

## Article

# Assessing *Trypanosoma cruzi* Parasite Diversity through Comparative Genomics: Implications for Disease Epidemiology and Diagnostics

Alicia Majeau, Laura Murphy, Claudia Herrera \* and Eric Dumonteil \*

Department of Tropical Medicine, School of Public Health and Tropical Medicine, Vector Borne Infectious Disease Research Center, Tulane University, New Orleans, LA 70112, USA; amajeau@tulane.edu (A.M.); lmurphy8@tulane.edu (L.M.)

\* Correspondence: edumonte@tulane.edu (E.D.); cherrera@tulane.edu (C.H.)

**Abstract:** Chagas disease is an important vector-borne neglected tropical disease that causes great health and economic losses. The etiological agent, *Trypanosoma cruzi*, is a protozoan parasite endemic to the Americas, comprised by important diversity, which has been suggested to contribute to poor serological diagnostic performance. Current nomenclature describes seven discrete typing units (DTUs), or lineages. We performed the first large scale analysis of *T. cruzi* diversity among 52 previously published genomes from strains covering multiple countries and parasite DTUs and assessed how different markers summarize this genetic diversity. We also examined how seven antigens currently used in commercial serologic tests are conserved across this diversity of strains. DTU structuration was confirmed at the whole-genome level, with evidence of sub-DTU diversity, associated in part to geographic structuring. We observed very comparable phylogenetic tree topographies for most of the 32 markers investigated, with clear clustering of sequences by DTU, and a few of these markers suggested some degree of intra-lineage diversity. At least three of the currently used antigens represent poorly conserved sequences, with sequences used in tests quite divergent from sequences in many strains. Most markers are well suited for estimating parasite diversity to DTU level, and a few are particularly well-suited to assess intra-DTU diversity. Analysis of antigen sequences across all strains indicates that antigenic diversity is a likely explanation for limited diagnostic performance in Central and North America.

**Keywords:** *Trypanosoma cruzi*; Chagas disease; strain diversity; polymorphism; antigen

**Citation:** Majeau, A.; Murphy, L.; Herrera, C.; Dumonteil, E. Assessing *Trypanosoma cruzi* Parasite Diversity through Comparative Genomics: Implications for Disease Epidemiology and Diagnostics. *Pathogens* **2021**, *10*, 212. <https://doi.org/10.3390/pathogens10020212>

Academic Editor: Vyacheslav Yurchenko

Received: 14 January 2021

Accepted: 11 February 2021

Published: 16 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Trypanosoma cruzi*, the etiological agent of Chagas disease, is a protozoan parasite that infects a wide range of mammals. It is transmitted in the feces of insects of the sub-family Triatominae during blood-feeding, congenitally, or orally via contaminated food or drink. The parasite is endemic to and widely distributed throughout the Americas; it has been reported as far south as northern Chile [1] and Argentinian Patagonia [2], and as far north in the United States as Illinois [3]. Thirty to forty percent of individuals who are infected with the parasite go on to develop chronic disease, marked by cardiomyopathies or gastrointestinal disorders such as megaesophagus or megacolon. The highest disease burden lies in Latin America and estimates suggest that around six million people may be infected globally [4]. With approximately 70 million persons at risk, Chagas disease costs 806,170 disability-adjusted life years (DALYs) annually, as well as about \$627 million in patient care [5].

*Trypanosoma cruzi* presents an extensive genetic diversity [6], and based on limited sets of genetic markers, it is currently divided into seven major genetic lineages, or discrete typing units (DTUs), known as TcI through TcVI and the more recently described

TcBat [7,8], which remains poorly studied using molecular markers [9]. DTUs TcV and TcVI have originated from hybridization events between TcII and TcIII, and they inherited their kinetoplast (mitochondrial) maxicircle DNA from TcIII [10–12]. Accordingly, TcII, TcV, and TcVI are closely related and can be difficult to differentiate with some markers, such as 24S $\alpha$  [13]. DTU TcI has been the most extensively studied and it has been further subdivided into TcIa-e [14–16] based on sequences from the spliced-leader, or mini-exon. Indeed, a recent genome-wide analysis of TcI genomes demonstrated clear intra-DTU heterogeneity, with strong evidence of geographic clustering [17]. Genetic diversity has also been reported within TcII [18–21], and both TcIII and TcIV have been proposed to be genetically structured between North and South America [22,23].

Understanding both the full extent of the parasite's genetic diversity and the geographical range of the various lineages is important, as strain diversity may be associated with the variety of clinical manifestations and particular parasite transmission cycles [2,7,9,14,24–27]. Phylogeography also likely plays a role in the significant discordance among serological diagnostics in some regions. Indeed, most current commercial serological diagnostic tests are based on limited sets of antigens from South American strains [28,29], which limits their diagnostic performance in North and Central America [30–33], most likely due to a failure of the antigens to react with antibodies induced by genetically diverse strains of the parasite. For example, in a study comparing three commercial and two in-house enzyme-linked immunosorbent assays (ELISAs) in Mexico, even the most sensitive of the three commercial tests only detected half of the seropositive cases confirmed by Western blot [30]. Antigen conservation for vaccine development has also been an issue, as immune selection pressure drives antigenic variation of major antigen families such as the trans-sialidase [34,35], and strain specific immunity has been detected with some vaccine candidates from this family [36]. Other antigens such as 1F8/Tc24 may be more conserved and useful for vaccine or diagnostic development [37]. Furthermore, there is some evidence that certain strains may be associated with more severe outcomes of disease development [38] and drug susceptibility [39–41], and fully understanding the eco-epidemiology of various genetic lineages and the composition of parasite strains infecting patients [42] may allow for more precise estimates of risk to human health and patient prognosis.

The parasite is predominantly described as diploid, with recently reported haploid genomes of 55 Mb [43] and 79 Mb [44] in length from DTUs TcI and TcV, respectively. However, the genome is markedly plastic, with genome size and karyotypic variation between strains, irregular genetic exchange, and reports of widespread aneuploidy [45,46]. The first *T. cruzi* genome was sequenced in 2005 [47], but limited comparisons of *T. cruzi* genome sequences have been performed so far, as few additional genomes were available [48,49]. These initial comparisons confirmed that extensive portions of the genome comprise repetitive sequences, making assembly complicated [48–50]. A “core compartment” of the genome has been identified, comprising conserved and hypothetical conserved genes, together with a “non-syntenic disruptive compartment” comprising the major multigene families of trans-sialidases, mucin-associated surface proteins (MASPs) and mucins [48]. The genetic structure of the parasite follows a “predominant clonal evolution” pattern [51], described by significant linkage disequilibrium, overrepresented multilocus genotypes, near-clades, and Russian doll patterns corresponding to the DTU subdivisions [51,52] as well as lesser near-clades within each DTU [14,18,51]. However, evidence of sexual reproduction and recombination has recently been observed within TcI, which, together with high levels of inbreeding, maintains the apparent clonal structure [53,54] but may allow the parasite to generate strain diversity.

The recent sequencing of a growing number of parasite genomes from across the Americas now allows for extensive evaluation of *T. cruzi* genomic diversity and represents an unprecedented resource for diagnostic and vaccine development. In this study, we took advantage of these sequences to perform the first large-scale analysis of *T. cruzi* genomic diversity among countries and across parasite DTUs. This comparative ge-

nomics approach also allowed investigating how different markers summarize this genetic diversity and may be used for rapid genotyping of parasites in field studies. Finally, we also examined how some of the main antigens currently used in several commercial serologic tests are conserved across this diversity of strains, as this may affect diagnostic test performance across the Americas.

## 2. Results

### 2.1. Whole Genome Comparison

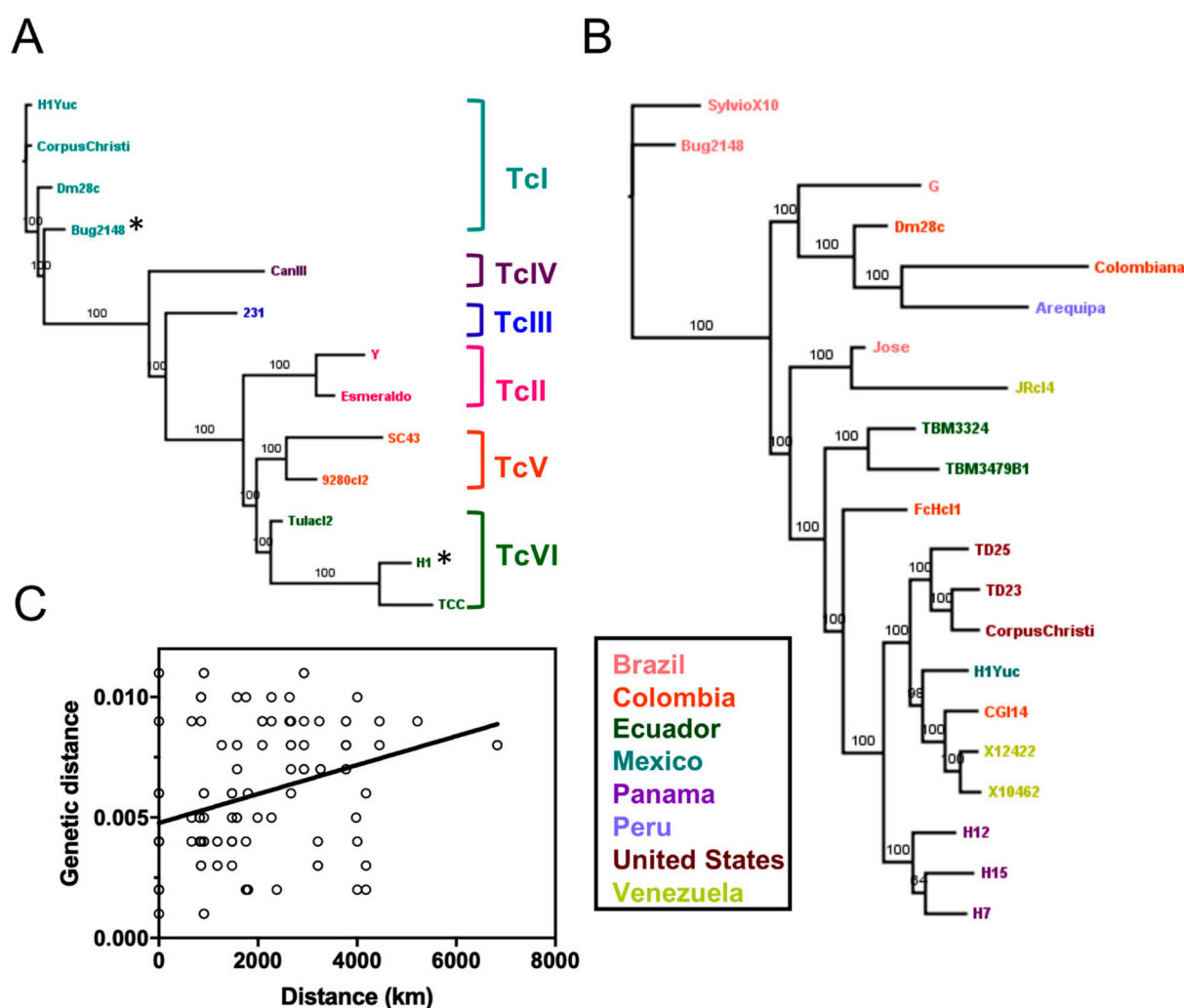
We used a data set of *T. cruzi* whole genome sequences from 52 strains (Table 1), covering a wide range of geographic origin (US, Mexico, Panama, Colombia, Chile, Venezuela, Ecuador, Bolivia, Brazil), DTUs (TcI to TcVI), and hosts (mammalian hosts and triatomine vectors). We performed full genome alignments and built maximum likelihood phylogenetic trees to assess the relationships among *T. cruzi* strains. Using a representative subset of 13 genomes covering six DTUs, we observed a clear clustering of strains according to their DTU (Figure 1A) confirming that *T. cruzi* DTU genetic structure is present at the whole genome level across all six DTUs. In addition, phylogenies based on 22 TcI genomes resulted in a strong genetic structuring, with some clustering according to the geographic origin of the strains, as previously reported by Talavera et al. [17], although not exclusively, as indicated by testing isolation by distance (Figure 1B,C, Mantel test,  $p = 0.0002$ ,  $R = 0.10$ ). For example, while TcI genome sequences from the US or Panama formed single clusters, those from Colombia or Venezuela could be found in very distinct clusters. Finally, whole genome analysis further indicated that the genome of “Bug2148” clearly belonged to TcI, and was more closely related to the Dm28c and SylvioX10 strains (Figure 1), while this strain was initially sequenced as TcV [43]. On the other hand, strain H1 from Panama was associated with the TcVI DTU cluster rather than to TcI as initially classified [17].

