Trypanosoma cruzi Has 32 Chromosomes: A Telomere-to-Telomere Assembly Defines Its Karyotype.

Greif, G.[1#], Díaz-Viraqué, F.[1], Chiribao, M.L.[1,2], Sanz-Rodriguez, Carlos E.[1], Robello, C.[1,2#]

[1]Laboratorio de Interacciones Hospedero-patógeno — UBM, Institut Pasteur de Montevideo, Montevideo, Uruguay.
[2]Departamento de Bioquímica, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay.
[#]Correspondence: robello@pasteur.edu.uy; ggreif@pasteur.edu.uy

Keywords: *Trypanosoma cruzi*; Chromosome complement; Karyotype; TriTryps; Telomere-to-Telomere.

Abstract

Trypanosoma cruzi, the causative agent of Chagas disease, exhibits remarkable genomic plasticity and possesses an expanded genome rich in multigene families. However, its precise chromosomal composition has remained elusive due to challenges in resolving repetitive regions. Using PacBio HiFi long-read sequencing, we assembled the complete nuclear genome of the Dm28c strain into 32 telomere-to-telomere chromosomes. The assembly revealed conserved chromosomal structures and synteny patterns when compared to the independently sequenced Dm25 strain, both of which belong to clade A (TcI). This indicates that the 32-chromosome karyotype is a stable and conserved feature of *T. cruzi*. Through a comprehensive annotation pipeline, we refined gene content and resolved haplotypes, resulting in an approximately 25% reduction in core gene redundancy. All chromosomes demonstrate a conserved compartmentalized architecture consisting of core, disruptive, and newly characterized subtelomeric regions. Subtelomeres, enriched in RHS, TS, and DGF-1 genes but lacking MASP and mucins, create a distinct third genomic compartment. Telomere analysis showed significant length heterogeneity, which may influence subtelomeric gene regulation. Overall, this work defines the complete chromosomal complement of *T. cruzi*, establishing a solid framework for comparative genomics and enabling detailed studies into genome organization, antigenic variability, and evolutionary dynamics across strains and clades. This assembly redefines the genomic reference landscape, paving the way for future research on *T. cruzi's* fundamental biology, pathogenic diversity, and adaptive mechanisms.

Introduction.

Trypanosoma cruzi, the causative agent of Chagas disease, belongs to the family Trypanosomatidae, which is composed exclusively of parasitic organisms. Three of these, known as the "TriTryps"—*Leishmania spp.*, *Trypanosoma brucei*, and *T. cruzi*—pose significant challenges to human and animal health and have been extensively studied. They share several biological features, including RNA editing, polycistronic transcription, a single mitochondrion with kinetoplast DNA (kDNA), *trans*-splicing, and the absence of chromosomal condensation during mitosis. Additionally, each of these species possesses unique characteristics. In *T. cruzi*, the most notable is its genome expansion due to multigene families, the majority of which encode surface GPI-anchored proteins. These include trans-sialidases/TS (Schenkman et al., 1991; Cazzulo and Frasch, 1992; Buscaglia et al., 1999; Freitas et al., 2011; Rubin and Schenkman, 2012; Freire-de-Lima et al., 2015), mucins (Frasch, 2000; Buscaglia et al., 2006), mucin associated surface proteins/MASP (De Pablos and Osuna, 2012; Leão et al., 2022), GP63 (Alvarez et al., 2012), disperse gene family 1/DGF-1 (Wincker et al., 1992; Ramirez, 2023), and retroposon hot spot/RHS (Bringaud et al., 2002; Bernardo et al., 2020). This expansion in *T. cruzi* represents an evolutionary acquisition linked to antigenic variability, invasion, infectivity, and immune system evasion, among other factors.

Regarding the karyotype of *T. cruzi*, the absence of chromosomal condensation during mitosis has impeded the use of classical cytogenetic studies. Instead, pulsed-field gel electrophoresis (PFGE) emerged decades ago as a valuable tool for determining their karyotypes. Using this approach, it was established that *T. brucei* has 11 chromosomes (Melville et al., 1998; Berriman et al., 2005), *Leishmania major*, *L. infantum*, and *L. donovani* each have 36 (Wincker et al., 1996) and *L. mexicana* and *L. braziliensis* contain 34 and 35 chromosomes, respectively (Britto et al., 1998). However, the exact chromosome number of *T. cruzi* remains unknown. PFGE studies consistently indicate that chromosomal sizes vary among different strains, ranging from 0.5 to 3 Mb, and estimate it may have between 20 and 40 chromosomes (Henriksson et al., 1990, 1995, 1996; Santos et al., 1997; Vargas et al. 2004, Galindo et al., 2007; Souza et al., 2011). The challenge in determining its precise number, compared to *Leishmania* and *T. brucei*, arises from several factors. First, *T. cruzi* shows significant variability in chromosomal sizes across strains, which limits the exact resolution of well-defined bands in PFGE. Additionally, its genome exhibits extreme plasticity, characterized by multigene families, repetitive elements, and retroelements (Pita et al., 2019). Furthermore, homologous chromosome pairs often differ in size, leading to a complex and unresolved banding pattern.

After completing the genome projects for these three organisms (the "TriTryps"), many of the results obtained through PFGE were confirmed for *L. major* (Ivens et al., 2005)

and *T. brucei* (Berriman et al., 2005). These studies also established a comparative chromosomal map between these two species, revealing regions of high synteny and providing insights into their common ancestor (El-Sayed et al., 2005b). However, *T. cruzi* could not be included in this comparison, as the scaffolds obtained were highly fragmented due to the repetitive nature of its genome, leading to a high incidence of regional collapse (El-Sayed et al., 2005a). Consequently, although syntenic regions were identified as shared among the TriTryps, chromosomal mappings systematically excluded *T. cruzi*. Later, Weatherly et al. (2009) proposed that *T. cruzi* possesses 41 chromosomes, based on BAC-end sequencing combined with chromosomal co-location and synteny with *T. brucei* and *Leishmania*, enabling the assembly of 41 *in silico* chromosomes. Although this estimation was derived from genome sequences prone to collapse due to repetitive regions, it has served as a valuable reference for subsequent studies.

The advent of long-read sequencing marked a significant advancement in studying complex genomes, such as *T. cruzi*, which contain numerous repetitive elements and multicopy genes that are highly similar or even identical and arranged in tandem. The use of PacBio or Nanopore methodologies enabled researchers to determine a more complete genome landscape by overcoming repetitive regions and generating scaffolds that, in some cases, have lengths compatible with the expected size of chromosomes (Berná et al., 2018; Callejas-Hernandez et al., 2018; Díaz-Viraqué et al., 2019; Wang et al 2021; Hakim et al., 2024). These studies provided a comprehensive view of genome organization and opened the door for comparative genomic analyses among the TriTryps. For instance, it was determined that multigene families encoding surface proteins are not situated in subtelomeric regions, as previously proposed, but clustered along different chromosomes. In this regard, a key discovery was the identification of two distinct genomic compartments in *T. cruzi*: the core and disruptive regions (Berná et al., 2018), the latter being non-syntenic with other trypanosomatids and housing hundreds of gene variants encoding surface proteins such as trans-sialidases (TS), mucins, and mucin-associated surface proteins (MASP). Furthermore, recent findings have shown that these genomic compartments correlate with specific three-dimensional chromatin organizations designated as C and D, which represent lower and higher levels of chromatin compaction, respectively, which in addition affects global gene expression (Lima et al., 2022; Díaz-Viraqué et al., 2023).

However, the absence of telomeric sequences at their ends precludes their classification as chromosomes. The sole exception is the genome of the Dm25 strain, which comprises 24 scaffolds with telomeres at both ends, confirming their identity as complete chromosomes (Sanchez et al., 2024). This significant advancement is primarily attributed to PacBio HiFi technology.

Considering these precedents, we aimed to determine the complete karyotype of *T. cruzi*, elucidate the chromosomal organization of its core and disruptive compartments, and characterize the structural features of its telomeric and subtelomeric regions.

Results and Discussion

1- Genome assembly of the 32 chromosomes of *Trypanosoma cruzi*

The reference strain Dm28c was chosen for this study due to its extensive characterization in terms of cell cycle dynamics (Contreras et al., 1988), genomic organization (Berná et al., 2018; Díaz-Viraqué et al., 2023), proteome (Parodi-Talice et al., 2004; Parodi-Talice et al., 2007), and transcriptome (Díaz-Viraqué et al., 2023; Berná et al., 2017). Total DNA was sequenced using PacBio HiFi technology, generating 301,047 reads, 2.93 Gb of sequenced bases, and an average read quality of Q29 (**Table 1**). The assembly was conducted with the HiFiAsm algorithm, yielding an estimated genome size of 36.18 Mb with 80× coverage. The assembled genome comprises 32 scaffolds with an average GC content of 51.06% (**Table 1**). Analysis of the scaffold ends identified telomeric sequences (CCCTAA/TTAGGG) (Barros et al., 2012) at both termini of all the 32 scaffolds adjacent to a 189bp junction (Chiurillo et al., 1999), indicating that they represent complete chromosomes. In summary, we established that *T. cruzi* has 32 chromosomes, which were sequentially numbered from 1 to 32, from longest to shortest (**Figure 1**). Chromosome lengths range from 2.49 to 0.58 Mb (**Table 1, Figure 1B**). Based on the assembly presented here and deposited in GenBank, chromosome orientation is defined with the 5′ to 3′ direction running from left to right. This convention is followed throughout the manuscript when describing chromosomal features and performing comparative analyses.

