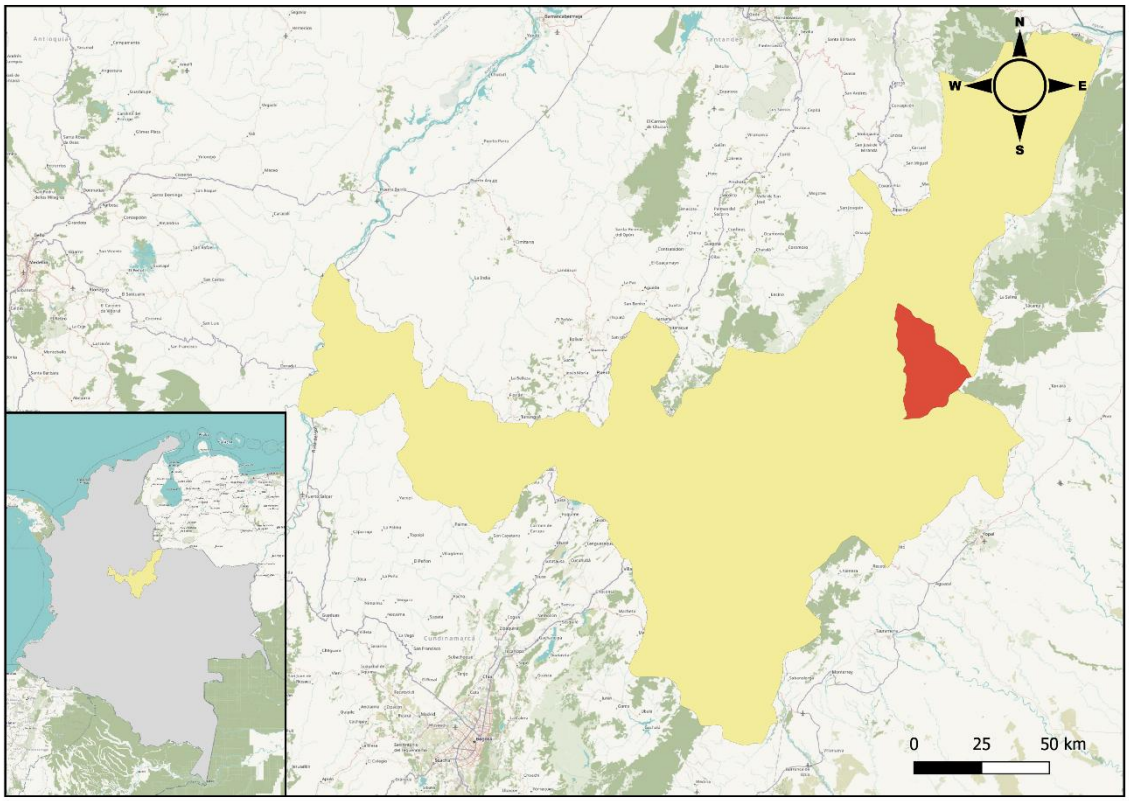
Despite these challenges, our study highlights the critical role of collaboration and advanced sequencing technologies in global efforts to control Chagas disease. This genome assembly serves as a foundational tool for future research, enabling studies that can offer insights into various ecological and biological traits of *T. dimidiata*. Through such studies, researchers may better understand the vector's adaptability to selective pressures, like insecticides and abrupt environmental changes (71,72), as well as uncover genomic factors that drive its blood-feeding behavior and long mobility ranges (133). Additionally, this resource supports efforts to resolve longstanding taxonomic uncertainties surrounding the species and to carry out population genomic studies (46). From an epidemiological perspective, the genome assembly provides a valuable basis for establishing genomic surveillance networks in regions where *T. dimidiata* plays a primary role in *T. cruzi* transmission. Such networks would allow public health agencies to develop proactive interventions informed by genomic epidemiology, helping to prevent parasite transmission before it impacts communities (134). Ultimately, this study lays a strong foundation for research and public health strategies that address both the biological complexity and public health significance of *T. dimidiata*.

In conclusion, we have produced a high-quality annotated genome assembly for *Triatoma dimidiata*, a key vector of Chagas disease. The final assembly consists of 10 pseudochromosomes corresponding to the organism's 10 autosomal chromosomes (Fig. 3; Table 4). Over 50% of the genome is composed of repetitive elements (Supplementary Table 1), and we have provided a robust estimate of coding sequences, informed by data from closely related species (Table 7). This high-quality assembly will serve as an invaluable resource for future research, with the ultimate goal of improving vector control measures and halting the transmission of *T. cruzi*.

The high-quality, chromosome-level annotated genome assembly of *Triatoma dimidiata* presented in this study marks a significant advancement in genomic resources for this important vector of *Trypanosoma cruzi*, the causative agent of Chagas disease. By generating haplotype-resolved assemblies with high contiguity and completeness, this study addresses a critical gap in genomic data, as *T. dimidiata* previously lacked a reliable reference genome. The assembly's consistency with previous estimates of genome size and chromosome number further supports its accuracy, enhancing its value for downstream applications such as comparative genomics, evolutionary biology, and vector control studies.