**Table 1.** *Trypanosoma cruzi* strains and genomes.

Strain	DTU	Accession No.	Host of Origin	Country
JRcl4	TcI	SRX290934, SRX180959, SRX180957, SRX031468, SRX031469	Human	Venezuela
Jose	TcI	SRR2057803	Human	Brazil
Colombiana	TcI	SRX871218	Human	Colombia
Arequipa	TcI	SRR1793866	Triatomine	Peru
Dm28c*	TcI	GCA_003177105.1	Opossum	Colombia
TBM3519W1	TcI	SRR3676270	Triatomine	Ecuador
TBM3324	TcI	SRR3676267	Triatomine	Ecuador
TBM3406B1	TcI	SRR3676268	Triatomine	Ecuador
TBM3479B1	TcI	SRR3676269	Triatomine	Ecuador
FcHcl1	TcI	SRR3676318	Human	Colombia
V1	TcI	SRR3676313	Triatomine	Panama
V2	TcI	SRR3676314	Triatomine	Panama
V3	TcI	SRR3676315	Triatomine	Panama
H1Yuc	TcI	SRX1851500	Human	Mexico
H1	TcVI**	SRR3676277	Human	Panama
H2	TcI	SRR3676278	Human	Panama
H3	TcI	SRR3676279	Human	Panama
H4	TcI	SRR3676280	Human	Panama
H5	TcI	SRR3676281	Human	Panama
H6	TcI	SRR3676282	Human	Panama
H7	TcI	SRR3676283	Human	Panama

H9	TcI	SRR3676285	Human	Panama
H12	TcI	SRR3676309	Human	Panama
H14	TcI	SRR3676310	Human	Panama
H15	TcI	SRR3676312	Human	Panama
TD23	TcI	SRR3676272	Triatomine	USA
TD25	TcI	SRR3676273	Triatomine	USA
Corpus Christi	TcI	SRX1054555	Human	USA
Bug2148*	TcI**	NMZN00000000	Human	Brazil
X12422	TcI	SRR3676275	Human	Venezuela
X10462	TcI	SRR3676274	Human	Venezuela
CGI14	TcI	SRX1851527	Human	Colombia
SylvioX10	TcI	ADWP00000000	Human	Brazil
G*	TcI	MKKV00000000	Opossum	Brazil
Esmeraldo	TcII	SRX022423, SRX022425, SRX022426, SRX022427, SRX022428, SRX022429, SRX271443, SRX271444, SRX2015243, SRX2015244	Human	Brazil
S11	TcII	SRX3453751	Human	Brazil
S15	TcII	SRX3453752	Human	Brazil
S154a	TcII	SRX3453753	Human	Brazil
S23b	TcII	SRX3453755	Human	Brazil
S1162a	TcII	SRX3453754	Human	Brazil
S92a	TcII	SRX3453757	Human	Brazil
S44a	TcII	SRX3453756	Human	Brazil
Y	TcII	SRR1796718, SRR1797819	Human	Brazil
Ycl4*	TcII	GCA_003594405.1	Human	Brazil
231*	TcIII	OGCJ00000000	Human	Brazil
CanIII	TcIV	SRR1996498, SRR1996501	Human	Brazil
9280 cl2	TcV	SRR1996492, SRR1996493, SRR1996496, SRR1996497, SRR1996502	Human	Bolivia
SC43*	TcV	GCA_015455285.1	Triatomine	Bolivia
CLBrenner*	TcVI	GCA_000209065.1	Triatomine (culture derivate)	Brazil
Tula cl2	TcVI	SRX268890, SRX268891, SRX268892, SRX268893, SRX268894, SRX268895, SRX268896	Human	Chile
TCC*	TcVI	GCA_003177095.1	Human (culture derivate)	Chile

\* Genome assemblies were used for these strains. \*\* Strains Bug2148 and H1 had been previously identified as belonging to TcV and TcI, respectively, but as shown in this study belong to TcI and TcVI, respectively.

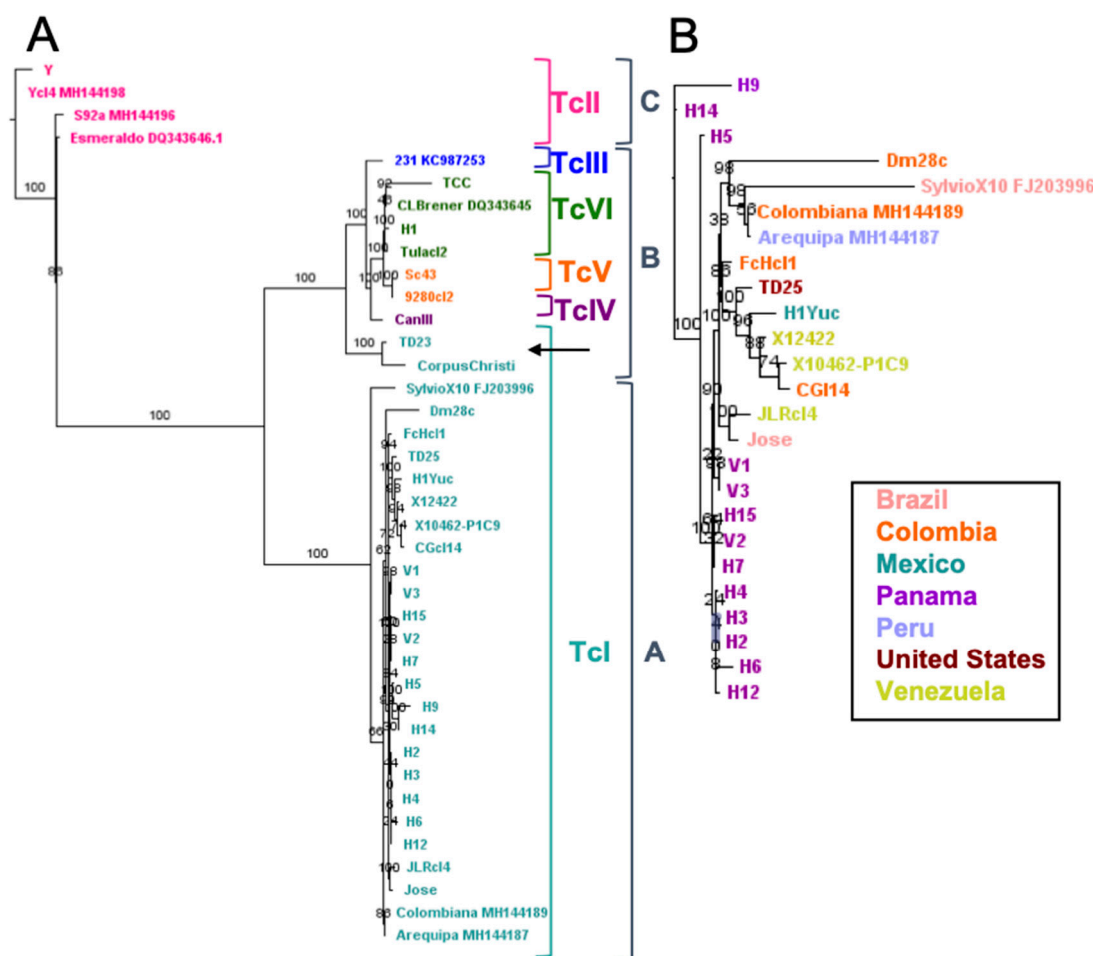


**Figure 1.** Whole genome analyses of strains representing all discrete typing units (DTUs). Phylogenetic trees constructed from whole genome alignments demonstrate a clear clustering by DTU (A) as well as strong sub-structuring within DTU TcI (B), with some geographic association (C) by Mantel test ( $p = 0.0002$ ,  $R = 0.10$ ). Asterisks indicate strains that were previously assigned to a different DTU.

Sequence analysis also allowed the assembly of full kinetoplast maxicircle sequences from a large number of parasite strains, for comparison with their nuclear genomes. Phylogenetic analysis of maxicircle sequences was in agreement with the previous identification of three major clades for kinetoplast maxicircle DNA [12], with clades A and C including TcI and TcII DTUs, respectively, and clade B the other DTUs (Figure 2A). However, a further genetic structuration was also detected within clade B. Indeed, TcV and TcVI hybrid strains segregated into separate clusters, distinct from both TcIII and TcIV DTUs.

Thus, genetic structuration according to parasite DTUs could also be observed at the level of the kinetoplast maxicircle. Furthermore, detailed analysis of TcI strains within maxicircle Clade A indicated a clear sub-structuring (Figure 2B), which mirrored in part that observed at the nuclear genome level (Figure 1B), strengthening the conclusion of the existence of genetic structuration within TcI. While there are not enough genomes of other DTUs to reach this same conclusion of genetic structuration within DTU for these lineages, there does appear to be intra-lineage diversity in DTU TcII and DTU TcVI in the maxicircle sequence analysis, again similar to what is observed at the nuclear genome level. Interestingly, analysis of the maxicircle sequences also revealed two clear cases of kinetoplast introgression for TcI strains from Texas (TD23 and Corpus Christi), kinetoplast sequences of which clustered within clade B (Figure 2A) but separate from DTUs

TcIII, TcIV, TcV and TcVI, while their nuclear genomes are closely related with that from TD25 strain (Figure 1B) which harbors a Clade A kinetoplast typical of TcI DTU (Figure 2A).