Efforts to achieve chromosomal completion have been underway since 2005, hindered by the limitations of current sequencing methods in resolving repetitive regions. The introduction of PacBio HiFi sequencing has now made this feasible. Although PFGE studies produced distinct bands, they were not stoichiometric since overlapping bands were often found in most *T. cruzi* strains (Henriksson et al., 1996; Santos et al., 1997). The chromosome sizes reported by PFGE across different clades ranged from 0.46 to 3.27 Mb (Henriksson et al., 1990, 1996; Santos et al., 1997; Vargas et al., 2004; Henriksson et al., 1995; Galindo et al., 2007; Souza et al., 2011). Notably, the karyotype of Dm28c examined by PFGE revealed a longest band of 2.50 Mb and a shortest band of 0.57 Mb (Souza et al., 2011), with sizes consistent with those found in this study (0.58 to 2.49 Mb; **Figure 1** and **Table 1**).

Altogether, these results confirm that the Dm28c genome is fully assembled into 32 complete chromosomes, whose sizes and structure are consistent with previous karyotypic estimates, marking a milestone in the chromosomal resolution of *T. cruzi*.

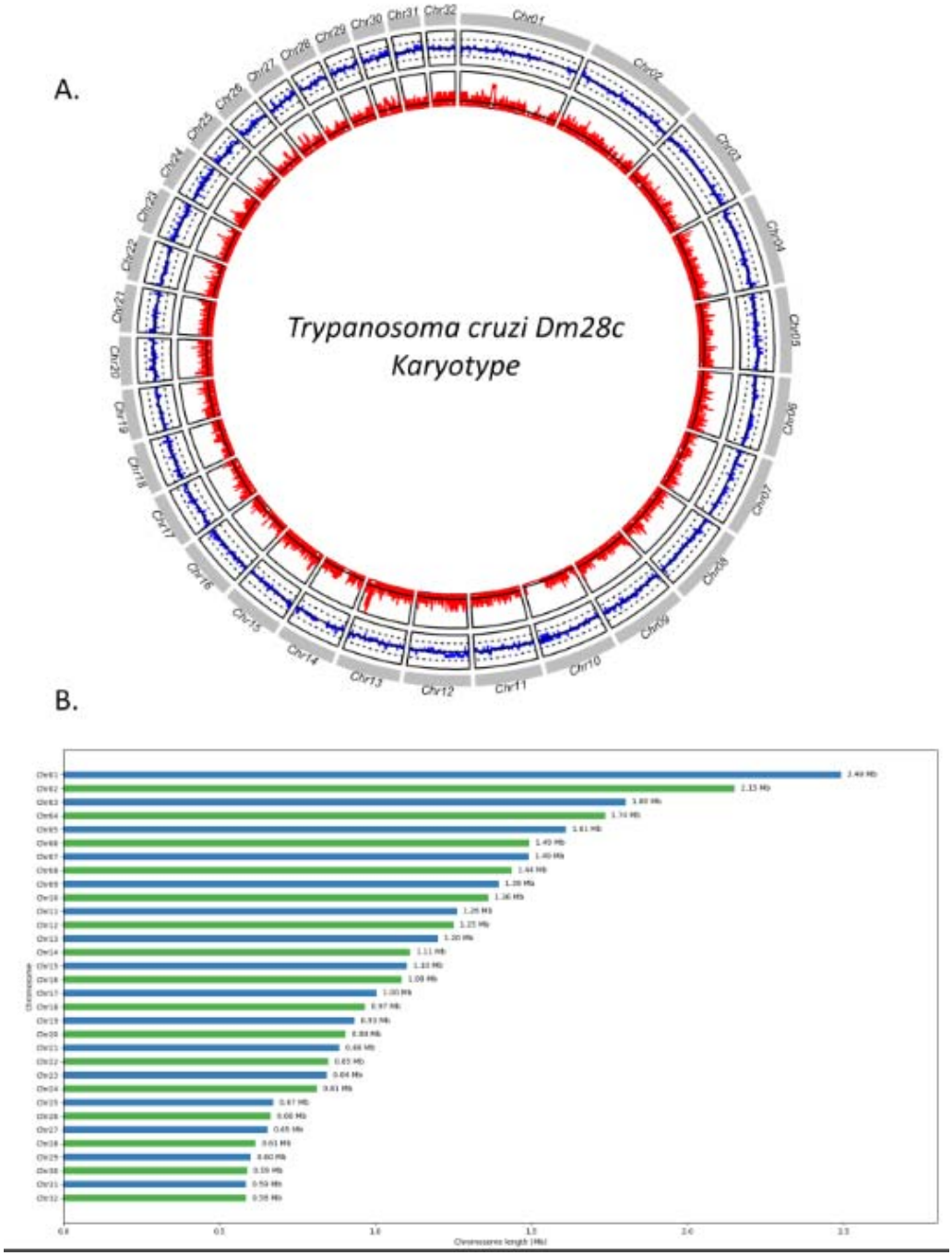

Figure 1. The karyotype of *T. cruzi*. A. Circos plot karyotype representation. The outer grey lines represent chromosomes; the middle circle represents the percentage of GC content along chromosomes (the black line indicates 50% GC, and the dotted black lines indicate 25% and 75% GC content). The inner circle in red indicates gene density calculated using overlapping windows of 5000 bases, sliding by 500 bases. **B.** Bar charts showing the *T. cruzi* chromosome lengths in Mb.

Table 1. Sequencing, genome assembly statistics. Final genome assembly features and annotation.

Elements	Metric	Value
Sequencing Statistics	Number of reads	301047
	Number of bases sequenced	2.92 Gb
	N50	10.68 Kb
	Average read quality	Q29
Genome Assembly Statistics	Number of contigs	51
	Total length of assembly	46.64 Mb
	G+C content	51.35
	Average coverage HiFi reads	62.7
	N50	1.2
Genome	Number of chromosomes	32
	Genome length	36.18 Mb
	Longest chromosome	2.49 Mb
	Shortest chromosome	0.58 Mb
	Average coverage HiFi reads	80.80
	N50	1.26 Mb
	G+C content	51.06
Number of genes	Protein-coding genes	13825
	RHS	683
	Trans-sialidases	1185
	Mucin	384
	MASP	530
	DGF-1	248
	Transposable elements	1816
	Putative pseudogenes	3218

2- Conservation of the 32-Chromosome Karyotype Across *T. cruzi* Strains

Given that it is widely accepted that *T. cruzi* exhibits significant genomic plasticity (Reis-Cunha and Bartholomeu, 2018; Cruz-Saavedra et al., 2022), we aimed to determine whether this molecular karyotype is specific to the strain studied or reflects a broader characteristic of the species. To explore this, we compared our assembly with the genome assembly of the Dm25 strain. It is important to note that, although the strain names are similar, they do not denote clones from the exact origin, as they were isolated in different regions and at different times (Sanchez et al., 2024; Contreras et al., 1988). When we conducted a comparative analysis using the chromosomes of Dm28c_T2T as a reference, we observed that Dm25 also possesses 32 chromosomes (**Figure 2A**). Furthermore, with the Dm28c_T2T genome, we resolved all previously assembled pseudo-chromosomes (chromosomes lacking telomeres at both ends) into more than two contigs in haplotype 1 of the Dm25 strain. For instance, Chromosome 13 in Dm28c_T2T corresponds to contig c1 of pseudo-chromosome 16 and contig c2 from pseudo-chromosome 30 in Dm25 (**Figure 2 A and B**).

Notably, the chromosomal complement coincides and exhibits structural conservation when analyzing the GC content (**Figure 2 C and D**). This high degree of conservation strongly indicates that a karyotype with 32 chromosomes is a characteristic feature of the *T. cruzi* species rather than a trait exclusive to these two strains. However, it is essential to note that from a phylogenetic perspective, *T. cruzi* consists of two major groups (1 and 2) and four clades - A to D - (Berná et al., 2021), traditionally categorized as DTUs (Discrete Typing Units) TcI to TcVI (Zingales et al., 2009). Dm28c and Dm25 are part of clade A (TcI), highlighting the necessity for future studies to concentrate on clades B, C, and D, as well as the B/C hybrids (TcV and TcVI), to determine whether this chromosomal organization is conserved across the entire species.

Determining the chromosomal complement of *T. cruzi* represents a significant milestone that provides a solid foundation for investigating genome organization, synteny, and evolutionary dynamics. A complete karyotype enables high-resolution comparative genomics, facilitating the identification of structural variations associated with pathogenicity, host adaptation, and drug resistance. This advancement is also key to understanding how genomic plasticity influences the life cycle, transmission, and virulence of the parasite, ultimately contributing to better strategies for tackling Chagas disease.

