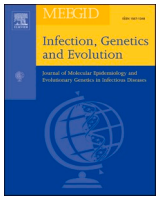
Infection, Genetics and Evolution 115 (2023) 105504

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Revisiting gene typing and phylogeny of *Trypanosoma cruzi* reference strains: Comparison of the relevance of mitochondrial DNA, single-copy nuclear DNA, and the intergenic region of mini-exon gene

Christian Barnabé [a,*], Simone Frédérique Brenière [a], Soledad Santillán-Guayasamín [a,b], Emmanuel J.P. Douzery [c], Etienne Waleckx [a,d,e]

[a] Institut de Recherche pour le Développement (IRD), UMR INTERTRYP IRD, CIRAD, University of Montpellier, Montpellier, France
[b] Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[c] Institut des Sciences de l'Évolution de Montpellier (ISEM), CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France
[d] Laboratorio de Parasitología, Centro de Investigaciones Regionales "Dr Hideyo Noguchi", Universidad Autónoma de Yucatán, Mérida, Mexico
[e] ACCyC, Asociación Chagas con Ciencia y Conocimiento, A. C, Orizaba, Mexico

ARTICLE INFO

Keywords:
Trypanosoma cruzi
DTU
Marker
Phylogeny
NGS
mini-exon

ABSTRACT

Chagas disease is a widespread neglected disease in Latin America. *Trypanosoma cruzi*, the causative agent of the disease, is currently subdivided into six DTUs (discrete typing units) named TcI-TcVI, and although no clear association has been found between parasite genetics and different clinical outcomes of the disease or different transmission cycles, genetic characterization of *T. cruzi* strains remains crucial for integrated epidemiological studies. Numerous markers have been used for this purpose, although without consensus. These include mitochondrial genes, single or multiple-copy nuclear genes, ribosomal RNA genes, and the intergenic region of the repeated mini-exon gene. To increase our knowledge of these gene sequences and their usefulness for strain typing, we sequenced fragments of three mitochondrial genes, nine single-copy nuclear genes, and the repeated intergenic part of the mini-exon gene by Next Generation Sequencing (NGS) on a sample constituted of 16 strains representative of *T. cruzi* genetic diversity, to which we added the corresponding genetic data of the 38 *T. cruzi* genomes fully sequenced until 2022.

Our results show that single-copy nuclear genes remain the gold standard for characterizing *T. cruzi* strains; the phylogenetic tree from concatenated genes (3959 bp) confirms the six DTUs previously recognized and provides additional information about the alleles present in the hybrid strains. In the tree built from the three mitochondrial concatenated genes (1274 bp), three main clusters are identified, including one with TcIII, TcIV, TcV, and TcVI DTUs which are not separated. Nevertheless, mitochondrial markers remain necessary for detecting introgression and heteroplasmy. The phylogenetic tree built from the sequence alignment of the repeated mini-exon gene fragment (327 bp) displayed six clusters, but only TcI was associated with a single cluster. The sequences obtained from strains belonging to the other DTUs were scattered into different clusters. Therefore, while the mini-exon marker may bring, for some biological samples, some advantages in terms of sensibility due to its repeated nature, mini-exon sequences must be used with caution and, when possible, avoided for *T. cruzi* typing and phylogenetic studies.

1. Introduction

Chagas disease is a parasitic disease caused by *Trypanosoma cruzi*, principally transmitted through the contaminated feces of blood-sucking insects of the Reduviidae family. Besides the vectorial route, the parasite can also be transmitted orally (food-borne), from a mother to her newborn (congenital), by blood transfusion, or organ transplantation. The disease is part of the "Neglected Tropical Diseases" group, while six to seven million people are estimated to be infected. Its geographical range is mainly Latin America, but it is found in other parts of the world,

* Corresponding author.
E-mail addresses: christian.barnabe@ird.fr (C. Barnabé), soledad.santillan@unil.ch (S. Santillán-Guayasamín), emmanuel.douzery@umontpellier.fr (E.J.P. Douzery), etienne.waleckx@ird.fr (E. Waleckx).

https://doi.org/10.1016/j.meegid.2023.105504
Received 12 July 2023; Received in revised form 12 September 2023; Accepted 19 September 2023
Available online 20 September 2023

mainly the USA and Europe, due to the migration of people from endemic areas (Albajar-Vinas and Jannin, 2011). *T. cruzi* exhibits extraordinary genetic diversity. Six genetic groups called DTUs (Discrete Typing Units) (Tibayrenc, 2003), consensually named TcI to TcVI (Zingales et al., 2012), are generally recognized. One seventh DTU can be added, Tcbat, which is closely related to TcI and firstly sampled in bats (Pinto et al., 2012). Notably, a distinct TcIV lineage named TcIV-USA (equivalent to TcIV North, Lopez-Dominguez et al., 2022) has also been proposed (Flores-Lopez et al., 2022). The evolution relationships between these DTUs are not fully elucidated, but TcV and TcVI are clearly hybrids between TcII and TcIII; TcIII and TcIV could also originate from ancient hybrids between TcI and TcII; finally, TcI and TcII appear as the ancestral lineages of all DTUs of *T. cruzi* (Breniere et al., 2016).