**Figure 2.** Analysis of kinetoplast maxicircle sequences. Three major clades were observed (A), in agreement with previous identification. Clades A and C include DTUs TcI and TcII respectively, and Clade B includes the other DTUs, with a structuration within this clade by DTU. Two clear cases of kinetoplast introgression (arrow) for TcI strains from Texas are present. Detailed analysis of TcI strains (B) indicated sub-structuring, similarly in part to the nuclear genome.

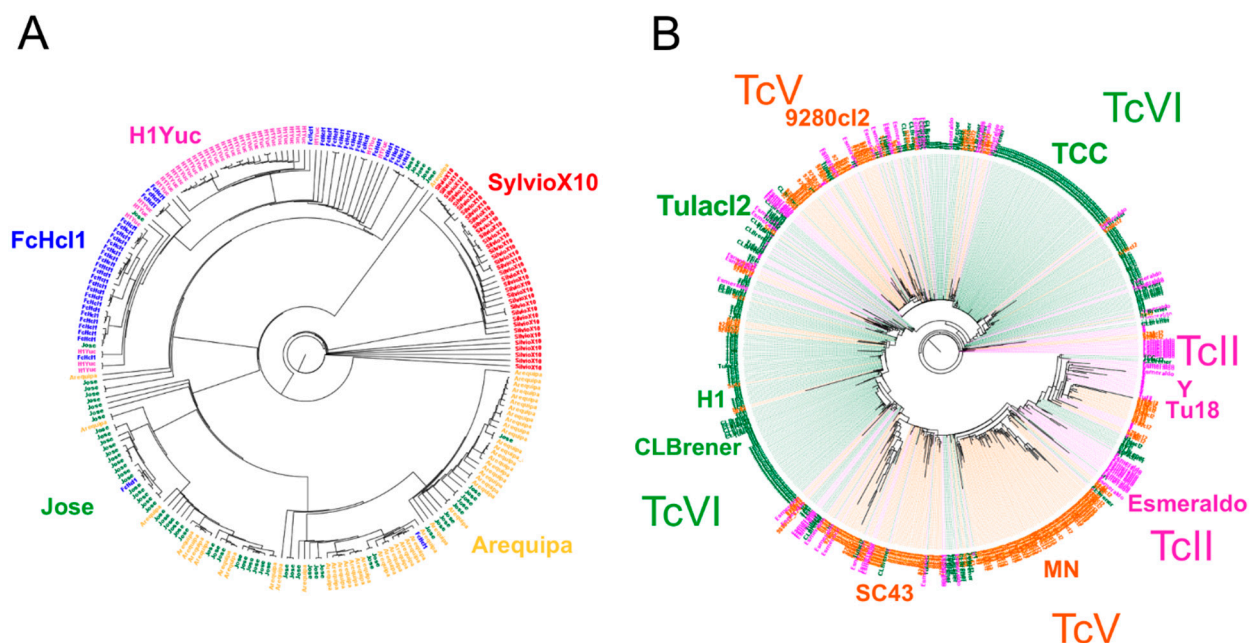
## 2.2. Phylogenetic Relationships Using Single and Multiple Markers

We then assessed the ability of some previously used markers to elucidate the parasite genomic diversity. Indeed, a large number of genetic markers have been proposed over the years to describe the extent of *T. cruzi* parasite diversity [2,7,55–58], including RNA genes, multiple single copy protein coding genes, as well as several genes for hypothetical conserved proteins. However, no extensive comparison of their ability to discriminate parasite strains has been performed. We identified nucleotide sequences corresponding to 30 individual markers (Supplementary Table S1) using basic local alignment search tool (BLAST) searches of the full genome sequence dataset and GenBank sequences. Individual markers varied in size from 118 to 2260 bp, and we constructed phylogenetic trees from each individual marker based on maximum likelihood (Supplementary Figure S1). We observed very comparable tree topographies for most markers, with clear clustering of sequences by DTU, except for TcV and TcVI, which often clustered with TcII and TcIII, in agreement with their hybrid nature. However, most markers including the 24S RNA, MSH2 or TcSD5D, which have been extensively used for genotyping [7,22,55,56], showed limited or no sequence variation within DTUs. A few markers such as the flagellum-adhesion glycoprotein or Beta-adaptin did, however,



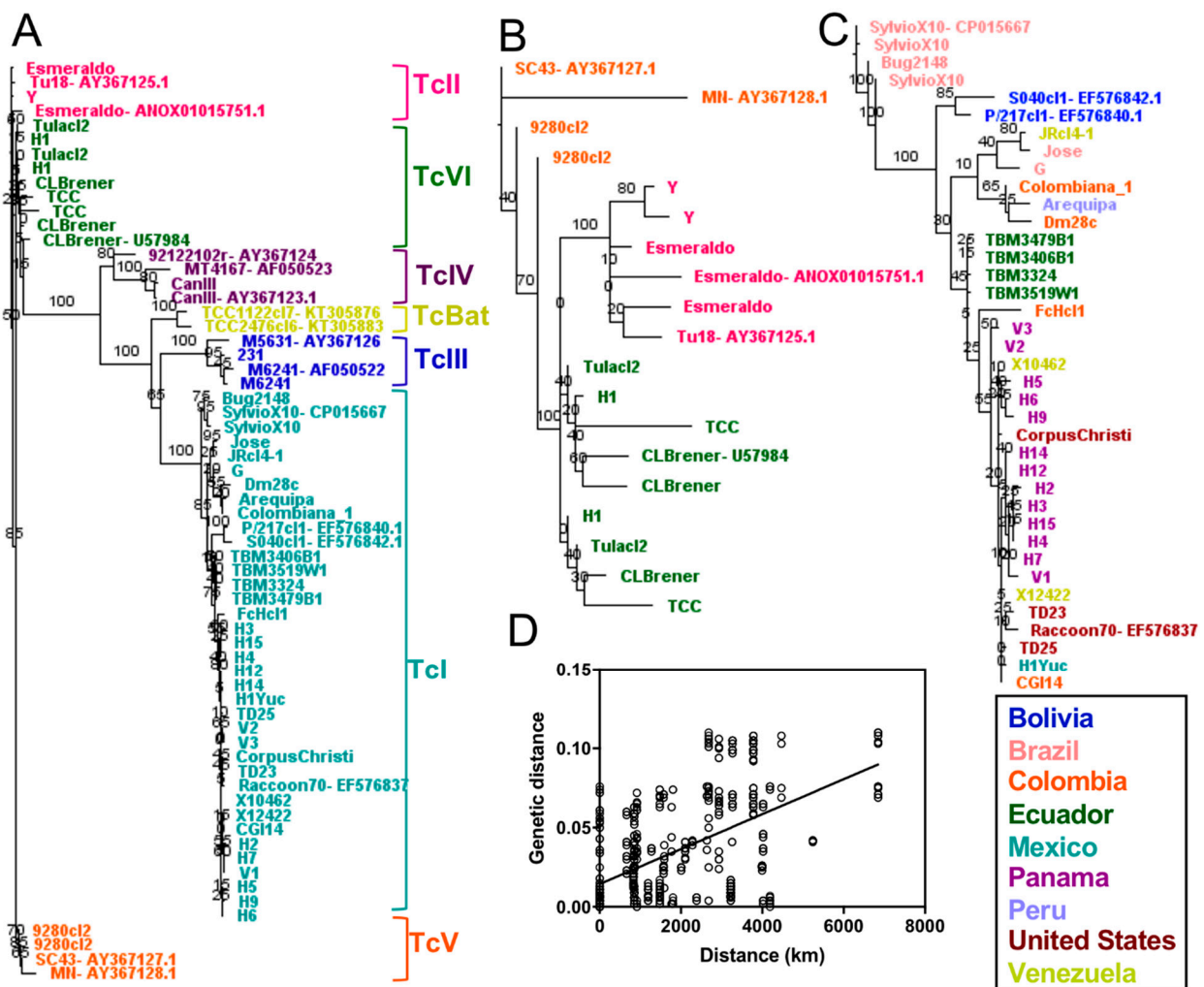
clearly resolve DTUs while also evidencing some level of intra-lineage variation, particularly for TcI and TcII. These markers may thus provide finer resolution for the detailed assessment of parasite genetic diversity beyond the major lineages.

We also evaluated how the mini-exon spliced leader marker discriminated among strains, as it has been one of the most extensively used marker [7,14,16,58–62]. BLAST analysis confirmed that it is a high copy number sequence, with up to 800 copies per genome, depending on the strain (Supplementary Table S2). We evaluated how multiple copies of the mini-exon sequences may overlap or blend among strains and DTUs. Phylogenetic analysis indicated that most copies clearly clustered for each strain and each DTU, following a Russian doll pattern, although some strains presented multiple clusters of mini-exon sequences (Figure 3). A notable exception was for a small number of sequences from TcII strains, which clustered with both TcV and TcVI sequences, in agreement with the hybrid origin of these DTUs (Figure 3B).



**Figure 3.** Analysis of multiple copies of mini-exon sequences. Phylogenetic analysis of Mini-exon sequences was performed for TcI (A) and TcII, TcV and TcVI (B). The names of the strains included in each tree are indicated (Arequipa, Jose, FcHcl1, H1Yuc, and SylvioX10 for TcI, and CLBrener, Tulacl2, 9280cl2, TCC, Y, Tu18, Esmeraldo, MN and SC43 for TcII, TcV and TcVI). Most copies of mini-exon sequences cluster by strain and by DTU in a Russian doll pattern, although some strains present multiple clusters and a small number of TcII sequences clustered with either TcV or TcVI sequences, in agreement with the hybrid nature of the latter DTUs.