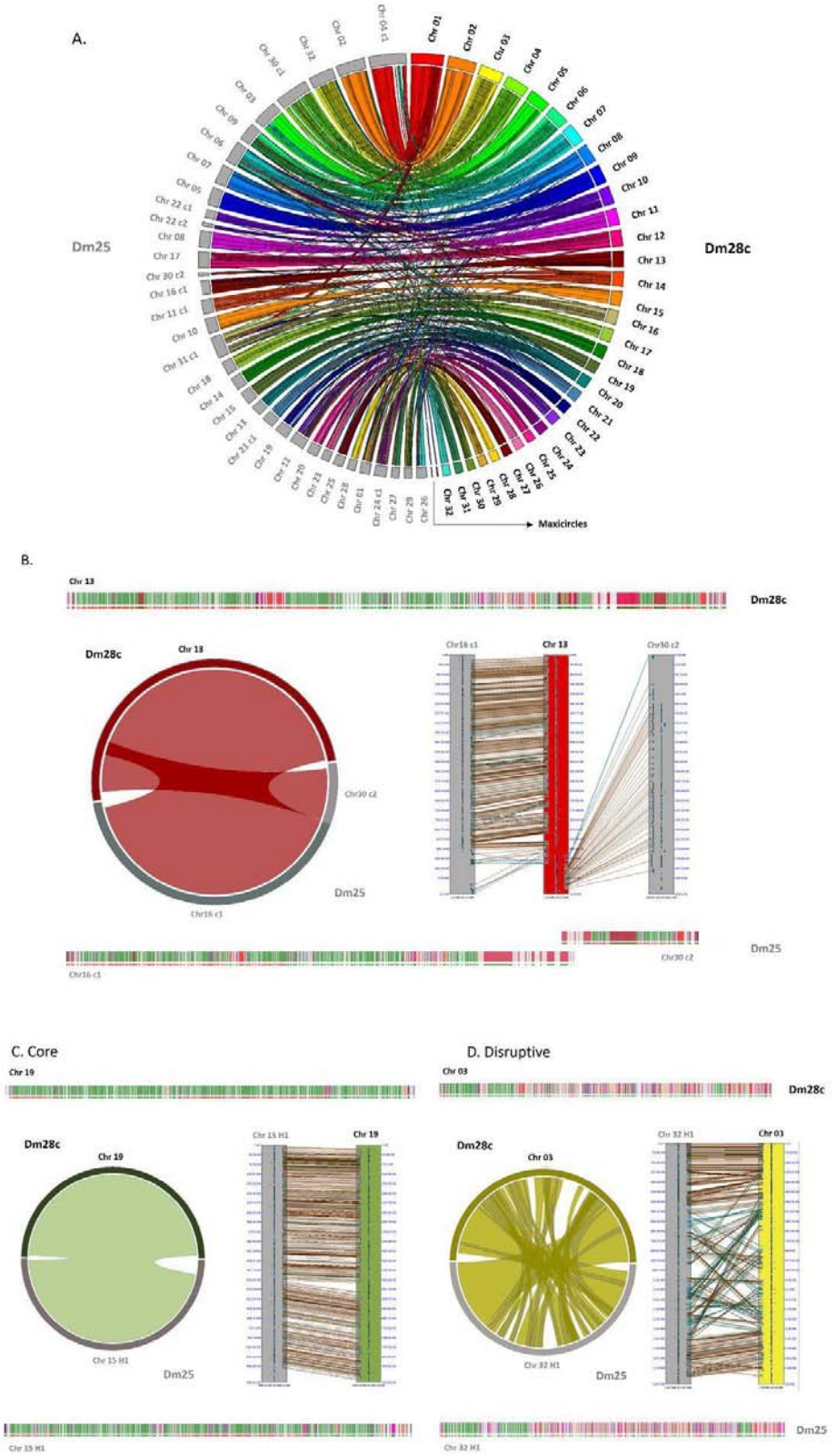
A.
Dm25
Dm28c
Maxicircles
B.
Chr 13
Dm28c
Chr 13
Chr30 c2
Dm25
Chr16 c1
Chr16 c1
Chr 13
Chr30 c2
Chr30 c2
Dm25
Chr16 c1
C. Core
Chr 19
Dm28c
Chr 19
Chr 15 H1
Chr 19
Chr 15 H1
Dm25
Chr 15 H1
D. Disruptive
Chr 03
Dm28c
Dm28c
Chr 03
Chr 32 H1
Chr 03
Chr 32 H1
Dm25
Dm25
Chr 32 H1

Figure 2. Comparison of the Dm28c and Dm25 genomes. A. The 32 chromosomes of *Dm28c* (colored) were mapped against all contigs present in haplotype 1 of *Dm25* (in gray). All contigs in the Dm25 assembly are represented in the 32 chromosomes of the Dm28c strain. **B.** The fragmented chromosome in Dm25 is assembled into the T2T chromosome in Dm28c. At the top (Dm28c) and bottom (Dm25) are schematic representations of chromosomes retrieved from the web page. In the middle, Symap results: Left panel: circos synteny plot; Right panel: synteny blocks. **C.** Example of synteny in the core chromosome and **D.** in the disruptive chromosome. At the top and bottom are schematic representations of chromosomes. In the middle, the results of synteny blocks obtained with Symap are represented by a circos plot and chromosomal representation similar to **C.** In all figures, gene representation colors align with the web interface's color code (https://cruzi.pasteur.uy/): RHS (brown), Trans-sialidase (orange-red), DGF-1 (red), Mucin and MASP (shades of blue), GP63 (orange), Conserved genes (green), and Pseudogenes (magenta).

3- Genome Annotation

The new pipeline described in Materials and Methods (**Supplementary Figure 1A**) allowed us to re-annotate the Dm28c genome. We noticed minor changes from our previous annotation (Berná et al., 2018) concerning the number and category of annotated features of proteins. The most significant difference is the nearly 25% reduction in the number of core genes (12230 to 9720; **Supplementary Table 1**). This decrease can be attributed to the fact that in our previous assembly, genes located in regions where haplotypes could be separated appeared twice. In contrast, those in collapsed haplotypes appeared only once.

In this work, PacBio HiFi data enabled the accurate separation of haplotype genomes and improved the determination of the number of genes in the haploid genome. Additionally, variations were observed in the number of genes from different gene families, with some increasing and others decreasing. This can be linked to the interaction of two opposing factors arising from PacBio HiFi sequencing: enhanced resolution—or decreased fragmentation—in repetitive regions (e.g., tandem, direct, and inverted repeats) and a more precise estimation of gene count due to haplotype separation.

4. Chromosomal organization of core and disruptive compartments

It is now well established that the genome of *T. cruzi* is compartmentalized (Berná et al., 2018; Lima et al., 2022; Rosón et al., 2022; Vitarelli et al., 2024). The core compartment exhibits synteny with other trypanosomatids. In contrast, the disruptive compartment consists of multigene families coding for trans-sialidases, mucins, and MASP and is characterized by the loss of synteny with other trypanosomatids. We previously demonstrated that these regions are not in the subtelomeres, unlike what was previously postulated, but are distributed throughout the genome (Berná et al., 2018). In **Figure 3A**, we present the landscape of chromosomes from the perspective of compartmentalization. All chromosomes display the characteristic core and disruptive compartment disposition. To calculate the percentage of core/disruptive

regions per chromosome, considering GC content, we utilized the disruptomics application (Balouz et al., 2025). This allowed us to categorize the chromosomes as core (over 60% of genes classified as core), disruptive (over 60% of genes classified as disruptive), and mixed (no category exceeds 60%) (**Table 2**).

Table 2. Chromosome classification by structure. Chromosome name, length in bases, core, disruptive, GC percentages, and chromosome classification were included.

Chromosome	Length (bases)	Core Percentage	Disruptive Percentage	GC Percentage	Chromosome Classification
Chr 01	2492231	68,38	31,62	49,89	Core
Chr 02	2151424	77,86	22,14	49,01	Core
Chr 03	1802204	13,99	86,01	53,13	Disruptive
Chr 04	1735261	29,17	70,83	52,02	Disruptive
Chr 05	1610701	77,20	22,80	49,02	Core
Chr 06	1492875	34,07	65,93	52,02	Disruptive
Chr 07	1490547	27,31	72,69	52,70	Disruptive
Chr 08	1436163	63,69	36,31	49,97	Core
Chr 09	1394721	58,36	41,64	49,99	Mixed
Chr 10	1361090	31,69	68,31	51,26	Disruptive
Chr 11	1261511	59,67	40,33	50,04	Mixed
Chr 12	1250845	33,92	66,08	53,50	Disruptive
Chr 13	1200139	61,10	38,90	50,28	Core
Chr 14	1111135	41,50	58,50	51,91	Mixed
Chr 15	1100981	80,11	19,89	48,54	Core
Chr 16	1084488	25,26	74,74	52,74	Disruptive
Chr 17	1002814	41,19	58,81	51,78	Mixed
Chr 18	965874	63,91	36,09	50,26	Core
Chr 19	933234	80,62	19,38	48,87	Core
Chr 20	903723	37,92	62,08	51,26	Disruptive
Chr 21	883590	40,55	59,45	51,46	Mixed
Chr 22	849045	48,91	51,09	50,55	Mixed
Chr 23	843191	18,27	81,73	53,26	Disruptive
Chr 24	811364	41,06	58,94	51,77	Mixed
Chr 25	671853	31,70	68,30	52,89	Disruptive
Chr 26	663682	19,99	80,01	53,10	Disruptive
Chr 27	653362	48,21	51,79	51,29	Mixed
Chr 28	614808	55,64	44,36	50,90	Mixed
Chr 29	599670	41,27	58,73	51,65	Mixed
Chr 30	588781	17,35	82,65	54,24	Disruptive
Chr 31	585579	48,00	52,00	51,44	Mixed
Chr 32	584768	70,45	29,55	49,73	Core