This assembly paves the way for various functional studies, including the identification of genetic markers related to insecticide resistance, behavior, and ecological adaptation. These insights are essential for developing targeted vector control strategies and for exploring the genetic basis of key traits that influence *T. dimidiata*'s role as a vector, as has been done with other species (71,72). Additionally, the repeat content analysis reveals significant differences in genome architecture across species, highlighting the dynamic nature of repetitive elements and their role in genome evolution (135).

Supplementary material



Supplementary figure 1. This map shows the municipality of Socotá, Boyacá, Colombia, highlighted in red as the capture zone for *Triatoma dimidiata* individuals. These insects were collected between October and December 2021 by the Secretaría de Salud de Boyacá during an active search in peridomestic areas. The captured individuals were subsequently used to establish colonies maintained by the Biología y Control de Enfermedades Infecciosas (BCEI) research group at the Universidad de Antioquia in Medellín, Colombia.

Supplementary table 1. Composition of repetitive elements within the genome of *Triatoma dimidiata*, categorized by element type, number of elements, total length (in base pairs), and percentage of the genome they occupy.

Element class			Number of elements	Length occupied	Percentage of sequence
Retroelements			782911	221351575 bp	17.55 %
	SINEs		165972	21787451 bp	1.73 %
	Penelope		1468	262578 bp	0.02 %
	LINEs		558684	162667246 bp	12.90 %
		CRE/SLACS	0	0 bp	0.00 %
		L2/CR1/Rex	57420	19997969 bp	1.59 %
		R1/LOA/Jockey	80335	23961266 bp	1.90 %

		R2/R4/NeSL	6868	2300437 bp	0.18 %
		RTE/Bov-B	112144	35312179 bp	2.80 %
		L1/CIN4	22	1233 bp	0.00 %
	LTR elements		58255	36896878 bp	2.93 %
		BEL/Pao	14555	15126810 bp	1.20 %
		Ty1/Copia	12810	3688232 bp	0.29 %
		Gypsy/DIRS1	28459	16900514 bp	1.34 %
		Retroviral	2431	1181322 bp	0.09 %
DNA transposons			393865	70219838 bp	5.57 %
	Hobo-activator		72089	18015025 bp	1.43 %
	Tc1-IS630-Pogo		251706	37527873 bp	2.98 %
	En-Spm		0	0	0.00 %
	MULE-MuDR		8	568 bp	0.00 %
	PiggyBac		38909	4334543 bp	0.34 %
	Tourist/Harbinger		9306	1526444 bp	0.12 %
	Other (Mirage, P-element, Transib)		47	3099 bp	0.00 %
Rolling-circles			745653	170352547 bp	13.51 %
Unclassified			816888	135722523 bp	10.76 %
Total interspersed repeats				427556514 bp	33.90 %
Small RNA			138734	17090523 bp	1.36 %
Satellites			24	1465 bp	0.00 %
Simple repeats			721502	28984167 bp	2.30 %
Low complexity			117937	5577522 bp	0.44 %

References

1. PAHO. Chagas disease - PAHO/WHO | Pan American Health Organization [Internet]. 2024 [cited 2024 Oct 8]. Available from: https://www.paho.org/en/topics/chagas-disease

2. WHO. Chagas disease in Latin America: an epidemiological update based on 2010 estimates. Wkly Epidemiol Rec. 2015;90(06):33–44.

3. Cucunubá ZM, Gutiérrez-Romero SA, Ramírez JD, Velásquez-Ortiz N, Ceccarelli S, Parra-Henao G, et al. The epidemiology of Chagas disease in the Americas. Lancet Reg Health – Am [Internet]. 2024 Sep 1 [cited 2024 Oct 21];37. Available from: https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(24)00208-4/fulltext

4. Benjamin RJ, Stramer SL, Leiby DA, Dodd RY, Fearon M, Castro E. Trypanosoma cruzi infection in North America and Spain: evidence in support of transfusion transmission (CME). Transfusion (Paris). 2012;52(9):1913–21.

5. Bern C. Chagas' Disease. Longo DL, editor. N Engl J Med. 2015 Jul 30;373(5):456–66.

6. Coura JR. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions - A comprehensive review. Mem Inst Oswaldo Cruz. 2014 Dec 2;110:277–82.

7. Howard E, Xiong X, Carlier Y, Sosa-Estani S, Buekens P. Frequency of the congenital transmission of rypanosoma cruzi: a systematic review and meta-analysis. BJOG Int J Obstet Gynaecol. 2014;121(1):22–33.

8. Huprikar S, Bosserman E, Patel G, Moore A, Pinney S, Anyanwu A, et al. Donor-Derived Trypanosoma cruzi Infection in Solid Organ Recipients in the United States, 2001–2011. Am J Transplant. 2013;13(9):2418–25.