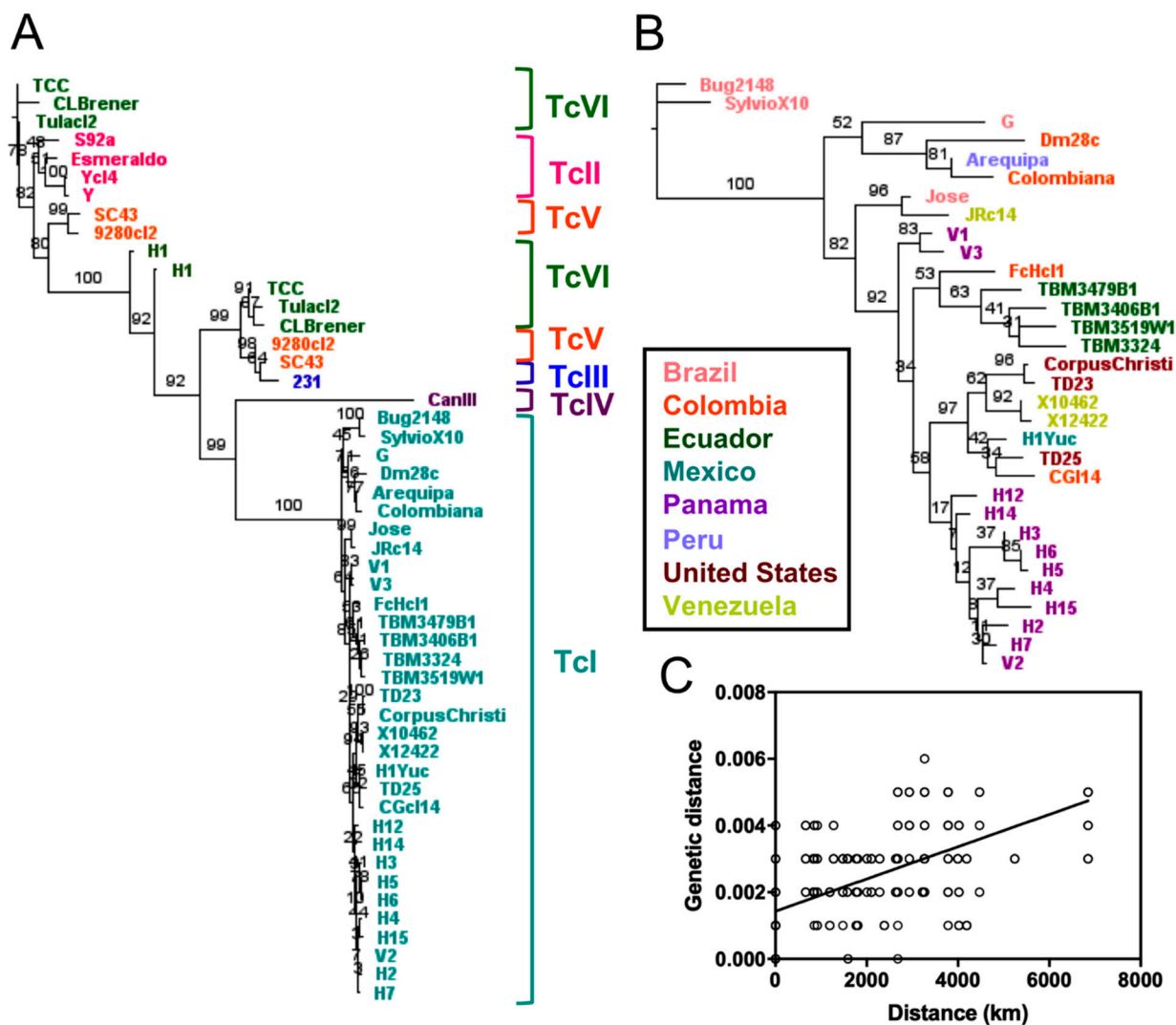
Detailed phylogenetic analysis using a single dominant mini-exon sequence from each strain resolved well all DTUs (Figure 4), including DTUs TcII, TcV and TcVI, which were clearly identified and well separated from TcIII (Figure 4A,B). In addition, the mini-exon revealed particular intra-lineage diversity for TcI (Figure 4C) and hinted at a similar diversity within other DTUs for which multiple sequences were available. Within TcI, this genetic diversity was mostly associated with the geographic origin of the strains with a strong isolation by distance (Figure 4D, Mantel test,  $p = 0.0002$ ,  $R = 0.54$ ). The mini-exon sequences were well suited to assess *T. cruzi* parasite diversity at the DTU level (Mantel test,  $p = 0.0001$ ,  $R = 0.70$ ) as well as within DTUs (Mantel test,  $p = 0.0002$ ,  $R = 0.61$ ) and summarized well the diversity observed above at the whole genome level. Mini-exon sequence from strain “Bug2148” also clustered with sequences from TcI DTU, in agreement with the whole genome analysis above. Similarly, the mini-exon sequences of strain H1 from Panama clustered with DTU TcVI.



**Figure 4.** Phylogenetic analysis of a single dominant mini-exon sequence per strain. Phylogenetic analyses included all DTUs (A), TcII, TcV and TcVI (B) and TcI only (C). All DTUs are clearly resolved with this marker and intra-lineage diversity was particularly apparent for DTU TcI (C). A clear isolation by distance was observed within TcI (D), Mantel test ( $p = 0.0002$ ,  $R = 0.544$ ).

We also assessed the usefulness of phylogenetic trees constructed from concatenated sequences of most of the available markers ( $n = 30$  markers), corresponding to approximately 26 kbp of sequence for each strain (Figure 5). These trees showed defined clusters for all DTUs (Figure 5A), as observed for whole genomes and the mini-exon marker. However, this approach allowed identifying separate haplotypes for TcV and TcVI hybrid strains, corresponding to a TcII-like haplotype and TcIII-like haplotype, respectively. Though the TcII-like concatenation demonstrated higher congruence (Mantel test,  $p = 0.0001$ ,  $R = 0.71$ ), both this and the TcIII-like concatenation (Mantel test,  $p = 0.0003$ ,  $R = 0.59$ ) were similar overall to the whole genome analysis. Trees from concatenated markers also confirmed that “Bug2148” clustered with TcI strains and H1 from Panama clustered with TcV/TcVI strains (Figure 5A). As in the analyses above, intra-lineage diversity could be observed based on the concatenated sequences for those DTUs for which sequences from multiple strains are available, particularly TcI and to a lesser extent TcII and TcVI. Accordingly, a separate analysis of TcI sequences (Figure 5B) confirmed a strong genetic structuration of TcI parasites rather comparable to the whole genome analysis (Mantel test,  $p = 0.0001$ ,  $R = 0.75$ ), with multiple well-supported clusters, which mostly corresponded to the geographic origin of the strains (Figure 5C, Mantel test,  $p = 0.0001$ ,  $R = 0.35$ ). Thus, the use of multiple markers provided limited additional information compared to the mini-exon sequence alone.





**Figure 5.** Phylogenetic trees from concatenated sequences of most available markers.  $n = 30$  markers corresponding to ~26 kbp. Clear clustering by DTU was observed (A). A separate analysis of TcI sequences (B) showed multiple well-supported clusters within DTU, with diversity again mostly corresponding to geographic origin (C) by Mantel test, ( $p = 0.0001$ ,  $R = 0.35$ ).

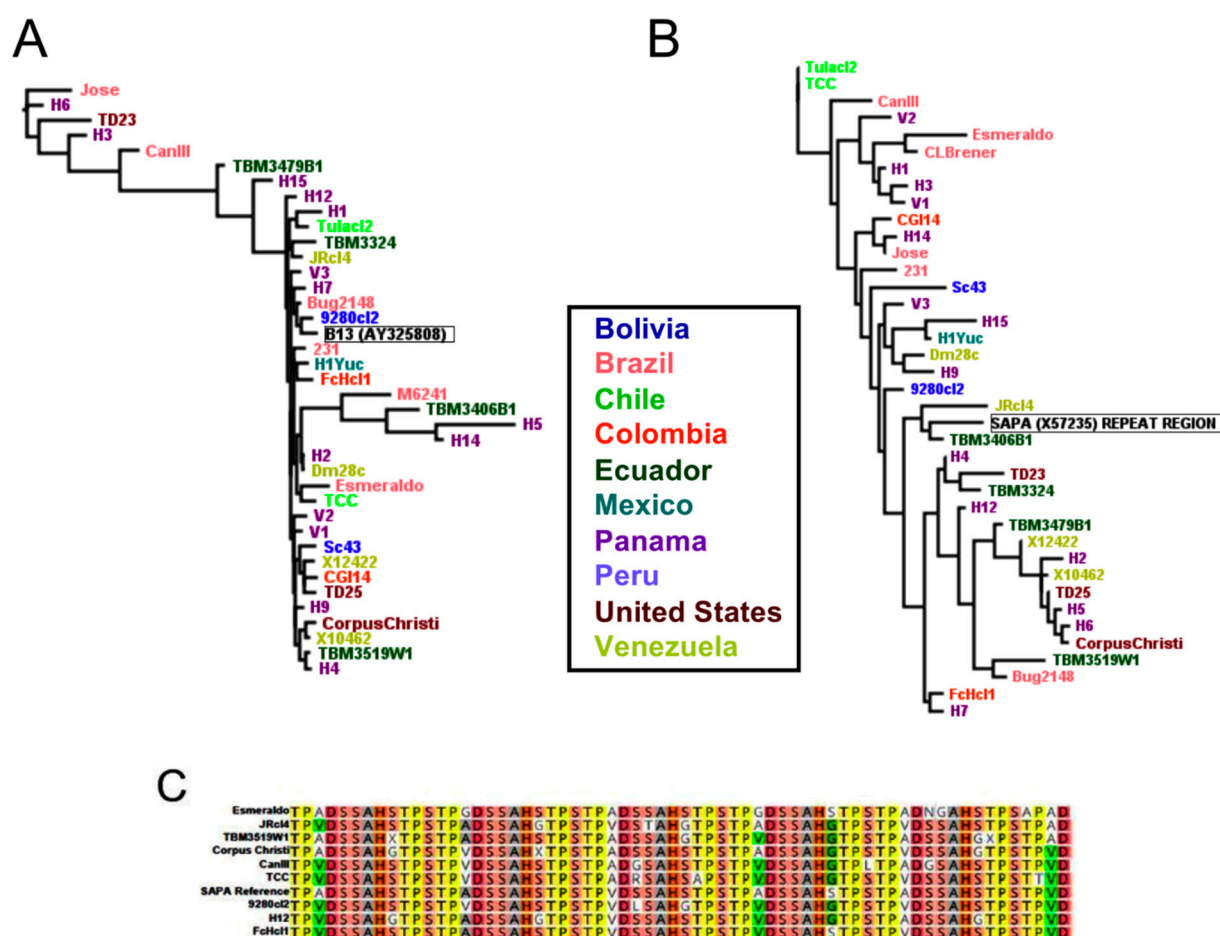
### 2.3. Diagnostic Antigen Conservation

The high genomic diversity of *T. cruzi*, both among DTUs as well as according to their geographic origin within DTUs, supports the hypothesis that limited antigen conservation may explain the variable performance of several current commercial serological diagnostic tests in some geographic areas. To further test this hypothesis, we performed BLAST searches of the available genomes for seven antigen sequences that are a subset of antigens used in some of these tests (Table 2), to assess the conservation of their predicted translation. Antigen1, Antigen36, and KMP-11 which correspond to small repeated sequences, were the most conserved, with alignments of peptide sequences representing all DTUs demonstrating an average 95.3% pairwise identity for Antigen1, 98.3% pairwise identity for Antigen36, and 95.6% pairwise identity for KMP-11. The sequences used in diagnostic tests were most closely related to those from strains from south America and from multiple DTUs but diverged more from sequences from Central and North American strains (Supplementary Figure S2).