Classification: %Core > 60% =Core. % Disrutpive> 60%=Disruptive. Otherwise=Mixed

The graphic representation clearly illustrates this correlation (**Figure 3A**), identifying 9 C (example in **Figure 3B**), 12 D (**Figure 3C),** and 11 mixed chromosomes. Notably, Dm25 chromosomes display an almost identical pattern (**Figures 3B and 3C -boxes-, Supplementary Figure 2)**. This finding suggests that not only is the number of chromosomes conserved in *T. cruzi* strains but also their architecture**,** which contrasts with the paradigm that genomic plasticity dynamically influences chromosome size. This correlation is maintained even in the disruptive chromosomes (**Figure 3C**), reinforcing the concept that the karyotype of *T. cruzi* is more stable than expected. The disruptive compartment represents an adaptive specialization for mammalian infection. It exhibits very low expression across all parasite stages, which is directly related to the significant chromatin compaction, except for specific genes encoding surface proteins that show elevated expression in trypomastigotes (Díaz-Viraqué et al., 2023). The high proportion of disruptive chromosomes constitutes a significant reservoir of spatially clustered genes. Their chromosomal organization ensures a coordinated spatial distribution, potentially facilitating expression-switching events. As these genes encode surface antigens, this organization promotes antigenic variability, ultimately enhancing the efficiency of immune evasion.

As mentioned, a remarkable correspondence between chromosomes was found when comparing Dm28c to Dm25 (**Figure 2A**). The corresponding chromosomes in both strains fall into the same classification: core chromosomes remain core, disruptive chromosomes remain disruptive, and mixed chromosomes remain mixed. Synteny is highly preserved in the core compartment (**Figure 2C**), whereas, in the disruptive compartment, some differences can be visualized while still conserving the overall architecture (**Figure 2D**). This can be easily visualized when the GC content is analyzed. A notable difference in the GC content of both compartments was previously found (Berná et al., 2018). When the chromosomes of both strains were analyzed for their GC content, the pattern obtained was almost identical (**Figure 3B and C**). Taken together, these results suggest that genome plasticity in *T. cruzi* is not primarily driven by chromosomal rearrangements. Instead, genome plasticity in *T. cruzi* will likely occur at the intrachromosomal level, which deserves further investigation.
The conservation of chromosome architecture suggests a stable genomic framework that may be functionally relevant across strains. This structural conservation implies evolutionary constraints that preserve gene order, regulatory landscapes, and chromatin organization despite the well-documented plasticity of the *T. cruzi* genome. Such stability provides a reliable foundation for comparative genomics, enabling the identification of lineage-specific rearrangements, and may reflect selective pressures to maintain core functions related to parasite survival, infectivity, and host interaction. Moreover, it highlights the potential of chromosome-level assemblies to refine our understanding of genome evolution and its impact on biological traits.

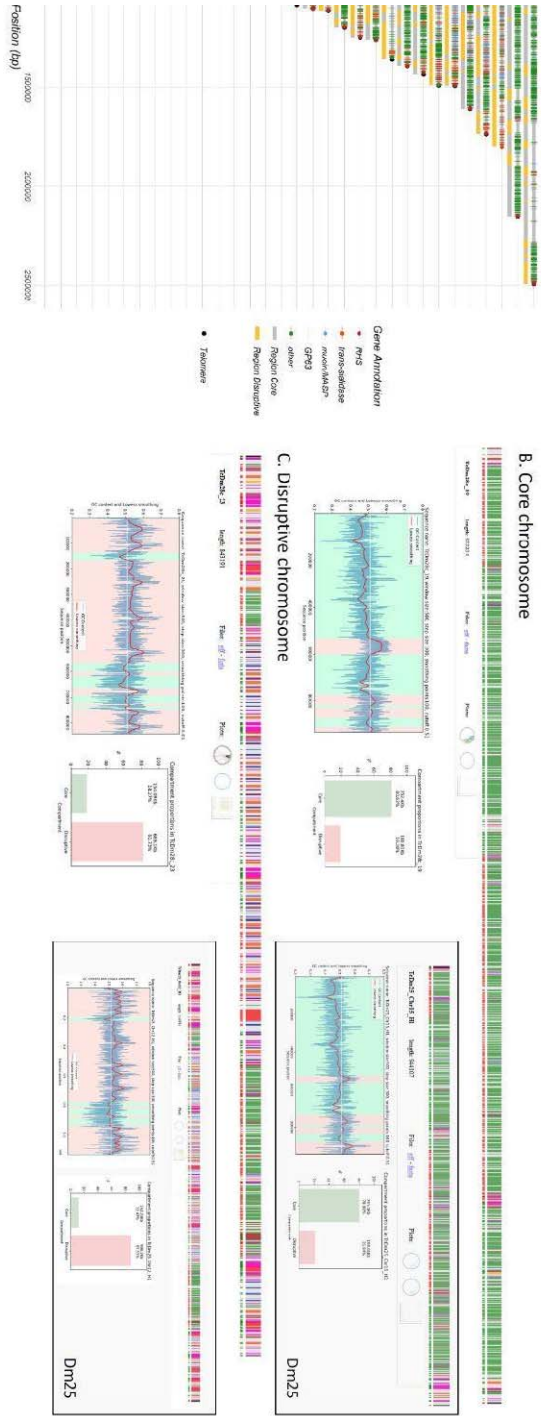

Figure 3. A. Representation e, Core
and disruptive regions are s sented
as black dots, and the f within
chromosomes, color-coded osome
predominantly classified as ization
into core and disruptive re scaglia,
2025). The graph on the righ ws the
same profile for the homol ng the
profile of a chromosome cla

5- The telomeres of *Trypanosoma cruzi*

Telomeres are specialized nucleoprotein structures located at the ends of eukaryotic chromosomes that protect genome integrity by maintaining chromosomal stability. In *T. cruzi*, we examined these chromosomal termini and confirmed the presence of the expected hexameric repeat (CCCTAA/TTAGGG), immediately followed by the 189bp junction (**Figure 4**). Analysis of telomeric sequences across all chromosomes reveals significant length variability, ranging from 44 bp (chromosome 13) to 13,299 bp (chromosome 10). Although long-read sequencing is not the gold standard for precise telomeric length quantification, this variability aligns with findings from Southern blot studies, which previously reported extensive telomeric length heterogeneity both within and between *T. cruzi* strains (Chiurillo et al., 1999; Freitas et al., 1999). Notably, in the *T. cruzi* Dm28c strain, telomeric lengths vary between 1 and 20 kb (Freitas et al., 1999). The telomeres of *T. cruzi* share compositional and length variability characteristics with those of *T. brucei* (Cano, 2001), which have been implicated in chromosome stability (Li et al., 2021). In *T. brucei*, telomere length can influence the expression of subtelomeric genes, including those involved in antigenic variation (Hovel-Miner et al., 2012). Although *T. cruzi* does not undergo antigenic switching due to the absence of *VSG* genes, its subtelomeric compartment (see below) harbors gene families that may play critical roles in host-parasite interactions. Investigating whether telomere length modulates their expression could provide insights into *T. cruzi* adaptability and virulence.

Overall, our findings underscore the pronounced heterogeneity in telomere length in *T. cruzi*, raising questions about its functional significance. Further research is required to elucidate the mechanisms governing telomere maintenance and their impact on genome stability and host-pathogen interactions.

6- The Subtelomeric Compartment

Subtelomeres are chromosomal regions adjacent to telomeres, typically characterized by low gene density and high genomic plasticity, which play a key role in variability and adaptation across diverse organisms. As previously noted, research on *T. cruzi* chromosomal ends has primarily relied on cloning and targeted analysis techniques. In this study, we aimed to characterize the 64 subtelomeres of *T. cruzi* to determine their features and composition. To achieve this, we employed complementary approaches. First, we assessed gene density in the first 50 Kb contiguous to each telomere (**Figure 4**). A global analysis of all chromosomes revealed a marked decrease in gene density in the regions adjacent to the 189 bp junction, allowing us to define the putative subtelomeric domains. We then manually examined the gene composition of these regions and found that RHS was the most frequent gene adjacent to the 189 bp junction, occurring in over 80% of cases (**Figures 4A, B, and C**).

Additionally, RHS, TS, and DGF-1 genes or pseudogenes were significantly enriched in these regions (**Figure 5A**). Based on this, we evaluated gene composition in sliding windows of five genes across entire chromosomes, quantitatively confirming that these three gene families were the most frequently identified at chromosomal ends beyond the 189 bp junction, followed by different core genes and retroelements. These patterns of gene density and composition were consistent among core, mixed, and disruptive chromosomes. **Figures 5B** and **5C** illustrate two representative cases: chromosome 19, classified as core, and chromosome 23, classified as disruptive. It is noteworthy that all genes within the subtelomeric regions are oriented toward the chromosomal ends.