To explore the genetic diversity and identify these different genetic groups, many molecular markers have been used, such as the pioneering isoenzymes, mitochondrial genes, single or multiple copy nuclear genes, ribosomal RNA genes, and their internal transcribed spacer 2 (ITS2) region, as well as the intergenic sequence of the mini-exon gene, a tandemly repeated gene involved in the process of RNA splicing which is a singularity of Kinetoplastidae (Breniere et al., 2016).

Although to date, no strong and clear association has been demonstrated between genetic data and biological and medical properties of the parasite (Ragone et al., 2012; Monteiro et al., 2013), the genetic characterization of *T. cruzi* strains remains a major epidemiological element to be considered. This is particularly important in a framework of integrated epidemiology (Tibayrenc, 1998), taking into account the different actors of transmission cycles, such as host, vector, and parasite diversities (Orantes et al., 2018). Recently, amplicon-based Next-Generation Sequencing (metabarcoding) has proven to be a powerful and highly sensitive approach that allows the simultaneous identification of the transmission cycle components, using only from the vectors as biological material (Dumonteil et al., 2018; Hernández-Andrade et al., 2020). This approach consists in sequencing a large number of taxonomic-informative markers selected from these different above-mentioned biological compartments. As this method is based on the use of deep sequencing (Next Generation Sequencing, NGS), it allows for the sequencing of repeated elements of the genomes and, to some extent, determining the relative proportions of the different haplotypes of repeated sequences, unlike previous methods (such as Sanger-type sequencing), which are in practice limited to provide only the dominant haplotype of a repeated fragment of a genome.

The aim of our study was to evaluate the congruence of phylogenetic trees inferred from three kinds of markers based on sequences obtained from amplicon-based and whole genome sequencing NGS data and among the most used for *T. cruzi* population genetics and phylogeny: mitochondrial genes, single-copy nuclear genes and the intergenic region of the mini-exon gene. We show that mitochondrial genes and single-copy nuclear genes give relevant phylogenies compatible with the knowledge of the evolution and typing of *T. cruzi*. On the contrary, the phylogeny based on the intergenic region of the mini-exon gene is difficult to interpret and unable to allow unequivocal DTU identification.

2. Materials and methods

2.1. T. cruzi DNA samples

We selected 16 samples of *T. cruzi* DNA available in our lab, previously extracted by standard methods from epimastigote cultures of strains generally cloned in the laboratory, which have been used in various studies and which encompass most of the species variability. They included all DTUs (three TcI, four TcII, one TcIII, two TcIV, three TcV, and two TcVI), and one *Trypanosoma cruzi marinkellei* strain, known to be the *T. cruzi* most closely related taxon (Stevens et al., 1999), and that was used as outgroup in the different phylogenetic trees. Strain names, geographic and host origins are shown in Table 1.

2.2. PCR processing

Fifteen markers were amplified: i) three from mitochondrial genes: Cytochrome oxidase subunit 2 (*COII*), Cytochrome B (*CytB*), and NADH dehydrogenase subunit 1 (*ND1*); ii) eleven from putative single-copy nuclear genes (Yeo et al., 2011; Espinosa-Alvarez et al., 2018): Dihydrofolate reductase-thymidylate synthase (*DHFR-TS*), Glucose-6-phosphate isomerase (*GPI*), Glutathione peroxidase (*GPX*), 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGCR*), Aminopeptidase, putative (*AP*), Pyruvate dehydrogenase (*PDK*), Pumilio/PUF RNA binding protein (*RB*), Rho-like GTP binding protein, putative (*RHO1*), Iron superoxide dismutase A (*SODA*), Serine/threonine-protein phosphatase PP1, putative (*STPP*), and Trypanothione reductase (*TR*); and iii) the intergenic region of the trans-spliced leader mini-exon gene. Primers, fragment lengths, and annealing temperatures, as well as the chromosomic localization of the nuclear genes, are presented in Table 2. All PCRs followed the same process. The mixture was composed of 48 µl of Green Taq Mix buffer containing 1. 5 mM MgCl2, 0. 2 mM of dNTP mix, 0.4 µM of each primer; and 1 IU of Taq plus 2 µl of DNA sample (≈ 50 ng/µl). The cycle program started with an initial denaturation at 95 °C for 2 min, followed by 30 cycles (95 °C 30 s, annealing temperature 30 s, 72 °C 30 s), and finished with a final elongation at 72 °C 5 min followed by 4 °C 20 min.