**Table 2.** Current commercial serological diagnostic antigens evaluated for sequence diversity across strains.

Antigen	Accession No.	Current Diagnostic
Ag1 [63]	M21330	Chagatest recombinant (Wiener), T-detect (InBios)
Ag36 [63]	M21331	Chagatest recombinant (Wiener), T-detect (InBios)
Ag30 [63,64]	n/a	Chagatest recombinant (Wiener), T-detect (InBios)
KMP-11 [63,65]	n/a	T-detect (InBios)
B13 [66]	AY325808	Stat-Pak (ChemBio)
SAPA [63,67]	X57235	Chagatest recombinant (Wiener), T-detect (InBios)
TcH49 [66]	L09564	Stat-Pak (ChemBio)

Other antigens including the shed acute phase antigen (SAPA) repeat region, Anti-30, TcH49, and B13 presented more limited sequence conservation across all strains, with respective pairwise identities of 77.0%, 88.2%, 94.9% and 69.2%. The reference antigen currently used in serology was quite different from many sequences from other strains (Figure 6A,B). Although sequence diversity of the SAPA antigen was high, it did not seem to be structured according to geographic origin of the strains or parasite DTU, and this diversity was observed for both the repeat region of the protein that is used in diagnostic tests (Figure 6B,C) or the full length protein (Supplementary Figure S2). Overall, the poor conservation of these antigens across parasite DTU and geographic diversity may explain the limited performance of tests based on these antigens in Central and North America.



**Figure 6.** Comparison of antigen sequence diversity across strains for currently utilized diagnostic antigens. Antigen B13 (A) and the SAPA repeat region (B,C), demonstrated limited sequence conservation across strains, although diversity does not seem to be clearly structured according to geography.

### 3. Discussion

The genetic diversity of *T. cruzi* has been detected a long time ago [68,69], but understanding its full extent and the geographical ranges of the various lineages remains critical, as strain diversity may be associated with the diversity of clinical manifestations and parasite transmission cycles. Here, we performed the first extensive analysis of *T. cruzi* genomes from multiple DTUs and geographic origins, to shed light on the phylogeography of the parasite.

A first key finding was that *T. cruzi* genetic structure corresponding to DTUs is maintained at the nuclear genome level. While this subdivision has been previously established with a small number of genetic markers [7], it is important to see it reflected at the whole genome level as this confirms that it corresponds to a strong evolutionary signal of *T. cruzi* genetic structure and strain diversity, and a useful classification of strains. However, we acknowledge that given the limitations in the number of available complete genomes from non-TcI strains, further confirmation of this result with additional TcIII and TcIV strains included is needed. In spite of a different inheritance mechanism and different evolutionary pressures, this structuration according to DTU was also reflected to some extent in the kinetoplast maxicircle sequences. Nonetheless, analysis of kinetoplast maxicircle also revealed clear examples of introgression. The close relatedness of the two strains displaying this introgression, as well as the apparently ancient origin of their kinetoplast suggests that they may have derived from a common ancestor, and it raises the question of the mechanisms and frequency of such kinetoplast introgression. Indeed, while *T. cruzi* population structure has been assumed to be mostly clonal [52,70,71], recent studies pointed to sexual reproduction and recombination at rates that reject a purely clonal model, associated with a high inbreeding [53,54]. Such sexual reproduction may provide opportunities for different inheritance patterns for kinetoplast DNA.

Within DTUs, a strong genetic structuring of *T. cruzi* was also detected within TcI, in agreement with a previous study [17]. This structuring corresponded in part to the geographic origin of the strains, but not exclusively, suggesting that multiple factors contribute to the diversity of strains within TcI. Indeed, multiple sublineages were detected within TcI from the same geographic region and possibly correspond to the previously detected groups TcIa to TcIe based on mini-exon sequences [14–16], which appear in distinct genomic clusters, but genomes from additional strains are needed to assess this correspondence. These multiple sublineages from the same country or region may be a result of distinct epidemiological and transmission cycles resulting from different hosts and environments. Alternatively, certain strains isolated in a particular region may represent introductions to the area as a result of host movement. Similarly, TcII, and to some extent TcVI, appear to exhibit intra-DTU diversity as well, and comparison of a greater number of strains from all DTUs will be necessary to assess their intra-DTU diversity and genetic structure.

While whole genomes certainly provide the most complete evaluation of *T. cruzi* diversity, our analysis of single and concatenated markers provides strong evidence that the extent of genomic diversity can be captured through limited sequence information. This would be particularly useful for relatively simple genotyping procedures of a very large number of strains to further study parasite diversity in endemic regions. Indeed, most of the single gene markers evaluated here, such as TcSD5D or the 24S RNA gene, can capture parasite diversity among DTU well, confirming their usefulness for such classification. Nonetheless, most of these markers are too conserved to detect significant intra-DTU diversity, and thus lead to an underestimation of parasite diversity. The mini-exon appears as a particularly good marker to assess both DTU and intra-DTU diversity. Although the polymorphic nature of the mini-exon has called into question its ability to summarize strain diversity because of intra-strain variation among orthologous copies, our results demonstrate that it recapitulates well multi-markers and whole genome diversity. Indeed, its intra- and inter-genome sequence diversity appears in

agreement with the Russian doll pattern proposed earlier to describe *T. cruzi* population structure [51]. The low mini-exon copy number observed in some strains likely represented limited sequencing coverage rather than a true lack of gene copies, as the spliced leader is essential for parasite viability [72,73], again pointing to limitations with currently available data. Therefore, our results indicate that to tease out local eco-epidemiology of the parasite, it is possible to focus on limited markers for a larger number of samples rather than on a high number of markers for fewer samples. At the same time, access to a greater number of genomes or reference sequences from those DTUs and geographic regions not well-represented in the literature is crucial to fully describe intra-lineage diversity for all DTUs.

Our analyses also indicated some unexpected discrepancies for some strains. First, H1 strain from Panama, was initially classified as belonging to TcI [17], but it rather belongs to the TcVI DTUs, and it is thus the first TcVI genome from a strain from Central America. Second, the genome of strain “Bug2148”, initially reported as belonging to TcV [43], unambiguously clustered with TcI strains for all markers tested as well as at the whole genome level. As multiple studies with different markers have previously shown that the original Bug2148 strain indeed belonged to TcV [74,75], a likely explanation is that a different strain (belonging to TcI) was actually sequenced instead of Bug2148.

With such limited degree of genome conservation across all DTUs, diagnostics are likely to result in false negatives in particular regions where genetically diverse parasite strains may be circulating. Our analysis of some of the antigens currently used in diagnostic tests indicate that several antigens are poorly conserved across various parasite strains and DTUs. Thus, these tests are suited to detect human infections with a limited diversity of DTUs and strains but may not detect infection with strain variants. This may explain the limited performance of current tests in North America [30–33] and highlights the need for new tests able to detect infections with the full diversity of *T. cruzi* parasites.

#### 4. Conclusion

In conclusion, whole genome comparison confirmed the strong genetic structure of *T. cruzi* strains, in agreement with its current subdivision into well-established DTUs [76]. However, some significant intra-DTU diversity is emerging, associated in part to geographic structuring. Although not all tested markers perform equally well, most are well suited for estimating parasite diversity to DTU level, and a few such as the mini-exon, are particularly well suited to assess intra DTU diversity as well. Analysis of antigen sequences across all strains confirms that antigenic diversity is a likely explanation for limited diagnostic performance in Central and North America. Identifying new antigens conserved across the extent of diversity may lead to improved serological diagnostics.

#### 5. Materials and Methods

##### 5.1. *Trypanosoma cruzi* Genome Sequences and Markers

Genome sequences from 52 previously described parasite strains were included in this study, covering TcI to TcVI (Table 1). Notably, no whole genome sequence of TcBat is currently available.

For strains CL-Brener, Dm28c, 231, TCC, SC43, S11, Ycl4, S23b, S92a, S15, S162a, S44a, S154, G, and Bug2148 we used previously published assemblies [18,44,47,48,77–79]. For all the remaining strains, we used raw sequence reads from the SRA database [80]. These were newly assembled in Geneious 11.1.5, using Dm28c or TCC as reference genomes, as these represent the most completely assembled genomes to date [48]. The use of new assemblies allowed having a more homogenous dataset that minimized potential bias due to differences in sequencing methodologies and assemblies. These new genome assemblies from each strain were used to generate strain-specific databases in Geneious 11, and these were searched using the BLAST for 35 markers, most of which represent single copy sequences that had been used before for identifying *T. cruzi* genotypes and

evolutionary relationships among DTUs (Supplementary Table S1). Raw sequence reads were similarly used to assemble kinetoplast maxicircle sequences, using 231 (KC987253), CLBrener (DQ343645), Esmeraldo (DQ343646) and SylvioX10 (FJ203996) maxicircle sequences as references.

### 5.2. Whole Genome Alignments

Whole genome alignments were constructed using the progressiveMauve algorithm and default settings [81]. Thirteen genomes representing all DTUs except TcBat and 22 genomes representing only TcI were separately aligned to identify homologous genome regions, referred to as locally collinear blocks (LCBs) in Mauve. The extent of LCBs was compared across genomes to assess conservation among all six DTUs, as well as among TcI genomes, and concatenated LCB sequences were used to elaborate maximum likelihood trees using PHYML as implemented in Geneious 11. Isolation by distance was tested using Mantel test as implemented in Past3 software [82] and pair-wise geographic distances among strains were based on their country of origin.

### 5.3. Single Marker Analysis

Multiple sequence alignments were made for all available sequences of each individual marker using MUSCLE as implemented in Geneious. Phylogenetic trees based on maximum likelihood as implemented in PHYML were constructed using the Phylogeny.fr tool [83] or IQ-Tree [84] for each marker. For the analysis of the mini-exon sequences, additional reference sequences were downloaded from GenBank [85] database (Tu18: AY367125.1, Esmeraldo: ANOX01015457.1, CLBrener: U57984, 9122102r: AY367124, MT4167: AF050523, CanIII: AY367123.1, TCC1122cl7: KT305876, TCC2476cl6: KT305883, M5631: AY367126, M6241: AF050522, SylvioX10: CP015667, P/217cl1: EF576840.1, S040cl1: EF576842, Raccoon70: EF576837, SC43: AY367127.1, MN: AY367128.1). As the mini-exon is a multicopy gene of approximately 600 bp with variable copy number depending on strain, any BLAST return of at least 250 bp was considered for analysis. To evaluate how multiple copies of the mini-exon sequences may overlap or blend among strains and DTUs, parasite strains with sufficient genome coverage to allow for identification of a large numbers of mini exon paralogous sequences of sufficient length were analyzed separately. Phylogenetic trees were edited using FigTree 1.4.4 for final rendering [86]. The Mantel test, as implemented in PAST software [82], was used to determine congruence between trees based on whole genomes and trees based on the mini-exon, both for all DTU trees and for TcI only trees.