The inner boundary of the subtelomere was arbitrarily defined as the first nucleotide of the coding sequence (CDS) of the most internal gene within the low gene density region. This operational definition would provide a consistent reference point for future analyses.

The specific subtelomeres previously reported and analyzed through cloning and sequencing strategies showed the enrichment of RHS and TS genes (Chiurillo et al., 1999; Freitas et al., 1999; Kim et al., 2005). Our results indicate that those findings, which focused on a limited number of subtelomeres, can now be expanded to include all 64 subtelomeres.

One point worth highlighting is that, although the gene composition of subtelomeres resembles that of the disruptive compartment—due to the enrichment of multigene families—no MASP or mucin genes were found in any subtelomere, except for two highly disruptive chromosomes with a high abundance of these genes. This observation supports the view that subtelomeres constitute a distinct third compartment of the *T. cruzi* genome (**Figure 6**). This finding underscores the need for further investigation into the dynamics of subtelomeric regions and their functional roles, particularly in genome stability, adaptive processes, and regulatory mechanisms.

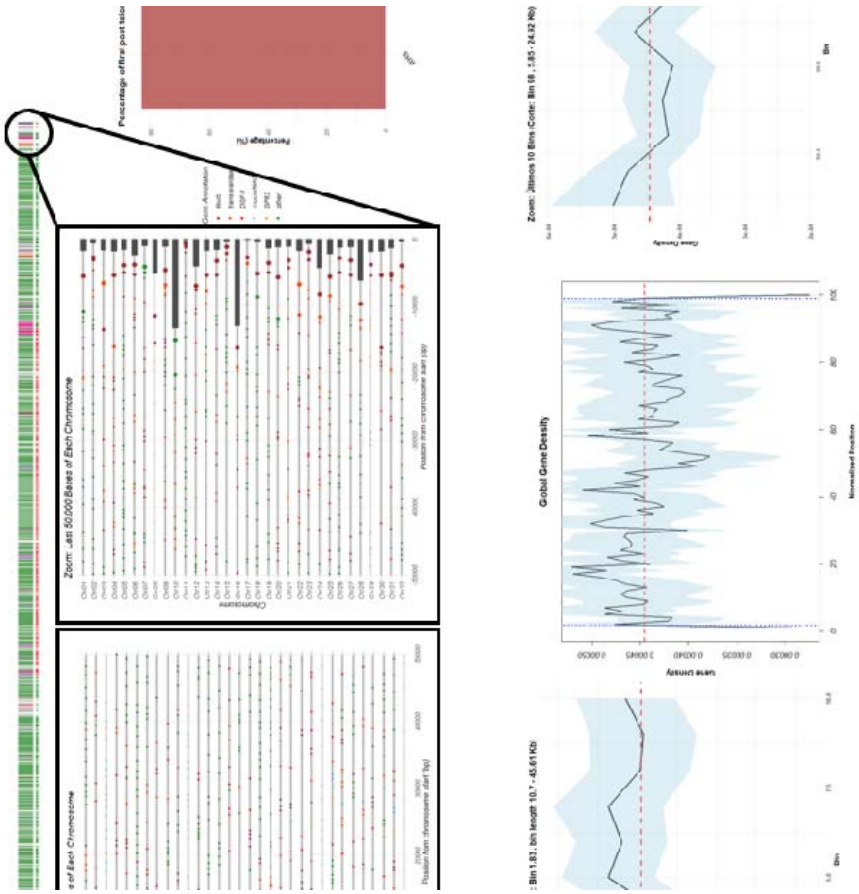

Figure 4. Telomeres and subtelomeres in *T. cruzi*. A. Representation of chromosomes with a zoom-in on the first and last 50 Kb. Dots indicate genes, and dark gray rectangles represent telomeric sequences. **B.** Data aggregated from all 32 chromosomes, with lengths normalized into 100 bins. Gene density is shown for the first 10 bins (left plot), the entire chromosome (center plot), and the last 10 bins (right plot). The red dotted line represents the mean gene density, while the blue line marks the crossing points at the beginning and end. **C.** Proportion of RHS, trans-sialidase, mucin/MASP, or conserved genes detected as first gene post-telomere sequence.

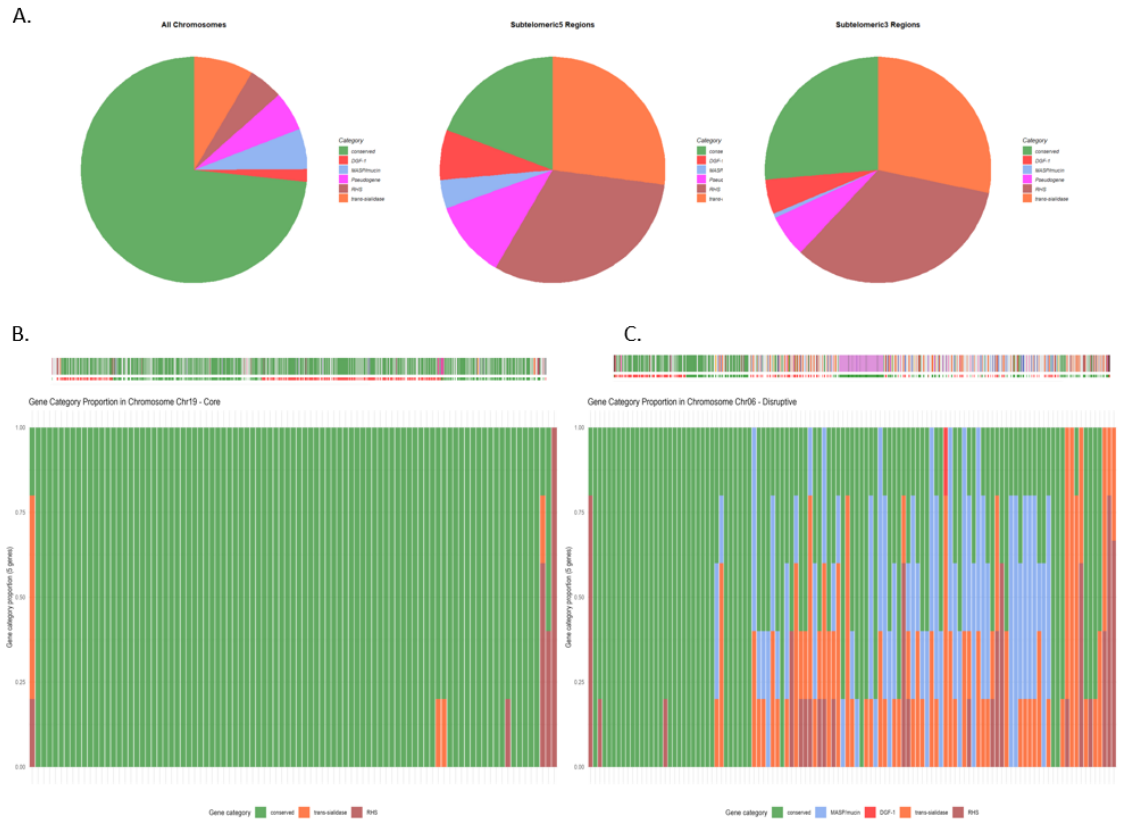

Figure 5. Gene density in core and disruptive chromosomes. Categories of genes (non-overlapping five-genes windows) were obtained. **A.** Proportions of gene classes in all chromosomes (left side), in subtelomeric 5' region (center) or subtelomeric 3' region (rigth). **B.** Core chromosome example. **C.** Disruptive chromosome.

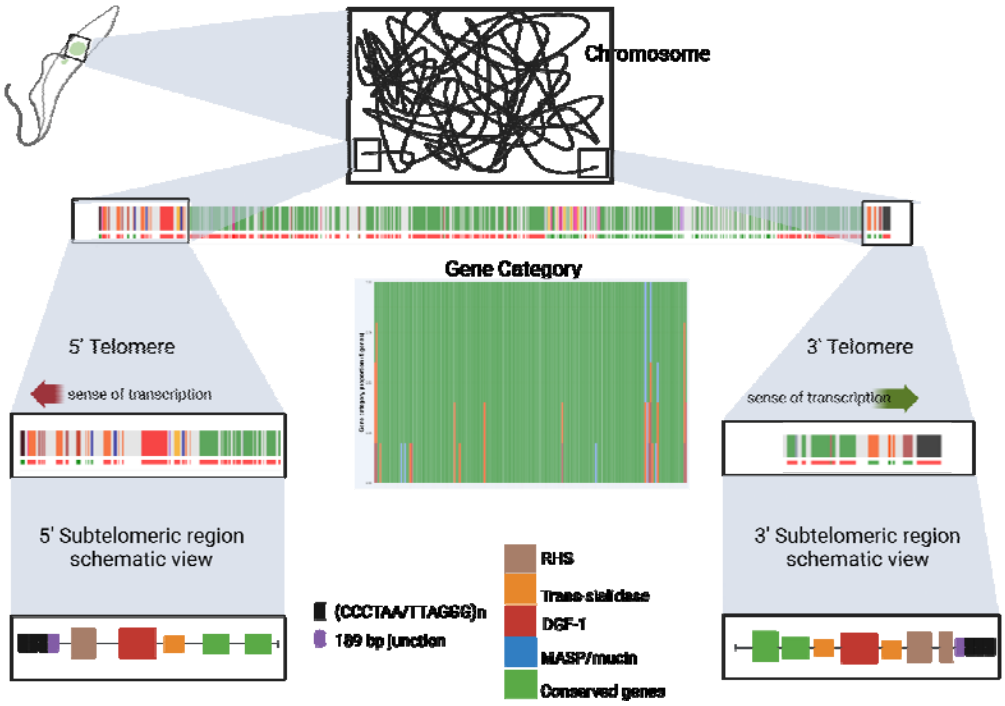

Figure 6. Subtelomeric compartment. Schematic representation of *T. cruzi* chromosome with a zoom in telomeres and subtelomeric region.