9. Klein MD, Proaño A, Noazin S, Sciaudone M, Gilman RH, Bowman NM. Risk factors for vertical transmission of Chagas disease: A systematic review and meta-analysis. Int J Infect Dis. 2021 Apr 1;105:357–73.

10. Lynn MK, Rodriguez Aquino MS, Cornejo Rivas PM, Kanyangarara M, Self SCW, Campbell BA, et al. Chagas Disease Maternal Seroprevalence and Maternal–Fetal Health Outcomes in a Parturition Cohort in Western El Salvador. Trop Med Infect Dis. 2023 Apr;8(4):233.

11. Rassi A, Marin-Neto JA. Chagas disease. The Lancet. 2010 Apr 17;375(9723):1388–402.

12. Robertson LJ, Havelaar AH, Keddy KH, Devleesschauwer B, Sripa B, Torgerson PR. The importance of estimating the burden of disease from foodborne transmission of Trypanosoma cruzi. PLoS Negl Trop Dis. 2024 Feb 8;18(2):e0011898.

13. Santana KH, Oliveira LGR, Barros de Castro D, Pereira M. Epidemiology of Chagas disease in pregnant women and congenital transmission of Trypanosoma cruzi in the Americas: systematic review and meta-analysis. Trop Med Int Health. 2020;25(7):752–63.

14. Cantillo-Barraza O, Medina M, Zuluaga S, Valverde C, Motta C, Ladino A, et al. Eco-epidemiological study reveals the importance of Triatoma dimidiata in the Trypanosoma cruzi transmission, in a municipality certified without transmission by Rhodnius prolixus in Colombia. Acta Trop. 2020 Sep 1;209:105550.

15. Dorn PL, Monroy C, Curtis A. Triatoma dimidiata (Latreille, 1811): A review of its diversity across its geographic range and the relationship among populations. Infect Genet Evol. 2007;7(2):343–52.

16. Velásquez-Ortiz N, Hernández C, Cantillo-Barraza O, Ballesteros N, Cruz-Saavedra L, Herrera G, et al. Trypanosoma cruzi Parasite Burdens of Several Triatomine Species in Colombia. Trop Med Infect Dis [Internet]. 2022 Dec [cited 2023 May 22];7(12). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9782637/

17. Bargues MD, Klisiowicz DR, Gonzalez-Candelas F, Ramsey JM, Monroy C, Ponce C, et al. Phylogeography and Genetic Variation of Triatoma dimidiata, the Main Chagas Disease Vector in Central America, and Its Position within the Genus Triatoma. PLoS Negl Trop Dis. 2008 May 7;2(5):e233.

18. Dorn PL, Calderon C, Melgar S, Moguel B, Solorzano E, Dumonteil E, et al. Two Distinct Triatoma dimidiata (Latreille, 1811) Taxa Are Found in Sympatry in Guatemala and Mexico. PLoS Negl Trop Dis. 2009 Mar 10;3(3):e393.

19. Gómez-Palacio A, Arboleda S, Dumonteil E, Townsend Peterson A. Ecological niche and geographic distribution of the Chagas disease vector, Triatoma dimidiata (Reduviidae: Triatominae): Evidence for niche differentiation among cryptic species. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2015 Dec;36:15–22.

20. Herrera-Aguilar M, Be-Barragán LA, Ramirez-Sierra MJ, Tripet F, Dorn P, Dumonteil E. Identification of a large hybrid zone between sympatric sibling species of *Triatoma dimidiata* in the Yucatan peninsula, Mexico, and its epidemiological importance. Infect Genet Evol. 2009 Dec 1;9(6):1345–51.

21. Monteiro FA, Peretolchina T, Lazoski C, Harris K, Dotson EM, Abad-Franch F, et al. Phylogeographic Pattern and Extensive Mitochondrial DNA Divergence Disclose a Species Complex within the Chagas Disease Vector Triatoma dimidiata. PLOS ONE. 2013 ago;8(8):e70974.

22. Gourbière S, Dumonteil E, Rabinovich JE, Minkoue R, Menu F. Demographic and Dispersal Constraints for Domestic Infestation by Non-Domiciliated Chagas Disease Vectors in the Yucatan Peninsula, Mexico. Am J Trop Med Hyg. 2008 Jan;78(1):133–9.

23. Polonio R, Ramirez-Sierra MJ, Dumonteil E. Dynamics and Distribution of House Infestation by *Triatoma dimidiata* in Central and Southern Belize. Vector-Borne Zoonotic Dis. 2009 Feb;9(1):19–24.

24. Ramirez-Sierra MJ, Herrera-Aguilar M, Gourbière S, Dumonteil E. Patterns of house infestation dynamics by non-domiciliated Triatoma dimidiata reveal a spatial gradient of infestation in rural villages and potential insect manipulation by Trypanosoma cruzi. Trop Med Int Health. 2010;15(1):77–86.