### 5.4. Concatenated Markers Analysis

Concatenations were made of marker sequences for each genome/strain. Because some marker sequences were absent from certain genomes due to limited sequencing coverage, a total of 30 markers were concatenated for 44 genomes. For those DTUs that are hybrid in origin, namely TcV and TcVI, two concatenations were made: one for the TcII-like sequences and one for the TcIII-like sequences. Concatenated sequences were aligned using MUSCLE, and maximum likelihood phylogenetic trees were constructed using PHYML as for single markers. Separate alignments and phylogenetic trees were also constructed for TcI and non-TcI parasite DTUs for increased resolution of within-DTU diversity. Isolation by distance was tested using Mantel test as implemented in Past3 software [82]. The Mantel test was also used to determine congruence between whole genome trees and concatenated trees for all DTUs and TcI only. For all-DTU concatenated trees, separate iterations of the test were run for the concatenation containing the TcII-like sequences and that containing the TcIII-like sequences.

### 5.5. Antigen Conservation

To evaluate the conservation of antigens that are used in current commercial serological diagnostic tests for *T. cruzi*, genome databases were searched via BLAST for sequences corresponding to seven antigens (Table 2).

Commercial tests represented include Chagas StatPak® (ChemBio Diagnostics Inc., Hauppauge, NY, USA) Trypanosoma Detect™ Rapid Test (InBios International Inc., Seattle, WA, USA), and Chagatest ELISA (Wiener Lab, Rosario, Argentina). StatPak® utilizes antigens B13, H49, and 1F8. 1F8 was excluded from our analysis as has been extensively characterized previously [37]. Trypanosoma Detect™ consists of a fusion peptide made up of portions of antigens 1, 30, 36, SAPA, Kmp-11, and multiepitope recombinant TcF. As TcF epitopes did not return BLAST results in numerous genomes, this antigen was excluded from further analysis. Wiener's Chagatest consists of antigens 1, 2, 13, 30, 36 and SAPA. Antigens 2 and 13 are very short peptide sequences (12 aa and 5 aa) that are identical to epitopes within B13 and TcD, part of the TcF multiepitope recombinant protein, so were not included separately in our analysis. DNA sequences were translated into protein sequences for each *T. cruzi* strain, using all six reading frames and selecting the translation that most closely matched the reference sequence annotations. All peptide sequences were aligned for each antigen, including the reference sequence, and trees were constructed from these alignments using PhyML with the LG (Le Gascuel) substitution model.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2076-0817/10/2/212/s1](http://www.mdpi.com/2076-0817/10/2/212/s1), Figure S1: Phylogenetic trees constructed for individual markers, Figure S2: Phylogenetic trees for antigen sequences used in current serologic diagnostics, Table S1: Markers used to assess *T. cruzi* diversity, Table S2: Mini-exon copy number in *T. cruzi* genomes.

**Author contributions:** A.M. performed the analyses and wrote the manuscript. L.M. contributed initial analyses and databases. E.D. and C.H. designed the study and analyses and revised the manuscript. All named authors meet the International Committee of Medical Journal Editors (IC-MJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant number 5R01HD94955) to CH and by the Louisiana Board of Regents through the Board of Regents Support Fund [# LESASF (2018-21)-RD-A-19] to ED.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** All genomic sequence data for this study are available at the TriTryp (<https://tritrypdb.org/tritrypdb/>) and SRA (<https://www.ncbi.nlm.nih.gov/sra>) and GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) databases.

**Conflicts of Interest:** All authors declare that they have no conflicts of interest.

## References

1. Browne, A.J.; Guerra, C.A.; Alves, R.V.; da Costa, V.M.; Wilson, A.L.; Pigott, D.M.; Hay, S.I.; Lindsay, S.W.; Golding, N.; Moyes, C.L. The contemporary distribution of Trypanosoma cruzi infection in humans, alternative hosts and vectors. *Sci. Data* **2017**, *4*, 170050.
2. Messenger, L.A.; Miles, M.A.; Bern, C. Between a bug and a hard place: Trypanosoma cruzi genetic diversity and the clinical outcomes of Chagas disease. *Expert Rev. Anti. Infect. Ther.* **2015**, *13*, 995–1029.
3. Vandermark, C.; Ziemann, E.; Boyles, E.; Nielsen, C.K.; Davis, C.; Jiménez, F.A. Trypanosoma cruzi strain TcIV infects raccoons from Illinois. *Memórias Inst. Oswaldo Cruz* **2018**, *113*, 30–37, doi:10.1590/0074-02760170230.
4. Fact Sheet: Chagas Disease [Internet]: PAHO; 2017. Available from: <https://www.paho.org/en/documents/chagas-americas-public-health-workers-2017> (accessed on 18 December 2020)



5. Lee, B.Y.; Bacon, K.M.; Bottazzi, M.E.; Hotez, P.J. Global economic burden of Chagas disease: A computational simulation model. *Lancet Infect. Dis.* **2013**, *13*, 342–348, doi:10.1016/s1473-3099(13)70002-1.
6. Messenger, L.A.; Llewellyn, M.S.; Bhattacharyya, T.; Franzén, O.; Lewis, M.D.; Ramírez, J.D.; Carrasco, H.J.; Andersson, B.; Miles, M.A. Multiple Mitochondrial Introgression Events and Heteroplasmy in *Trypanosoma cruzi* Revealed by Maxicircle MLST and Next Generation Sequencing. *PLOS Neglected Trop. Dis.* **2012**, *6*, e1584, doi:10.1371/journal.pntd.0001584.
7. Zingales, B.; Miles, M.A.; Campbell, D.A.; Tibayrenc, M.; Macedo, A.M.; Teixeira, M.M.; Schijman, A.G.; Llewellyn, M.S.; Lages-Silva, E.; Machado, C.R.; et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infect. Genet. Evol.* **2012**, *12*, 240–253, doi:10.1016/j.meegid.2011.12.009.
8. Marcili, A.; Lima, L.; Cavazzana, M.J.; Junqueira, A.C.V.; Veludo, H.H.; Da Silva, F.M.; Campaner, M.; Paiva, F.; Nunes, V.L.B.; Teixeira, M.M.G. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitol.* **2009**, *136*, 641–655, doi:10.1017/s0031182009005861.
9. Brenière, S.F.; Waleckx, E.; Barnabé, C. Over Six Thousand *Trypanosoma cruzi* Strains Classified into Discrete Typing Units (DTUs): Attempt at an Inventory. *PLOS Neglected Trop. Dis.* **2016**, *10*, e0004792, doi:10.1371/journal.pntd.0004792.
10. Ramírez, J.C.; Torres, C.; Curto, M.D.L.A.; Schijman, A.G. New insights into *Trypanosoma cruzi* evolution, genotyping and molecular diagnostics from satellite DNA sequence analysis. *PLOS Neglected Trop. Dis.* **2017**, *11*, e0006139, doi:10.1371/journal.pntd.0006139.
11. Lewis, M.D.; Llewellyn, M.S.; Yeo, M.; Acosta, N.; Gaunt, M.W.; Miles, M.A. Recent, Independent and Anthropogenic Origins of *Trypanosoma cruzi* Hybrids. *PLOS Neglected Trop. Dis.* **2011**, *5*, e1363, doi:10.1371/journal.pntd.0001363.
12. Barnabé, C.; Mobarec, H.I.; Jurado, M.R.; Cortez, J.A.; Brenière, S.F. Reconsideration of the seven discrete typing units within the species *Trypanosoma cruzi* a new proposal of three reliable mitochondrial clades. *Infect. Genet. Evol.* **2016**, *39*, 176–186, doi:10.1016/j.meegid.2016.01.029.
13. Burgos, J.M.; Altcheh, J.; Bisio, M.; Duffy, T.; Valadares, H.M.; Seidenstein, M.E.; Piccinalli, R.; Freitas, J.M.; Levin, M.J.; Macchi, L.; et al. Direct molecular profiling of minicircle signatures and lineages of *Trypanosoma cruzi* bloodstream populations causing congenital Chagas disease. *Int. J. Parasitol.* **2007**, *37*, 1319–1327, doi:10.1016/j.ijpara.2007.04.015.
14. Herrera, C.; Bargues, M.D.; Fajardo, A.; Montilla, M.; Triana, O.; Vallejo, G.A.; Guhl, F. Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. *Infect. Genet. Evol.* **2007**, *7*, 535–539, doi:10.1016/j.meegid.2006.12.003.
15. Falla, A.; Herrera, C.; Fajardo, A.; Montilla, M.; Vallejo, G.A.; Guhl, F. Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. *Acta Trop.* **2009**, *110*, 15–21, doi:10.1016/j.actatropica.2008.12.003.
16. Cura, C.I.; Mejía-Jaramillo, A.M.; Duffy, T.; Burgos, J.M.; Rodriguero, M.; Cardinal, M.V.; Kjos, S.; Gurgel-Gonçalves, R.; Blanchet, D.; De Pablos, L.M.; et al. *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. *Int. J. Parasitol.* **2010**, *40*, 1599–1607, doi:10.1016/j.ijpara.2010.06.006.
17. Talavera-López, C.; Messenger, L.A.; Lewis, M.D.; Yeo, M.; Reis-Cunha, J.L.; Bartholomeu, D.C.; Calzada, J.E.; Saldaña, A.; Ramírez, J.D.; Guhl, F.; et al. Repeat-driven generation of antigenic diversity in a major human pathogen *Trypanosoma cruzi*. *bioRxiv*. 2018:283531.
18. Reis-Cunha, J.L.; Baptista, R.P.; Rodrigues-Luiz, G.F.; Coqueiro-Dos-Santos, A.; Valdivia, H.O.; De Almeida, L.V.; Cardoso, M.S.; D'Ávila, D.A.; Dias, F.H.C.; Fujiwara, R.T.; et al. Whole genome sequencing of *Trypanosoma cruzi* field isolates reveals extensive genomic variability and complex aneuploidy patterns within TcII DTU. *BMC Genom.* **2018**, *19*, 816, doi:10.1186/s12864-018-5198-4.
19. Barnabé, C.; Brisse, S.; Tibayrenc, M. Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas disease: A multilocus enzyme electrophoresis approach. *Parasitology* **2000**, *120*, 513–526, doi:10.1017/s0031182099005661.
20. Baptista, R.d.P.; D'Ávila, D.A.; Segatto, M.; do Valle, Í.F.; Franco, G.R.; Valadares, H.M.S.; Dias Gontijo, E.; da Cunha Galvão, L.M.; Junho Pena, S.D.; Chiari E.; et al. Evidence of substantial recombination among *Trypanosoma cruzi* II strains from Minas Gerais. *Infect. Genet. Evol.* **2014**, *22*, 183–191.
21. Yeo, M.; Mauricio, I.L.; Messenger, L.A.; Lewis, M.D.; Llewellyn, M.S.; Acosta, N.; Bhattacharyya, T.; Diosque, P.; Carrasco, H.J.; Miles, M.A. Multilocus Sequence Typing (MLST) for Lineage Assignment and High Resolution Diversity Studies in *Trypanosoma cruzi*. *PLOS Neglected Trop. Dis.* **2011**, *5*, e1049, doi:10.1371/journal.pntd.0001049.
22. Roellig, D.M.; Savage, M.Y.; Fujita, A.W.; Barnabé, C.; Tibayrenc, M.; Steurer, F.J.; Yabsley, M.J. Genetic Variation and Exchange in *Trypanosoma cruzi* Isolates from the United States. *PLoS ONE* **2013**, *8*, e56198, doi:10.1371/journal.pone.0056198.
23. Llewellyn, M.S.; Lewis, M.D.; Acosta, N.; Yeo, M.; Carrasco, H.J.; Segovia, M.; Vargas, J.; Torrico, F.; Miles, M.A.; Gaunt, M.W. *Trypanosoma cruzi* IIc: Phylogenetic and Phylogeographic Insights from Sequence and Microsatellite Analysis and Potential Impact on Emergent Chagas Disease. *PLOS Negl. Trop. Dis.* **2009**, *3*, e510, doi:10.1371/journal.pntd.0000510.
24. Izeta-Alberdi, A.; Ibarra-Cerdeña, C.N.; Moo-Llanes, D.A.; Ramsey, J.M. Geographical, landscape and host associations of *Trypanosoma cruzi* DTUs and lineages. *Parasites Vectors* **2016**, *9*, 631, doi:10.1186/s13071-016-1918-2.
25. Pronovost, H.; Peterson, A.C.; Chavez, B.G.; Blum, M.J.; Dumonteil, E.; Herrera, C.P. Deep sequencing reveals multiclinality and new discrete typing units of *Trypanosoma cruzi* in rodents from the southern United States. *J. Microbiol. Immunol. Infect.* **2020**, *53*, 622–633, doi:10.1016/j.jmii.2018.12.004.