Materials and Methods

Genome Visualization and Web Interface

In previous work, we created a web interface for visualizing the genomes of Dm28c_2018 and TCC (http://bioinformatica.fcien.edu.uy/cruzi). To enhance data visualization and usability, we utilized that model again with minor modifications. The new Dm28cT2T genome assembly and annotation are accessible through https://cruzi.pasteur.uy/. As in the previous version, the direction of the polycistrons can be found underneath each represented chromosome. From each chromosome, the circos and Yass plots can be displayed, and the fasta and gff files can be downloaded (**Supplementary Figure 3**). As before, the additional functionalities were preserved, such as viewing gene annotations, retrieving nucleotide and amino acid sequences, and conducting various searches by annotation or keywords.

DNA extraction and Sequencing.

Cryopreserved epimastigotes of *T. cruzi* Dm28c_2018 (Berná et al., 2018), were grown in liver infusion tryptose (LIT) medium supplemented with 10% fetal bovine serum at 28° C; total DNA was extracted using the Quick DNA Universal kit and Quick DNA HMW MagBeads Kit (Zymo Research, USA), quantified with Qubit (Invitrogen, USA). Five µg of each extraction were pooled and sent to the Macrogen service to perform PacBio Hi-Fi sequencing. 301,047 reads were obtained (Read N50=10688), representing 2,926,558,665 bases (73X coverage for an estimated 40 Mb genome) with an average read quality of Q29.

Genome Assembly

The default HiFiAsm pipeline (16) (hifiasm -o Prefix -t 32 Reads.fq.gz) generated an initial assembly of 51 contigs and two haplotypes with 80 and 86 contigs, respectively. In this first step, 25 telomere-to-telomere chromosomes were recovered (18 of them were also T2T reconstructed in previous *T. cruzi* Dm25 HiFi genome assembly (Wang et al., 2021; Sanchez et al., 2024). Contigs with telomeres at 5' or 3' ends were manually assembled using previously assembled T2T chromosomes from the Dm25 strain as a reference. Terminal telomeric signals (CCCTAA/TTAGGG) were detected in all the chromosomes assembled, indicating T2T assembled chromosomes. Chromosomes were sorted and numbered in order of decreasing length and designated as chromosomes 1 to 32. Their sizes range from 2.5 Mbp (TcDm28c_01) to 0.6 Mbp (TcDm28c_32). Consistent with previous reports, the maxicircle (TcDm28c_aMaxicircle) has a length of 50,477 bp (bases) and belongs to the type "a" and was recovered in one contig. Finally, 47 contigs of the first HiFiAsm output were used to reconstruct the 32 complete telomere-to-telomere chromosomes.

Genome Annotation.

To annotate the coding sequences, we first trained Augustus (Stanke and Morgenstern, 2005; Stanke et al., 2006) using T. cruzi (previous annotation of the Dm28c genome), to produce the initial annotation GFF file. We then corrected the start and stop codons of each CDS utilizing the getorf method as previously described (Rice et al., 2000). All CDS were retrieved and annotated using MMseqs (Mirdita et al.,

2019) against the UniProt database (uniref_100). The coding sequences were categorized into three possible outcomes: no similarity to the UniProt database, similarity to the UniProt database, or truncated/multiple possible annotations (**Supplementary Figure 1A**). Busco (Manni et al., 2021) was used to evaluate the completeness of the gene space against the Euglenozoa database, indicating the quality of the assembly (**Supplementary Figure 1B**). The complete pipeline is summarized in **Supplementary Figure 1A.**

The web page with the new genome and its annotation was constructed based on a previous web page published by our group (Berná et al., 2018) and available at https://www.cruzi.pasteur.uy.

Tandem repeats and satellite sequences were annotated with the TRF algorithm (Benson, 1999). Satellite regions were retrieved from TRF output by filtering the length of repeats between 190 and 200 bp. Additionally, we tested our annotation by comparing it with the algorithm developed by Dean et al. (2025) for classifying MASP genes, obtaining equivalent results.

Bioinformatics.

All scripts used to produce data and figures were available on github (www.github.com/Gon1976/TcruziGenome). R scripts were used to estimate gene density along the chromosomes. Gene density was computed using the countOverlaps function from the GenomicRanges package in R (Lawrence et al., 2013). This analysis was performed for all chromosomes. Circos (Krzywinski et al., 2009) were used to create Figure 2A, using data from minimap2 comparison with Dm28c and Dm25 genomes. SyMap (Soderlund et al., 2011) was used for synteny mapping and circos in Figure 2B, C and D. Circlize package (Gu et al., 2014) in R was used to create Figure 1, including information on GC content (calculated using BioString package (Lawrence et al., 2013) and gene density. R ggplot package was used to create gene representation in the subtelomeres regions using gff data.

References

Alvarez VE, Niemirowicz GT, Cazzulo JJ. The peptidases of Trypanosoma cruzi: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. Biochim Biophys Acta. 2012 Jan;1824(1):195-206. doi: 10.1016/j.bbapap.2011.05.011.

Balouz, V., and C. Buscaglia. 2025. GCanner, a genome-wide GC composition tool for the unbiased assessment of *T. cruzi* genomic compartments. *Methods In Molecular Biology*, submitted.

Barros, R.R.M., M.M. Marini, C.R. Antônio, D.R. Cortez, A.M. Miyake, F.M. Lima, J.C. Ruiz, D.C. Bartholomeu, M.A. Chiurillo, J.L. Ramirez, and J.F. da Silveira. 2012. Anatomy and evolution of telomeric and subtelomeric regions in the human protozoan parasite Trypanosoma cruzi. *BMC Genom.* 13:229. doi:10.1186/1471-2164-13-229.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580. doi:10.1093/nar/27.2.573.

Berná, L., M.L. Chiribao, G. Greif, M. Rodriguez, F. Alvarez-Valin, and C. Robello. 2017. Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in Trypanosoma cruzi. *PeerJ.* 5:e3017. doi:10.7717/peerj.3017.

Berná, L., M. Rodriguez, M.L. Chiribao, A. Parodi-Talice, S. Pita, G. Rijo, F. Alvarez-Valin, and C. Robello. 2018. Expanding an expanded genome: long-read sequencing of Trypanosoma cruzi. *Microb. Genom.* 4:e000177. doi:10.1099/mgen.0.000177.

Berná, L., G. Greif, S. Pita, P. Faral-Tello, F. Díaz-Viraqué, R.D.C.M.D. Souza, G.A. Vallejo, F. Alvarez-Valin, and C. Robello. 2021. Maxicircle architecture and evolutionary insights into Trypanosoma cruzi complex. *PLoS Neglected Trop. Dis.* 15:e0009719. doi:10.1371/journal.pntd.0009719.

Bernardo, W.P., R.T. Souza, A.G. Costa-Martins, E.R. Ferreira, R.A. Mortara, M.M.G. Teixeira, J.L. Ramirez, and J.F.D. Silveira. 2020. Genomic Organization and Generation of Genetic Variability in the RHS (Retrotransposon Hot Spot) Protein Multigene Family in Trypanosoma cruzi. *Genes*. 11:1085. doi:10.3390/genes11091085.

Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, et al. 2005. The Genome of the African Trypanosome Trypanosoma brucei. *Science*. 309:416–422. doi:10.1126/science.1112642.

Bringaud, F., N. Biteau, S.E. Melville, S. Hez, N.M. El-Sayed, V. Leech, M. Berriman, N. Hall, J.E. Donelson, and T. Baltz. 2002. A New, Expressed Multigene Family Containing a Hot Spot for Insertion of Retroelements Is Associated with Polymorphic Subtelomeric Regions of Trypanosoma brucei. *Eukaryot. Cell*. 1:137–151. doi:10.1128/ec.1.1.137-151.2002.

Britto, C., C. Ravel, P. Bastien, C. Blaineau, M. Pagès, J.-P. Dedet, and P. Wincker. 1998. Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World Leishmania genomes. *Gene*. 222:107–117. doi:10.1016/s0378-1119(98)00472-7.

Buscaglia CA, Alfonso J, Campetella O, Frasch AC. 1999. Tandem amino acid repeats from Trypanosoma cruzi shed antigens increase the half-life of proteins in blood. Blood. Mar 15;93(6):2025-32.

Buscaglia CA, Campo VA, Frasch ACC, Di Noia JM. 2006. Trypanosoma cruzi surface mucins: host-dependent coat diversity. Nat Rev Microbiol. 2 4(3):229-236. doi: 10.1038/nrmicro1351.