25. Rebollar-Téllez EA, Reyes-Villanueva F, Escobedo-Ortegón J, Balam-Briceño P, May-Concha I. Abundance and nightly activity behavior of a sylvan population of Triatoma dimidiata (Hemiptera: Reduviidae: Triatominae) from the Yucatan, México. J Vector Ecol. 2009;34(2):304–10.

26. Farfán-García AE, Angulo-Silva VM. Triatoma dimidiata populations' (Hemiptera: Reduviidae: Triatominae) feeding behaviour in an endemic zone and related epidemiological implications. Rev Salud Pública. 2011 Feb;13(1):163–72.

27. Torres-Montero J, López-Monteon A, Dumonteil E, Ramos-Ligonio A. House infestation dynamics and feeding sources of Triatoma dimidiata in central Veracruz, Mexico. Am J Trop Med Hyg. 2012;86(4):677.

28. Velásquez-Ortiz N, Hernández C, Cantillo-Barraza O, Medina M, Medina-Alfonso M, Suescún-Carrero S, et al. Estimating the genetic structure of Triatoma dimidiata (Hemiptera: Reduviidae) and the transmission dynamics of Trypanosoma cruzi in Boyacá, eastern Colombia. PLoS Negl Trop Dis. 2022 Jul;16(7):e0010534.

29. Monroy C, Rodas A, Mejía M, Rosales R, Tabaru Y. Epidemiology of Chagas disease in Guatemala: infection rate of Triatoma dimidiata, Triatoma nitida and Rhodnius prolixus (Hemiptera, Reduviidae) with Trypanosoma cruzi and Trypanosoma rangeli (Kinetoplastida, Trypanosomatidae). Vol. 98, Memorias do Instituto Oswaldo Cruz. SciELO Brasil; 2003. p. 305–10.

30. Monroy MC, Bustamante DM, Rodas AG, Enriquez ME, Rosales RG. Habitats, Dispersion and Invasion of Sylvatic Triatoma dimidiata (Hemiptera: Reduviidae: Triatominae) in Petén, Guatemala. J Med Entomol. 2003 Nov 1;40(6):800–6.

31. Nakagawa J, Juárez J, Nakatsuji K, Akiyama T, Hernandez G, Macal R, et al. Geographical characterization of the triatomine infestations in north–central Guatemala. Ann Trop Med Parasitol. 2005 Apr 1;99(3):307–15.

32. Barbu C, Dumonteil E, Gourbière S. Optimization of Control Strategies for Non-Domiciliated Triatoma dimidiata, Chagas Disease Vector in the Yucatán Peninsula, Mexico. PLoS Negl Trop Dis. 2009 abr;3(4):e416.

33. Dumonteil E, Ruiz-Piña H, Rodriguez-Félix E, Barrera-Pérez M, Ramirez-Sierra MJ, Rabinovich JE, et al. Re-infestation of houses by Triatoma dimidiata after intra-domicile insecticide application in the Yucatán peninsula, Mexico. Mem Inst Oswaldo Cruz. 2004 May;99:253–6.

34. Ferral J, Chavez-Nuñez L, Euan-Garcia M, Ramirez-Sierra MJ, Najera-Vazquez MR, Dumonteil E. Comparative Field Trial of Alternative Vector Control Strategies for Non-Domiciliated Triatoma dimidiata. Am J Trop Med Hyg. 2010 Jan;82(1):60–6.

35. Pinto C. Valor do rostro e antenas na caracterização dos gêneros de Triatomídeos. Hemiptera, Reduvidioidea. Bol Biol. 1931;19:45–136.

36. Usinger RL. Notes and descriptions of neotropical Triatominae (Hemiptera, Reduviidae). Pan-Pac Entomol. 1941;17(2):49–57.

37. Lent H, Wygodzinsky P. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. Bull Am Mus Nat Hist. 1979;163(3):123–520.

38. Bustamante DM, Monroy C, Menes M, Rodas A, Salazar-schettino PM, Rojas G, et al. Metric Variation Among Geographic Populations of the Chagas Vector Triatoma dimidiata (Hemiptera: Reduviidae: Triatominae) and Related Species. J Med Entomol. 2004 May 1;41(3):296–301.

39. Calderón-Fernández GM, Girotti JR, Juárez MP. Cuticular Hydrocarbons of Triatoma dimidiata (Hemiptera: Reduviidae): Intraspecific Variation and Chemotaxonomy. J Med Entomol. 2011 Mar 1;48(2):262–71.

40. Catalá S, Sachetto C, Moreno M, Rosales R, Salazar-Schettino PM, Gorla D. Antennal Phenotype of Triatoma dimidiata Populations and Its Relationship with Species of phyllosoma and protracta Complexes. J Med Entomol. 2005 Sep 1;42(5):719–25.

41. Fernández GC, Juárez MP, Ramsey J, Schettino PMS, Monroy MC, Ordoñez R, et al. Cuticular Hydrocarbon Variability Among Triatoma dimidiata (Hemiptera: Reduviidae) Populations from Mexico and Guatemala. J Med Entomol. 2005 Sep 1;42(5):780–8.