26. Monje-Rumi, M.; Floridia-Yapur, N.; Zago, M.; Ragone, P.; Brandán, C.P.; Nuñez, S.; Barrientos, N.; Tomasini, N.; Diosque, P. Potential association of *Trypanosoma cruzi* DTUs TcV and TcVI with the digestive form of Chagas disease. *Infect. Genet. Evol.* **2020**, *84*, 104329, doi:10.1016/j.meegid.2020.104329.
27. Miles, M.A.; Póvoa, M.M.; Prata, A.; Cedillos, R.A.; De Souza, A.A.; Macedo, V. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' Disease? *Lancet* **1981**, *317*, 1338–1340.
28. Gruber, A.; Zingales, B. *Trypanosoma cruzi*: Characterization of Two Recombinant Antigens with Potential Application in the Diagnosis of Chagas' Disease. *Exp. Parasitol.* **1993**, *76*, 1–12, doi:10.1006/expr.1993.1001.
29. Carmona, S.J.; Nielsen, M.; Schafer-Nielsen, C.; Mucci, J.; Altcheh, J.; Balouz, V.; Tekiel, V.; Frasc, A.C.; Campetella, O.; Buscaglia, C.A.; et al. Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants. *Mol. Cell. Proteom.* **2015**, *14*, 1871–1884, doi:10.1074/mcp.m114.045906.
30. Guzmán-Gómez, D.; López-Monteon, A.; Lagunes-Castro, M.D.L.S.; Álvarez-Martínez, C.; Hernández-Lutzon, M.J.; Dumonteil, E.; Ramos-Ligonio, A. Highly discordant serology against *Trypanosoma cruzi* in central Veracruz, Mexico: Role of the antigen used for diagnostic. *Parasites Vectors* **2015**, *8*, 466, doi:10.1186/s13071-015-1072-2.
31. Buekens, P.; Althabe, F.; López, B.; Graiff, O.; Ramírez-Sierra, M.J.; Sosa-Estani, S.; Wesson, D.M.; Bustamante, N.; Alger, J.; Zúñiga, C.; et al. Congenital Transmission of *Trypanosoma cruzi* in Argentina, Honduras, and Mexico: An Observational Prospective Study. *Am. J. Trop. Med. Hyg.* **2018**, *98*, 478–485, doi:10.4269/ajtmh.17-0516.
32. Elmayan, A.; Tu, W.; Duhon, B.; Marx, P.; Wolfson, W.; Balsamo, G.; Herrera, C.; Dumonteil, E. High prevalence of *Trypanosoma cruzi* infection in shelter dogs from southern Louisiana, USA. *Parasites Vectors* **2019**, *12*, 1–8, doi:10.1186/s13071-019-3572-y.
33. Whitman, J.D.; Bulman, C.A.; Gunderson, E.L.; Irish, A.M.; Townsend, R.L.; Stramer, S.L.; Sakanari, J.A.; Bern, C. Chagas Disease Serological Test Performance in U.S. Blood Donor Specimens. *J. Clin. Microbiol.* **2019**, *57*, doi:10.1128/jcm.01217-19.
34. Martin, D.L.; Weatherly, D.B.; A Laucella, S.; A Cabinian, M.; Crim, M.T.; Sullivan, S.; Heiges, M.; Craven, S.H.; Rosenberg, C.S.; Collins, M.H.; et al. CD8+ T-Cell Responses to *Trypanosoma cruzi* Are Highly Focused on Strain-Variant trans-Sialidase Epitopes. *PLOS Pathog.* **2006**, *2*, e77, doi:10.1371/journal.ppat.0020077.
35. Weatherly, D.B.; Peng, D.; Tarleton, R.L. Recombination-driven generation of the largest pathogen repository of antigen variants in the protozoan *Trypanosoma cruzi*. *BMC Genom.* **2016**, *17*, 1–17, doi:10.1186/s12864-016-3037-z.
36. Haolla, F.A.; Claser, C.; De Alencar, B.C.; Tzelepis, F.; De Vasconcelos, J.R.; De Oliveira, G.; Silvério, J.C.; Machado, A.V.; Lannes-Vieira, J.; Bruna-Romero, O.; et al. Strain-specific protective immunity following vaccination against experimental *Trypanosoma cruzi* infection. *Vaccine* **2009**, *27*, 5644–5653, doi:10.1016/j.vaccine.2009.07.013.
37. Arnal, A.; Villanueva-Lizama, L.; Teh-Poot, C.; Herrera, C.; Dumonteil, E. Extent of polymorphism and selection pressure on the *Trypanosoma cruzi* vaccine candidate antigen Tc24. *Evol. Appl.* **2020**, *13*, 2663–2672, doi:10.1111/eva.13068.
38. Macedo AM, Oliveira RP, Pena SD. Chagas disease: Role of parasite genetic variation in pathogenesis. *Expert Rev. Mol. Med.* **2002**, *4*, 1–16.
39. Andrade, S.G.; Magalhães, J.B.; Pontes, A.L. Evaluation of chemotherapy with benznidazole and nifurtimox in mice infected with *Trypanosoma cruzi* strains of different types\*. *Bull. World Heal. Organ.* **1985**, *63*, 721–726.
40. Filardi, L.; Brener, Z. Susceptibility and natural resistance of *Trypanosoma cruzi* strains to drugs used clinically in Chagas disease. *Trans. R. Soc. Trop. Med. Hyg.* **1987**, *81*, 755–759, doi:10.1016/0035-9203(87)90020-4.
41. Zingales, B. *Trypanosoma cruzi* genetic diversity: Something new for something known about Chagas disease manifestations, serodiagnosis and drug sensitivity. *Acta Trop.* **2018**, *184*, 38–52, doi:10.1016/j.actatropica.2017.09.017.
42. Villanueva-Lizama, L.; Teh-Poot, C.; Majeau, A.; Herrera, C.; Dumonteil, E. Molecular Genotyping of *Trypanosoma cruzi* by Next-Generation Sequencing of the Mini-Exon Gene Reveals Infections With Multiple Parasite Discrete Typing Units in Chagasic Patients From Yucatan, Mexico. *J. Infect. Dis.* **2019**, *219*, 1980–1988, doi:10.1093/infdis/jiz047.
43. Callejas-Hernández, F.; Gironés, N.; Fresno, M. Genome Sequence of *Trypanosoma cruzi* Strain Bug2148. *Genome Announc.* **2018**, *6*, e01497–17, doi:10.1128/genomea.01497-17.
44. DeCuir, J.T.; Tu, W.; Dumonteil, E.; Herrera, C. Sequence of *Trypanosoma cruzi* reference strain SC43 nuclear genome and kinetoplast maxicircle confirms a strong genetic structure among closely related parasite discrete typing units. *Genome* **2020**, *1–7*, doi:10.1139/gen-2020-0092.
45. A Minning, T.; Weatherly, D.B.; Flibotte, S.; Tarleton, R.L. Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in *Trypanosoma cruzi* strains revealed by array comparative genomic hybridization. *BMC Genom.* **2011**, *12*, 139–11, doi:10.1186/1471-2164-12-139.
46. Souza, R.T.; Lima, F.M.; Barros, R.M.; Cortez, D.R.; Santos, M.F.; Cordero, E.M.; Ruiz, J.C.; Goldenberg, S.; Teixeira, M.M.G.; Da Silveira, J.F. Genome Size, Karyotype Polymorphism and Chromosomal Evolution in *Trypanosoma cruzi*. *PLoS ONE* **2011**, *6*, e23042, doi:10.1371/journal.pone.0023042.
47. El-Sayed, N.M.; Myler, P.J.; Bartholomeu, D.C.; Nilsson, D.; Aggarwal, G.; Tran, A.-N.; Ghedin, E.; Worthey, E.A.; Delcher, A.L.; Blandin, G.; et al. The Genome Sequence of *Trypanosoma cruzi*, Etiologic Agent of Chagas Disease. *Science* **2005**, *309*, 409–415, doi:10.1126/science.1112631.
48. Berná, L.; Rodriguez, M.; Chiribao, M.L.; Parodi-Talice, A.; Pita, S.; Rijo, G.; Alvarez-Valin, F.; Robello, C. Expanding an expanded genome: Long-read sequencing of *Trypanosoma cruzi*. *Microb. Genom.* **2018**, *4*, e000177, doi:10.1099/mgen.0.000177.