Callejas-Hernández, F., A. Rastrojo, C. Poveda, N. Gironès, and M. Fresno. 2018. Genomic assemblies of newly sequenced Trypanosoma cruzi strains reveal new genomic expansion and greater complexity. *Scientific Reports*. 8:14631. doi:10.1038/s41598-018-32877-2.

Cano, M.I.N. 2001. Telomere biology of Trypanosomatids: more questions than answers. *Trends Parasitol.* 17:425–429. doi:10.1016/s1471-4922(01)02014-1.

Cazzulo JJ, Frasch AC. 1992. SAPA/trans-sialidase and cruzipain: two antigens from Trypanosoma cruzi contain immunodominant but enzymatically inactive domains. FASEB J. Nov;6(14):3259-64. PMID: 1426764.

Chiurillo, M.A., I. Cano, J.F.D. Silveira, and J.L. Ramirez. 1999. Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite Trypanosoma cruzi. *Mol. Biochem. Parasitol.* 100:173–183. doi:10.1016/s0166-6851(99)00047-x.

Chiurillo, M.A., M.R.M. Santos, J.F.D. Silveira, and J.L. Ramírez. 2002. An improved general approach for cloning and characterizing telomeres: the protozoan parasite Trypanosoma cruzi as a model organism. *Gene*. 294:197–204. doi:10.1016/s0378-1119(02)00768-0.

Contreras, V.T., T.C. de Araújo-Jorge, M.C. Bonaldo, N. Thomaz, H.S. Barbosa, M. de N.S.L. de Meirelles, and S. Goldenberg. 1988. Biological aspects of the DM28C clone of Trypanosoma cruzi after metacylogenesis in chemically defined media. *Memórias Inst. Oswaldo Cruz*. 83:123–133. doi:10.1590/s0074-02761988000100016.

Cruz-Saavedra L, Schwabl P, Vallejo GA, Carranza JC, Muñoz M, Patino LH, Paniz-Mondolfi A, Llewellyn MS, Ramírez JD. 2022. Genome plasticity driven by aneuploidy and loss of heterozygosity in *Trypanosoma cruzi*. Microb Genom. Jun;8(6):mgen000843. doi: 10.1099/mgen.0.000843. Erratum in: Microb Genom. 2022 Jul;8(7). doi: 10.1099/mgen.0.000873.

De Pablos LM, Osuna A. 2012. Conserved regions as markers of different patterns of expression and distribution of the mucin-associated surface proteins of Trypanosoma cruzi. Infect Immun. 80(1):169-74. doi: 10.1128/IAI.05859-11.

Dean, A.A.C., Berná, L., Robello, C., Buscaglia, C.A., Balouz, V. 2025. An Algorithm for annotation and classification of *T. cruzi* MASP sequences: towards a better understanding of the parasite genetic variability. *BMC Genomics*. Feb 24;26(1):194.doi: 10.1186/s12864-025-11384-5.

Díaz-Viraqué, F., S. Pita, G. Greif, R. de C.M. de Souza, G. Iraola, and C. Robello. 2019. Nanopore sequencing significantly improves genome assembly of the protozoan parasite Trypanosoma cruzi. *Genome Biol. Evol.* 11:1952–1957. doi:10.1093/gbe/evz129.

Díaz-Viraqué, F., M.L. Chiribao, M.G. Libisch, and C. Robello. 2023. Genome-wide chromatin interaction map for Trypanosoma cruzi. *Nat. Microbiol.* 8:2103–2114. doi:10.1038/s41564-023-01483-y.

El-Sayed, N.M., P.J. Myler, D.C. Bartholomeu, D. Nilsson, G. Aggarwal, et al. 2005a. The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. 309:409–415. doi:10.1126/science.1112631.

El-Sayed, N.M., P.J. Myler, G. Blandin, M. Berriman, J. Crabtree, et al. 2005b. Comparative genomics of trypanosomatid parasitic protozoa. 309:404–409. doi:10.1126/science.1112181.

Frasch, A.C.C. 2000. Functional Diversity in the Trans-sialidase and Mucin Families in Trypanosoma cruzi. *Parasitol. Today*. 16:282–286. doi:10.1016/s0169-4758(00)01698-7.

Freire-de-Lima L, Fonseca LM, Oeltmann T, Mendonça-Previato L, Previato JO. 2015. The trans-sialidase, the major Trypanosoma cruzi virulence factor: Three decades of studies. Glycobiology. 25(11):1142-9. doi: 10.1093/glycob/cwv057. Epub 2015 Jul 29. PMID: 26224786.

Freitas, L.H.G., R.M. Porto, L.A. Pirrit, S. Schenkman, and A. Scherf. 1999. Identification of the telomere in Trypanosoma cruzi reveals highly heterogeneous telomere lengths in different parasite strains. *Nucleic Acids Res.* 27:2451–2456. doi:10.1093/nar/27.12.2451.

Freitas LM, dos Santos SL, Rodrigues-Luiz GF, Mendes TA, Rodrigues TS, Gazzinelli RT, Teixeira SM, Fujiwara RT, Bartholomeu DC. 2011. Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of Trypanosoma cruzi reveal an undetected level of complexity. PLoS One. 6(10):e25914. doi: 10.1371/journal.pone.0025914. Epub 2011 Oct 19.

Galindo, M., V. Sabaj, I. Espinoza, G.C. Toro, J. Búa, J. Grenet, R. López-Solís, A.M. Ruiz, and N. Galanti. 2007. Chromosomal size conservation through the cell cycle supports karyotype stability in Trypanosoma cruzi. *FEBS Lett.* 581:2022–2026. doi:10.1016/j.febslet.2007.03.093.

Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors. 2014. circlize implements and enhances circular visualization in R. *Bioinformatics*. 30:2811–2812. doi:10.1093/bioinformatics/btu393.

Hakim, J.M.C., S.A.G. Guarnizo, E.M. Machaca, R.H. Gilman, and M.R. Mugnier. 2024. Whole-genome assembly of a hybrid Trypanosoma cruzi strain assembled with Nanopore sequencing alone. *G3: Genes, Genomes, Genet.* 14:jkae076. doi:10.1093/g3journal/jkae076.

Henriksson, J., L. Åslund, R.A. Macina, B.M.F. de Cazzulo, J.J. Cazzulo, A.C.C. Frasch, and U. Pettersson. 1990. Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in Trypanosoma cruzi. *Mol. Biochem. Parasitol.* 42:213–223. doi:10.1016/0166-6851(90)90164-h.

Henriksson, J., L. Åslund, and U. Pettersson. 1996. Karyotype variability in Trypanosoma cruzi. *Parasitol. Today*. 12:108–114. doi:10.1016/0169-4758(96)80670-3.

Henriksson, J., B. Porcel, M. Rydåker, A. Ruiz, V. Sabaj, N. Galanti, J.J. Cazzulo, A.C.C. Frasch, and U. Pettersson. 1995. Chromosome specific markers reveal conserved linkage groups in spite of extensive chromosomal size variation in Trypanosoma cruzi. *Mol. Biochem. Parasitol.* 73:63–74. doi:10.1016/0166-6851(95)00096-j.

Hovel-Miner GA, Boothroyd CE, Mugnier M, Dreesen O, Cross GA, Papavasiliou FN. 2012. Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. PLoS Pathog. 8(8):e1002900. doi: 10.1371/journal.ppat.1002900.

Ivens, A.C., C.S. Peacock, E.A. Worthey, L. Murphy, G. Aggarwal, et al. 2005. The Genome of the Kinetoplastid Parasite, Leishmania major. *Science*. 309:436–442. doi:10.1126/science.1112680.

Kim, D., M.A. Chiurillo, N. El-Sayed, K. Jones, M.R.M. Santos, P.E. Porcile, B. Andersson, P. Myler, J.F. da Silveira, and J.L. Ramírez. 2005. Telomere and subtelomere of Trypanosoma cruzi chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of T. cruzi telomeres. *Gene*. 346:153–161. doi:10.1016/j.gene.2004.10.014.

Krzywinski, M., J. Schein, İ. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645. doi:10.1101/gr.092759.109.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M.T. Morgan, and V.J. Carey. 2013. Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* 9:e1003118. doi:10.1371/journal.pcbi.1003118.

Leão AC, Viana LA, Fortes de Araujo F, de Lourdes Almeida R, Freitas LM, Coqueiro-Dos-Santos A, da Silveira-Lemos D, Cardoso MS, Reis-Cunha JL, Teixeira-Carvalho A, Bartholomeu DC. Antigenic diversity of MASP gene family of Trypanosoma cruzi. Microbes Infect. 2022 Sep;24(6-7):104982. doi: 10.1016/j.micinf.2022.104982. Epub 2022 Apr 26. PMID: 35487471.

Li B. 2021. Keeping Balance Between Genetic Stability and Plasticity at the Telomere and Subtelomere of Trypanosoma brucei. Front Cell Dev Biol. 9:699639. doi: 10.3389/fcell.2021.699639. PMID: 34291053; PMCID: PMC8287324.

Lima, A.R.J., H.G. de S. Silva, S. Poubel, J.N. Rosón, L.P.O. de Lima, H.M. Costa-Silva, C.S. Gonçalves, P.A.F. Galante, F. Holetz, M.C.M. Motta, A.M. Silber, M.C. Elias, and J.P.C. da Cunha. 2022. Open chromatin analysis in Trypanosoma cruzi life forms highlights critical differences in genomic compartments and developmental regulation at tDNA loci. *Epigenetics Chromatin*. 15:22. doi:10.1186/s13072-022-00450-x.