42. Marcilla A, Bargues MD, Ramsey JM, Magallon-Gastelum E, Salazar-Schettino PM, Abad-Franch F, et al. The ITS-2 of the Nuclear rDNA as a Molecular Marker for Populations, Species, and Phylogenetic Relationships in Triatominae (Hemiptera: Reduviidae), Vectors of Chagas Disease. Mol Phylogenet Evol. 2001 Jan 1;18(1):136–42.

43. Panzera F, Ferrandis I, Ramsey J, Ordòñez R, Salazar-Schettino PM, Cabrera M, et al. Chromosomal variation and genome size support existence of cryptic species of Triatoma dimidiata with different epidemiological importance as Chagas disease vectors. Trop Med Int Health. 2006;11(7):1092–103.

44. Panzera F, Pérez R, Panzera Y, Ferrandis I, Ferreiro MJ, Calleros L. Cytogenetics and Genome Evolution in the Subfamily Triatominae (Hemiptera, Reduviidae). Cytogenet Genome Res. 2010 Apr 19;128(1–3):77–87.

45. Tamay-Segovia P, Alejandre-Aguilar R, Martínez F, Villalobos G, Zavala-Diaz de la Serna FJ, Torre P de la, et al. Two Triatoma dimidiata clades (Chagas disease vector) associated with different habitats in southern Mexico and Central America. Am J Trop Med Hyg. 2008;78(3):472–8.

46. Justi SA, Cahan S, Stevens L, Monroy C, Lima-Cordón R, Dorn PL. Vectors of diversity: Genome wide diversity across the geographic range of the Chagas disease vector *Triatoma dimidiata* sensu lato (Hemiptera: Reduviidae). Mol Phylogenet Evol. 2018 Mar 1;120:144–50.

47. Caranci AT, Grieco JP, Achee NL, Hoel DF, Bautista K, King R, et al. Distribution of Triatoma dimidiata sensu lato (Reduviidae: Triatominae) and Risk Factors Associated with Household Invasion in Northern Belize, Central America. J Med Entomol. 2022;59(2):764–71.

48. Pita S, Lorite P, Vela J, Mora P, Palomeque T, Thi KP, et al. Holocentric chromosome evolution in kissing bugs (Hemiptera: Reduviidae: Triatominae): diversification of repeated sequences. Parasit Vectors. 2017 Sep 6;10(1):410.

49. DNA Zoo. DNA Zoo. 2024 [cited 2024 Oct 9]. Rhodnius_prolixus. Available from: https://www.dnazoo.org/assemblies/rhodnius_prolixus

50. Liu Q, Guo Y, Zhang Y, Hu W, Li Y, Zhu D, et al. A chromosomal-level genome assembly for the insect vector for Chagas disease, Triatoma rubrofasciata. GigaScience. 2019 Aug 1;8(8):giz089.

51. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, Rivera-Pomar R, Monteiro FA, Minx P, et al. Genome of *Rhodnius prolixus* , an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc Natl Acad Sci. 2015 Dec;112(48):14936–41.

52. Peterson JK, MacDonald ML, Ellis VA. Genome report: First whole genome sequence of Triatoma sanguisuga (Le Conte, 1855), vector of Chagas disease. BioRxiv Prepr Serv Biol. 2024 Oct 8;2024.10.07.611472.

53. NCBI [Internet]. [cited 2024 Oct 30]. Triatoma sanguisuga genome assembly UDEL_Tsan_1.0. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_041753995.1/

54. NCBI [Internet]. [cited 2024 Oct 30]. Triatoma infestans genome assembly UVM_Tinf_1.0. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_011037195.1/

55. NCBI [Internet]. [cited 2024 Oct 30]. Panstrongylus geniculatus genome assembly ihPanGeni1.hap1.1. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_964188295.1/

56. Hernández-Vargas MJ, Gil J, Lozano L, Pedraza-Escalona M, Ortiz E, Encarnación-Guevara S, et al. Proteomic and transcriptomic analysis of saliva components from the hematophagous reduviid *Triatoma pallidipennis*. J Proteomics. 2017 Jun 6;162:30–9.

57. Brito T, Julio A, Berni M, Poncio L de C, Bernardes ES, Araujo H, et al. Transcriptomic and functional analyses of the piRNA pathway in the Chagas disease vector Rhodnius prolixus. PLoS Negl Trop Dis. 2018 Oct 10;12(10):e0006760.

58. Latorre-Estivalis JM, Robertson HM, Walden KKO, Ruiz J, Gonçalves LO, Guarneri AA, et al. The molecular sensory machinery of a Chagas disease vector: expression changes through imaginal moult and sexually dimorphic features. Sci Rep. 2017 Jan 6;7(1):40049.