49. Franzen, O.; Ochaya, S.; Sherwood, E.; Lewis, M.D.; Llewellyn, M.S.; Miles, M.A.; Andersson, B. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. *PLoS Negl. Trop. Dis.* **2011**, *5*, 984.
50. Díaz-Viraqué, F.; Pita, S.; Greif, G.; Souza, R.D.C.M.D.; Iraola, G.; Robello, C. Nanopore Sequencing Significantly Improves Genome Assembly of the Protozoan Parasite *Trypanosoma cruzi*. *Genome Biol. Evol.* **2019**, *11*, 1952–1957, doi:10.1093/gbe/evz129.
51. Tibayrenc, M.; Ayala, F.J. The population genetics of *Trypanosoma cruzi* revisited in the light of the predominant clonal evolution model. *Acta Trop.* **2015**, *151*, 156–165, doi:10.1016/j.actatropica.2015.05.006.
52. Bartholomeu DC, Buck GA, Teixeira SMR, El-Sayed NMA. Genetics of *Trypanosoma cruzi*: Nuclear genome. In: Telleria J, Tibayrenc M, Eds.; *American Trypanosomiasis Chagas disease One hundred Years of Research*. Elsevier: London, UK, 2010; p. 443–448.
53. Berry, A.S.F.; Salazar-Sánchez, R.; Castillo-Neyra, R.; Borrini-Mayori, K.; Chipana-Ramos, C.; Vargas-Maquera, M.; Ancca-Juarez, J.; Náquira-Velarde, C.; Levy, M.Z.; Brisson, D.; et al. Sexual reproduction in a natural *Trypanosoma cruzi* population. *PLOS Neglected Trop. Dis.* **2019**, *13*, e0007392, doi:10.1371/journal.pntd.0007392.
54. Schwabl, P.; Imamura, H.; Broeck, F.V.D.; Costales, J.A.; Miguashca-Sánchez, J.; Miles, M.A.; Andersson, B.; Grijalva, M.J.; Llewellyn, M.S. Meiotic sex in Chagas disease parasite *Trypanosoma cruzi*. *Nat. Commun.* **2019**, *10*, 1–14, doi:10.1038/s41467-019-11771-z.
55. Cosentino, R.O.; Agüero, F. A Simple Strain Typing Assay for *Trypanosoma cruzi*: Discrimination of Major Evolutionary Lineages from a Single Amplification Product. *PLOS Neglected Trop. Dis.* **2012**, *6*, e1777, doi:10.1371/journal.pntd.0001777.
56. Augusto-Pinto, L.; Teixeira, S.M.R.; Pena, S.D.J.; Machado, C.R. Single-nucleotide polymorphisms of the *Trypanosoma cruzi* MSH2 gene support the existence of three phylogenetic lineages presenting differences in mismatch-repair efficiency. *Genet.* **2003**, *164*, 117–126.
57. De Freitas, J.M.; Augusto-Pinto, L.; Pimenta, J.R.; Bastos-Rodrigues, L.; Goncalves, V.F.; Teixeira, S.M.; Chiari, E.; Junqueira, A.C.V.; Fernandes, O.; Macedo, A.M.; et al. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathog.* **2006**, *2*, e24.
58. Lewis, M.D.; Ma, J.; Yeo, M.; Carrasco, H.J.; Llewellyn, M.S.; Miles, M.A. Genotyping of *Trypanosoma cruzi*: Systematic Selection of Assays Allowing Rapid and Accurate Discrimination of All Known Lineages. *Am. J. Trop. Med. Hyg.* **2009**, *81*, 1041–1049, doi:10.4269/ajtmh.2009.09-0305.
59. Souto, R.P.; Fernandes, O.; Macedo, A.M.; Campbell, D.A.; Zingales, B. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Mol Biochem Parasitol.* **1996**, *83*, 141–152.
60. Fernandes, O.; Sturm, N.R.; Derré, R.; A Campbell, D. The mini-exon gene: A genetic marker for zymodeme III of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **1998**, *95*, 129–133.
61. Tomasini, N.; Lauthier, J.J.; Rumi, M.M.M.; Ragone, P.G.; D’Amato, A.A.A.; Brandan, C.P.; Cura, C.I.; Schijman, A.G.; Barnabé, C.; Tibayrenc, M.; et al. Interest and limitations of Spliced Leader Intergenic Region sequences for analyzing *Trypanosoma cruzi* I phylogenetic diversity in the Argentinean Chaco. *Infect. Genet. Evol.* **2011**, *11*, 300–307, doi:10.1016/j.meegid.2010.10.020.
62. McCarthy-Burke, C.; Taylor, Z.A.; Buck, G.A. Characterization of the spliced leader genes and transcripts in *Trypanosoma cruzi*. *Gene* **1989**, *82*, 177–189.
63. Houghton, R.L.; Stevens, Y.Y.; Hjerrild, K.; Guderian, J.; Okamoto, M.; Kabir, M.; Reed, S.G.; Leiby, D.A.; Morrow, W.J.W.; Lorca, M.; et al. Lateral Flow Immunoassay for Diagnosis of *Trypanosoma cruzi* Infection with High Correlation to the Radioimmunoprecipitation Assay. *Clin. Vaccine Immunol.* **2009**, *16*, 515–520, doi:10.1128/cvi.00383-08.
64. Ibáñez, C.F.; Affranchino, J.L.; Macina, R.A.; Reyes, M.B.; Leguizamon, S.; Camargo, M.E.; Åslund, L.; Pettersson, U.; Frasch, A.C. Multiple *Trypanosoma cruzi* antigens containing tandemly repeated amino acid sequence motifs. *Mol. Biochem. Parasitol.* **1988**, *30*, 27–33, doi:10.1016/0166-6851(88)90129-6.
65. Thomas, M.C.; Longobardo, M.V.; Carmelo, E.; Marañón, C.; Planelles, L.; E Patarroyo, M.; Alonso, C.; López, M.C. Mapping of the antigenic determinants of the T. cruzi kinetoplastid membrane protein-11. Identification of a linear epitope specifically recognized by human Chagasic sera. *Clin. Exp. Immunol.* **2001**, *123*, 465–471.
66. Ponce, C.; Ponce, E.; Vinelli, E.; Montoya, A.; De Aguilar, V.; Gonzalez, A.; Zingales, B.; Rangel-Aldao, R.; Levin, M.J.; Esfandiari, J.; et al. Validation of a Rapid and Reliable Test for Diagnosis of Chagas’ Disease by Detection of *Trypanosoma cruzi*-Specific Antibodies in Blood of Donors and Patients in Central America. *J. Clin. Microbiol.* **2005**, *43*, 5065–5068, doi:10.1128/jcm.43.10.5065-5068.2005.
67. Umezawa, E.S.; Nascimento, M.S.; Kesper, N., Jr.; Coura, J.R.; Borges-Pereira, J.; Junqueira, A.C.; Camargo, M.E. Immunoblot assay using excreted-secreted antigens of *Trypanosoma cruzi* in serodiagnosis of congenital, acute, and chronic Chagas’ disease. *J Clin Microbiol.* **1996**, *34*, 2143–2147.
68. Miles, M.A.; De Souza, A.A.A.; Pova, M.; Shaw, J.J.; Lainson, R.; Toye, P.J. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas’ disease in Amazonian Brazil. *Nat. Cell Biol.* **1978**, *272*, 819–821, doi:10.1038/272819a0.
69. Tibayrenc, M.; Ward, P.; Moya, A.; Ayala, F.J. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclinal structure. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 115–119, doi:10.1073/pnas.83.1.115.
70. Tibayrenc, M.; Ayala, F. *Trypanosoma cruzi* populations: More clonal than sexual. *Parasitol. Today* **1987**, *3*, 189–190, doi:10.1016/0169-4758(87)90179-7.

71. Tibayrenc, M.; Kjellberg, F.; Ayala, F.J. A clonal theory of parasitic protozoa: The population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2414–2418, doi:10.1073/pnas.87.7.2414.
72. Schimanski, B.; Nguyen, T.N.; Günzl, A. Characterization of a Multisubunit Transcription Factor Complex Essential for Spliced-Leader RNA Gene Transcription in *Trypanosoma brucei*. *Mol. Cell. Biol.* **2005**, *25*, 7303–7313, doi:10.1128/mcb.25.16.7303-7313.2005.
73. Kramer, S.; Marnef, A.; Standart, N.; Carrington, M. Inhibition of mRNA maturation in trypanosomes causes the formation of novel foci at the nuclear periphery containing cytoplasmic regulators of mRNA fate. *J. Cell Sci.* **2012**, *125*, 2896–2909, doi:10.1242/jcs.099275.
74. Bhattacharyya, T.; Brooks, J.; Yeo, M.; Carrasco, H.J.; Lewis, M.D.; Llewellyn, M.S.; Miles, M.A. Analysis of molecular diversity of the *Trypanosoma cruzi* trypomastigote small surface antigen reveals novel epitopes, evidence of positive selection and potential implications for lineage-specific serology. *Int. J. Parasitol.* **2010**, *40*, 921–928, doi:10.1016/j.ijpara.2010.01.002.
75. Oury, B.; Tarrieu, F.; Monte-Alegre, A.; Ouassiss, A. *Trypanosoma cruzi*: Sequence polymorphism of the gene encoding the Tc52 immunoregulatory-released factor in relation to the phylogenetic diversity of the species. *Exp. Parasitol.* **2005**, *111*, 198–206, doi:10.1016/j.exppara.2005.07.001.
76. Zingales, B.; Andrade, S.G.; Briones, M.R.S.; A Campbell, D.; Chiari, E.; Fernandes, O.; Guhl, F.; Lages-Silva, E.; Macedo, A.M.; Machado, C.R.; et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: Second revision meeting recommends TcI to TcVI. *Memórias do Instituto Oswaldo Cruz* **2009**, *104*, 1051–1054, doi:10.1590/s0074-02762009000700021.
77. Callejas-Hernández, F.; Rastrojo, A.; Poveda, C.; Gironès, N.; Fresno, M. Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. *Sci. Rep.* **2018**, *8*, 1–13, doi:10.1038/s41598-018-32877-2.
78. Baptista, R.P.; Reis-Cunha, J.L.; DeBarry, J.D.; Chiari, E.; Kissinger, J.C.; Bartholomeu, D.C.; Macedo, A.M. Assembly of highly repetitive genomes using short reads: The genome of discrete typing unit III *Trypanosoma cruzi* strain 231. *Microb. Genom.* **2018**, *4*, e000156, doi:10.1099/mgen.0.000156.
79. Bradwell, K.R.; Koparde, V.N.; Matveyev, A.V.; Serrano, M.G.; Alves, J.M.P.; Parikh, H.; Huang, B.; Lee, V.; Espinosa-Alvarez, O.; A Ortiz, P.; et al. Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma rangeli* to *Trypanosoma cruzi* strains of high and low virulence. *BMC Genom.* **2018**, *19*, 770, doi:10.1186/s12864-018-5112-0.
80. Leinonen, R.; Sugawara, H.; Shumway, M.; on behalf of the International Nucleotide Sequence Database Collaboration. The Sequence Read Archive. *Nucleic Acids Res.* **2010**, *39*, D19–D21, doi:10.1093/nar/gkq1019.
81. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **2010**, *5*, e11147, doi:10.1371/journal.pone.0011147.
82. Hammer, Ø.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 1–9.
83. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.-F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469, doi:10.1093/nar/gkn180.
84. Trifinopoulos, J.; Nguyen, L.-T.; Von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235, doi:10.1093/nar/gkw256.
85. Sayers, E.W.; Cavanaugh, M.; Clark, K.; Ostell, J.; Pruitt, K.D.; Karsch-Mizrachi, I. GenBank. *Nucleic Acids Res.* **2020**, *48*, D84–D86.
86. Rambaut, A.; Drummond, A. FigTree version 1.4. 0. 2012. (accessed on 9 July 2020)