Manni, M., M.R. Berkeley, M. Seppey, and E.M. Zdobnov. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* 1:e323. doi:10.1002/cpz1.323.

Melville, S.E., V. Leech, C.S. Gerrard, A. Tait, and J.M. Blackwell. 1998. The molecular karyotype of the megabase chromosomes of Trypanosoma brucei and the assignment of chromosome markers. *Mol. Biochem. Parasitol.* 94:155–173. doi:10.1016/s0166-6851(98)00054-1.

Mirdita, M., M. Steinegger, and J. Söding. 2019. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*. 35:2856–2858. doi:10.1093/bioinformatics/bty1057.

Parodi-Talice, A., R. Durán, N. Arrambide, V. Prieto, M.D. Piñeyro, O. Pritsch, A. Cayota, C. Cerveñansky, and C. Robello. 2004. Proteome analysis of the causative agent of Chagas disease: Trypanosoma cruzi. *Int. J. Parasitol.* 34:881–886. doi:10.1016/j.ijpara.2004.05.002.

Parodi-Talice, A., V. Monteiro-Goes, N. Arrambide, A.R. Avila, R. Duran, A. Correa, B. Dallagiovanna, A. Cayota, M. Krieger, S. Goldenberg, and C. Robello. 2007. Proteomic analysis of metacyclic trypomastigotes undergoing Trypanosoma cruzi metacyclogenesis. *J. Mass Spectrom.* 42:1422–1432. doi:10.1002/jms.1267.

Pita, S., Díaz-Viraqué, F., Iraola, G., & Robello, C. 2019. The Tritryps comparative repeatome: insights on repetitive element evolution in Trypanosomatid pathogens. Genome biology and evolution, 11(2), 546-551. doi: 10.1093/gbe/evz017.

Ramirez JL. An Evolutionary View of Trypanosoma cruzi Telomeres. 2020. Front Cell Infect Microbiol. 9:439. doi: 10.3389/fcimb.2019.00439.

Ramírez JL. 2023. The Elusive *Trypanosoma cruzi* Disperse Gene Protein Family (DGF-1). Pathogens. 12(2):292. doi: 10.3390/pathogens12020292. PMID: 36839564; PMCID: PMC9967923.

Reis-Cunha JL, Valdivia HO, Bartholomeu DC. 2018. Gene and Chromosomal Copy Number Variations as an Adaptive Mechanism Towards a Parasitic Lifestyle in Trypanosomatids. Curr Genomics. 19(2):87-97. doi: 10.2174/1389202918666170911161311.

Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet*. 16:276–277. doi:10.1016/s0168-9525(00)02024-2.

Rosón, J.N., M. de O. Vitarelli, H.M. Costa-Silva, K.S. Pereira, D. da S. Pires, L. de S. Lopes, B. Cordeiro, A.J. Kraus, K.N.T. Cruz, S.G. Calderano, S.P. Fragoso, T.N. Siegel, M.C. Elias, and J.P.C. da Cunha. 2022. H2B.V demarcates divergent strand-switch regions, some tDNA loci, and genome compartments in Trypanosoma cruzi and affects parasite differentiation and host cell invasion. *PLoS Pathog.* 18:e1009694. doi:10.1371/journal.ppat.1009694.

Rubin SS, Schenkman S. 2012. Trypanosoma cruzi trans-sialidase as a multifunctional enzyme in Chagas' disease. Cell Microbiol.14(10):1522-30. doi: 10.1111/j.1462-5822.2012.01831.x. Epub 2012 Jul 22. PMID: 22747789.

Sanchez, M.C.H., H.S.O. Zapata, B.D. Suarez, C. Ospina, H.J. Barbosa, J.C.C. Martinez, G.A. Vallejo, D.U. Montes, and J. Duitama. 2024. A phased genome assembly of a Colombian Trypanosoma cruzi TcI strain and the evolution of gene families. *Sci. Rep.* 14:2054. doi:10.1038/s41598-024-52449-x.

Santos, M.R., M.I. Cano, A. Schijman, H. Lorenzi, M. Vázquez, M.J. Levin, J.L. Ramirez, A. Brandão, W.M. Degrave, and J.F. da Silveira. 1997. The Trypanosoma cruzi Genome Project: Nuclear Karyotype and Gene Mapping of Clone CL Brener. *Memórias Inst. Oswaldo Cruz*. 92:821–828. doi:10.1590/s0074-02761997000600018.

Schenkman S, Jiang MS, Hart GW, Nussenzweig V. 1991. A novel cell surface trans-sialidase of Trypanosoma cruzi generates a stage-specific epitope required for invasion of mammalian cells. Cell. 65(7):1117-25. doi: 10.1016/0092-8674(91)90008-m.

Soderlund, C., Bomhoff, M., Nelson, W.M. 2011. SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* Vol. 39,10, e68 .https://doi.org/10.1093/nar/gkr123.

Souza, R.T., F.M. Lima, R.M. Barros, D.R. Cortez, M.F. Santos, E.M. Cordero, J.C. Ruiz, S. Goldenberg, M.M.G. Teixeira, and J.F. da Silveira. 2011. Genome Size, Karyotype Polymorphism and Chromosomal Evolution in Trypanosoma cruzi. *PLoS ONE*. 6:e23042. doi:10.1371/journal.pone.0023042.

Stanke, M., and B. Morgenstern. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33:W465–W467. doi:10.1093/nar/gki458.

Stanke, M., O. Schöffmann, B. Morgenstern, and S. Waack. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* 7:62. doi:10.1186/1471-2105-7-62.

Vargas, N., A. Pedroso, and B. Zingales. 2004. Chromosomal polymorphism, gene synteny, and genome size in T. cruzi I and T. cruzi II groups. *Mol. Biochem. Parasitol.* 138:131–141. doi:10.1016/j.molbiopara.2004.08.005.

Vitarelli, M. de O., T.A. Franco, D. da S. Pires, A.R.J. Lima, V.L. Viala, A.J. Kraus, I. de L.M.J. de Azevedo, J.P.C. da Cunha, and M.C. Elias. 2024. Integrating high-throughput analysis to create an atlas of replication origins in Trypanosoma cruzi in the context of genome structure and variability. *mBio*. 15:e00319-24. doi:10.1128/mbio.00319-24.

Wang, W., D. Peng, R.P. Baptista, Y. Li, J.C. Kissinger, and R.L. Tarleton. 2021. Strain-specific genome evolution in Trypanosoma cruzi, the agent of Chagas disease. *PLoS Pathog.* 17:e1009254. doi:10.1371/journal.ppat.1009254.

Wincker P, Murto-Dovales AC, Goldenberg S. 1992. Nucleotide sequence of a representative member of a Trypanosoma cruzi dispersed gene family. Mol Biochem Parasitol. 55(1-2):217-20. doi: 10.1016/0166-6851(92)90142-7. PMID: 1435871.

Wincker P, Ravel C, Blaineau C, Pages M, Jauffret Y, Dedet JP, Bastien P. 1996. The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. Nucleic Acids Res. 24(9):1688-94. doi: 10.1093/nar/24.9.1688. PMID: 8649987; PMCID: PMC145848.

Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, Miles MA, Romanha AJ, Sturm NR, Tibayrenc M, Schijman AG; Second Satellite Meeting. A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz. 2009 Nov;104(7):1051-4. doi: 10.1590/s0074-02762009000700021.

Acknowledgments

We thank Matías Rodriguez (Institute of Bioinformatics, Faculty of Medicine, University of Münster, Münster, Germany) for facilitating the code for the webpage and for his significant assistance with the annotation scripts; L. Berna (I. Pasteur Montevideo, Uruguay) for providing the first version of the script used to detect telomeres; A. C. Dean, V. Balouz, and C. Buscaglia (Universidad Nacional de San Martín, Argentina) for sharing the GCScanner algorithm and the MASP annotation pipeline before publication; J.L. Ramirez (Instituto de Estudios Avanzados-IDEA and Universidad Central de Venezuela) for valuable discussions and suggestions regarding chromosomes, telomeres, and subtelomeres; and all members of Laboratorio de Interacciones Hospedero-Patógeno — UBM (Institut Pasteur de Montevideo) for their helpful comments and engaging discussions. This work was supported by the Research Council United Kingdom Grand Challenges Research Fund (GCRF) 'A Global Network for Neglected Tropical Diseases' MR/P027989/1 (CR); Agencia Nacional de Investigación e Innovación (ANII, Uruguay) DCI-ALA/2011/023–502, 'Contrato de apoyo a las políticas de innovación y cohesión territorial'; Fondo para la Convergencia Estructural del Mercado Común del Sur (FOCEM) 03/1 (C.R.); Pasteur Network ACIP 2024 n°1846 / Chromacruzi (FDV); and Agencia Nacional de Investigación e Innovación (ANII, Uruguay), FCE_3_2022_1_172653 (FDV). GG, FDV, MLC, CS, and CR are members of the Sistema Nacional de Investigadores (SNI-ANII, Uruguay) and the "Programa de Desarrollo de Ciencias Básicas" (PEDECIBA, Uruguay).