59. Carvalho DB de, Congrains C, Chahad-Ehlers S, Pinotti H, Brito RA de, Rosa JA da. Differential transcriptome analysis supports Rhodnius montenegrensis and Rhodnius robustus (Hemiptera, Reduviidae, Triatominae) as distinct species. PLOS ONE. 2017 abr;12(4):e0174997.

60. Ribeiro JMC, Andersen J, Silva-Neto MAC, Pham VM, Garfield MK, Valenzuela JG. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. Insect Biochem Mol Biol. 2004 Jan 1;34(1):61–79.

61. Assumpção TCF, Francischetti IMB, Andersen JF, Schwarz A, Santana JM, Ribeiro JMC. An insight into the sialome of the blood-sucking bug *Triatoma infestans*, a vector of Chagas' disease. Insect Biochem Mol Biol. 2008 Feb 1;38(2):213–32.

62. Martínez-Barnetche J, Lavore A, Beliera M, Téllez-Sosa J, Zumaya-Estrada FA, Palacio V, et al. Adaptations in energy metabolism and gene family expansions revealed by comparative transcriptomics of three Chagas disease triatomine vectors. BMC Genomics. 2018 Apr 27;19(1):296.

63. Ribeiro JMC, Genta FA, Sorgine MHF, Logullo R, Mesquita RD, Paiva-Silva GO, et al. An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug, Rhodnius prolixus. PLoS Negl Trop Dis. 2014 ene;8(1):e2594.

64. Defferrari MS, Da Silva SR, Orchard I, Lange AB. A Rhodnius prolixus Insulin Receptor and Its Conserved Intracellular Signaling Pathway and Regulation of Metabolism. Front Endocrinol [Internet]. 2018 Dec 6 [cited 2024 Oct 27];9. Available from: https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2018.00745/full

65. Medeiros MN, Logullo R, Ramos IB, Sorgine MHF, Paiva-Silva GO, Mesquita RD, et al. Transcriptome and gene expression profile of ovarian follicle tissue of the triatomine bug *Rhodnius prolixus*. Insect Biochem Mol Biol. 2011 Oct 1;41(10):823–31.

66. Traverso L, Lavore A, Sierra I, Palacio V, Martinez-Barnetche J, Latorre-Estivalis JM, et al. Comparative and functional triatomine genomics reveals reductions and expansions in insecticide resistance-related gene families. PLoS Negl Trop Dis. 2017 Feb 15;11(2):e0005313.

67. Zumaya-Estrada FA, Martínez-Barnetche J, Lavore A, Rivera-Pomar R, Rodríguez MH. Comparative genomics analysis of triatomines reveals common first line and inducible immunity-related genes and the absence of Imd canonical components among hemimetabolous arthropods. Parasit Vectors. 2018 Jan 22;11(1):48.

68. Brito RN, Geraldo JA, Monteiro FA, Lazoski C, Souza RCM, Abad-Franch F. Transcriptome-based molecular systematics: Rhodnius montenegrensis (Triatominae) and its position within the Rhodnius prolixus–Rhodnius robustus cryptic–species complex. Parasit Vectors. 2019 Jun 17;12(1):305.

69. Lee SG, Na D, Park C. Comparability of reference-based and reference-free transcriptome analysis approaches at the gene expression level. BMC Bioinformatics. 2021 Oct 21;22(11):310.

70. Huang M, Kingan S, Shoue D, Nguyen O, Froenicke L, Galvin B, et al. Improved high quality sand fly assemblies enabled by ultra low input long read sequencing. Sci Data. 2024 Aug 24;11(1):918.

71. International Glossina Genome Initiative, Attardo GM, Abila PP, Auma JE, Baumann AA, Benoit JB, et al. Genome Sequence of the Tsetse Fly (Glossina morsitans): Vector of African Trypanosomiasis. Science. 2014 Apr 25;344(6182):380–6.

72. Matthews B, O D, Sb K, S K, I A, Je C, et al. Improved reference genome of Aedes aegypti informs arbovirus vector control. Nature [Internet]. 2018 Nov [cited 2024 Oct 13];563(7732). Available from: https://pubmed.ncbi.nlm.nih.gov/30429615/

73. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu Z (Jake), et al. Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science. 2007 Jun 22;316(5832):1718–23.

74. Zamyatin A, Avdeyev P, Liang J, Sharma A, Chen C, Lukyanchikova V, et al. Chromosome-level genome assemblies of the malaria vectors Anopheles coluzzii and Anopheles arabiensis. GigaScience. 2021 Mar 1;10(3):giab017.

75. Denton A, Yatsenko H, Jay J, kh, Howard C. Sanger Tree of Life Wet Laboratory Protocol Collection. 2023 Oct 2 [cited 2024 Oct 2]; Available from: https://www.protocols.io/view/sanger-tree-of-life-wet-laboratory-protocol-collec-cy5rxy56

76. Oatley G, Denton A, Howard C. Sanger Tree of Life HMW DNA Extraction: Automated MagAttract v.2. 2023 Sep 29 [cited 2024 Oct 2]; Available from: https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-extraction-automated-m-czjux4nw

77. Clayton-Lucey I, Howard C, Bates A. Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor®3 for LI PacBio. 2023 Sep 29 [cited 2024 Oct 2]; Available from: https://www.protocols.io/view/sanger-tree-of-life-hmw-dna-fragmentation-diagenod-czhmx346

78. Strickland M, Cornwell C, Howard C. Sanger Tree of Life Fragmented DNA clean up: Manual SPRI. 2023 Sep 29 [cited 2024 Oct 2]; Available from: https://www.protocols.io/view/sanger-tree-of-life-fragmented-dna-clean-up-manual-czhkx34w

79. Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. BMC Genomics. 2022 Feb 22;23(1):157.

80. Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 2017 Sep 1;33(17):2759–61.

81. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020 Mar 18;11(1):1432.

82. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021 Feb;18(2):170–5.

83. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinforma Oxf Engl. 2020 May 1;36(9):2896–8.

84. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Feb 1;10(2):giab008.

85. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) [Internet]. 2019 [cited 2024 Oct 2]. p. 314–24. Available from: https://ieeexplore.ieee.org/document/8820962

86. Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, et al. Pairtools: from sequencing data to chromosome contacts. bioRxiv. 2023 Feb 15;2023.02.13.528389.

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

88. Zimin AV, Salzberg SL. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLOS Comput Biol. 2020 Jun 26;16(6):e1007981.

89. ncbi/fcs [Internet]. NCBI - National Center for Biotechnology Information/NLM/NIH; 2024 [cited 2024 Oct 2]. Available from: https://github.com/ncbi/fcs

90. Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol. 2024 Feb 26;25(1):60.

91. Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. Bioinformatics. 2022 Sep 2;38(17):4214–6.

92. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 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 Oct 1;38(10):4647–54.

93. Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Consortium DT of L, Formenti G, Abueg L, et al. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio High Fidelity reads [Internet]. bioRxiv; 2023 [cited 2024 Oct 2]. p. 2022.12.23.521667. Available from: https://www.biorxiv.org/content/10.1101/2022.12.23.521667v2

94. Geneious [Internet]. [cited 2024 Oct 27]. Geneious | Bioinformatics Software for Sequence Data Analysis. Available from: https://www.geneious.com/

95. Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst. 2018 Feb 7;6(2):256.

96. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020 Apr 28;117(17):9451–7.

97. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2015 Nov 26;44(Database issue):D81.

98. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011 May 15;29(7):644–52.

99. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003 Oct 1;31(19):5654–66.

100. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008 Mar 1;24(5):637–44.

101. Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biol. 2006 Dec 7;4(1):41.

102. Dotson EM, Beard CB. Sequence and organization of the mitochondrial genome of the Chagas disease vector, Triatoma dimidiata. Insect Mol Biol. 2001;10(3):205–15.

103. Jung H, Ventura T, Chung JS, Kim WJ, Nam BH, Kong HJ, et al. Twelve quick steps for genome assembly and annotation in the classroom. PLOS Comput Biol. 2020 Nov 12;16(11):e1008325.

104. Whibley A, Kelley JL, Narum SR. The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. Mol Ecol Resour. 2021;21(3):641–52.

105. Hjelmen CE. Genome size and chromosome number are critical metrics for accurate genome assembly assessment in Eukaryota. Genetics. 2024 Aug 1;227(4):iyae099.

106. Azevedo SLC de, Catanho M, Guimarães ACR, Galvão TC. Genomic surveillance: a potential shortcut for effective Chagas disease management. Mem Inst Oswaldo Cruz. 2023 Jan 20;117:e220164.

107. Flynn JM, Ahmed-Braimah YH, Long M, Wing RA, Clark AG. High-Quality Genome Assemblies Reveal Evolutionary Dynamics of Repetitive DNA and Structural Rearrangements in the Drosophila virilis Subgroup. Genome Biol Evol. 2024 Jan 1;16(1):evad238.

108. Housman G, Tung J. Next-generation primate genomics: New genome assemblies unlock new questions. Cell. 2023 Dec 7;186(25):5433–7.

109. Inzaule SC, Tessema SK, Kebede Y, Ouma AEO, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. Lancet Infect Dis. 2021 Feb 12;21(9):e281.

110. Moss S, Pretorius E, Ceesay S, Hutchins H, da Silva ET, Ndiath MO, et al. Genomic surveillance of Anopheles mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. Parasit Vectors. 2024 Jan 4;17(1):10.

111. Gunter SM, Nelson A, Kneubehl AR, Justi SA, Manzanero R, Zielinski-Gutierrez E, et al. Novel species of Triatoma (Hemiptera: Reduviidae) identified in a case of vectorial transmission of Chagas disease in northern Belize. Sci Rep. 2024 Jan 16;14:1412.

112. Aguilera-Uribe M, Meza-Lázaro RN, Kieran TJ, Ibarra-Cerdeña CN, Zaldívar-Riverón A. 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 J Mol Epidemiol Evol Genet Infect Dis. 2020 Oct;84:104373.

113. Zou Y, Zhu W, Sloan DB, Wu Z. Long-read sequencing characterizes mitochondrial and plastid genome variants in Arabidopsis msh1 mutants. Plant J Cell Mol Biol. 2022 Nov;112(3):738–55.

114. Cigarroa-Toledo N, Baak-Baak CM, Chan-Pérez JI, Hernandez-Mena DI, Guardia KCA, Ocaña-Correa MF, et al. Dataset of assembly and annotation of the mitogenomes of *Triatoma dimidiata* and *Triatoma huehuetenanguensis* captured from Yucatán, México. Data Brief. 2024 Feb 1;52:109866.

115. Dijk EL van, Naquin D, Gorrichon K, Jaszczyszyn Y, Ouazahrou R, Thermes C, et al. Genomics in the long-read sequencing era. Trends Genet. 2023 Sep 1;39(9):649–71.

116. Butenko A, Lukeš J, Speijer D, Wideman JG. Mitochondrial genomes revisited: why do different lineages retain different genes? BMC Biol. 2024 Jan 25;22(1):15.

117. Dorn PL, de la Rúa NM, Axen H, Smith N, Richards BR, Charabati J, et al. Hypothesis testing clarifies the systematics of the main Central American Chagas disease vector, *Triatoma dimidiata* (Latreille, 1811), across its geographic range. Infect Genet Evol. 2016 Oct 1;44:431–43.

118. Mora P, Pita S, Montiel EE, Rico-Porras JM, Palomeque T, Panzera F, et al. Making the Genome Huge: The Case of Triatoma delpontei, a Triatominae Species with More than 50% of Its Genome Full of Satellite DNA. Genes. 2023 Jan 31;14(2):371.

119. Pita S, Panzera F, Mora P, Vela J, Cuadrado Á, Sánchez A, et al. Comparative repeatome analysis on Triatoma infestans Andean and Non-Andean lineages, main vector of Chagas disease. PLoS ONE. 2017 Jul 19;12(7):e0181635.

120. Loreto EL, Deprá M, Diesel JF, Panzera Y, Valente-Gaiesky VLS. Drosophila relics hobo and hobo-MITEs transposons as raw material for new regulatory networks. Genet Mol Biol. 2018 Mar 26;41(1 Suppl 1):198.

121. Montiel EE, Panzera F, Palomeque T, Lorite P, Pita S. Satellitome Analysis of Rhodnius prolixus, One of the Main Chagas Disease Vector Species. Int J Mol Sci. 2021 Jan;22(11):6052.

122. Pita S, Panzera F, Mora P, Vela J, Cuadrado Á, Sánchez A, et al. Comparative repeatome analysis on Triatoma infestans Andean and Non-Andean lineages, main vector of Chagas disease. PLOS ONE. 2017 Jul 19;12(7):e0181635.

123. Boulesteix M, Weiss M, Biémont C. Differences in Genome Size Between Closely Related Species: The Drosophila melanogaster Species Subgroup. Mol Biol Evol. 2006 Jan 1;23(1):162–7.

124. Gregory TR. Synergy between sequence and size in Large-scale genomics. Nat Rev Genet. 2005 Sep;6(9):699–708.

125. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018 Nov 19;19(1):199.

126. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2012 Jan;13(1):36–46.

127. Kazazian HH. Mobile Elements: Drivers of Genome Evolution. Science. 2004 Mar 12;303(5664):1626–32.

128. Guigó R. Genome annotation: From human genetics to biodiversity genomics. Cell Genomics. 2023 Aug 9;3(8):100375.

129. Brandies P, Peel E, Hogg CJ, Belov K. The Value of Reference Genomes in the Conservation of Threatened Species. Genes. 2019 Oct 25;10(11):846.

130. Espinosa E, Bautista R, Larrosa R, Plata O. Advancements in long-read genome sequencing technologies and algorithms. Genomics. 2024 May;116(3):110842.

131. Valenzuela JG, Aksoy S. Impact of vector biology research on old and emerging neglected tropical diseases. PLoS Negl Trop Dis. 2018 May 31;12(5):e0006365.

132. Worley KC, Richards S, Rogers J. The Value of New Genome References. Exp Cell Res. 2016 Dec 23;358(2):433.

133. Freitas L, Nery MF. Expansions and contractions in gene families of independently-evolved blood-feeding insects. BMC Evol Biol. 2020 Jul 17;20:87.

134. Pollett S, Fauver JR, Maljkovic Berry I, Melendrez M, Morrison A, Gillis LD, et al. Genomic Epidemiology as a Public Health Tool to Combat Mosquito-Borne Virus Outbreaks. J Infect Dis. 2020 Mar 28;221(Supplement_3):S308–18.

135. Sproul JS, Hotaling S, Heckenhauer J, Powell A, Marshall D, Larracuente AM, et al. Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges. Genome Res. 2023 Oct;33(10):1708.