2.3. Obtaining, selection and processing of NGS reads

The PCR products obtained for each marker were pooled for each sample. These pools were then cleaned up and sequenced on an NGS Illumina MiSeq V3 platform (2 × 300 pb), following library preparation for amplicon sequencing (dsDNA) by the company ABM (Richmond, Canada). The 301 bp reads were processed using VSEARCH v. 2. 17. 1 (Rognes et al., 2016), which is an open-source alternative to USEARCH (Edgar, 2010) for high-throughput search and clustering. All the reads were processed the same way. Briefly, for each sample, the processing steps were: a) merging of the two paired-end sequence reads (R1 and R2) into one merged sequence; b) sequence "dereplication", i.e., strictly identical sequences, defined as having the same length and the same

Table 1
Origin of the 16 *T. cruzi* strains sequenced in this study.

Strain	Country	Location	Host	DTU
B3	Brazil	Bahia, São Felipe	*Phyllostomus discolor*	*T. c. marinkellei*
OPS21cl11	Venezuela	Cojedes	*Homo sapiens*	TcI
P209cl93	Bolivia	Chuquisaca, Sucre	*Homo sapiens*	TcI
Tehcl2cl92	Mexico	Jalisco, Tehuantepec	*Triatoma* sp.	TcI
MAS1cl1	Brazil	Brasilia	*Homo sapiens*	TcII
TU18cl2	Bolivia	Potosi, Tupiza	*Triatoma infestans*	TcII
CBBcl3	Chile	Coquimbo, Tulahuen	*Homo sapiens*	TcII
Esmcl3Z2	Brazil	Bahia, São Felipe	*Homo sapiens*	TcII
X109/2	Paraguay	Makthlawaiya	*Canis familiaris*	TcIII
92122102R	USA	Georgia, Statesboro	*Procyon lotor*	TcIV
DogTheis	USA	uk*	*Canis familiaris*	TcIV
Bug2148cl1	Brazil	Rio Grande do Sul	*Triatoma infestans*	TcV
MNcl2	Chile	Coquimbo, Illapel	*Homo sapiens*	TcV
SC43cl1	Bolivia	Andres Ibañez, Santra Cruz	*Triatoma infestans*	TcV
CLBrener	Brazil	Rio Grande do Sul	*Triatoma infestans*	TcVI
Tulacl2	Chile	Coquimbo, Tulahuen	*Homo sapiens*	TcVI

* uk: unknown.

Table 2
Information of the 15 genes and corresponding primers used in this study and best model of multiple substitutions used for the construction of the trees.

Genes	Primers: name, sequence and melting temperature (Tm)	Length of PCR fragment		Annealing temperature for PCR	Best model of substitution[(1)]	[(2)]Localization on genome Brazil clone A4	[(2)]Localization on genome Y clone C6
		with primers	without primers				
Cytochrome oxidase subunit 2 (*COII*)	COIIn-F: AGTTTGTGGATACTTATTGGTTTGT (Tm = 56,4 °C) COIIn-R: CATAACTCGCTGCATTGYCC (Tm = 58,3 °C)	414 bp	369 bp	50 °C	TN93 + G	NR	NR
Cytochrome B (*CytB*)	CytB-lc-F: TGGTGTTTGYTTAGCRTGAATGT (Tm = 57,1 °C) CytB-lc-R: ACAAAAACAAAGTCKCTCRCA (Tm = 54,0 °C)	535 bp	491 bp	50 °C	HKY + G	NR	NR
NADH dehydrogenase subunit 1 (*ND1*)	ND1-lc-F: AGTAGTAYRAATGTTTGTGCAGTAG (Tm = 58,1 °C) ND1-lc-R: AGCAACYATGCGYACATTATTTT (Tm = 55,3 °C)	462 bp	414 bp	50 °C	HKY + G	NR	NR
Dihydrofolate reductase-thymidylate synthase (*DHFR-TS*)	Dhfr-F: ACTGCTGTGGTTTCTACGGG (Tm = 59,4 °C) Dhfr-R: AAGGGTGTGCACCAACTCTC (Tm = 59,4 °C)	496 bp	456 bp	54 °C	K2	Ch. 8 CM026590.1: 200196–200,691	Ch. 16 CM026641.1: 772431–772,926
Glucose-6-phosphate isomerase (*GPI*)	GPIF: CGCCATGTTGTGAATATTGG (Tm = 55,3 °C) GPIR: GGCGGACCACAATGAGTATC (Tm = 59,4 °C)	405 bp	365 bp	50 °C	T92	Ch. 20 CM026602.1: 546795–547,199	Ch. 35 CM026660.1: 226606–227,010
Glutathione peroxidase (*GPX*)	Gpx-F: TGTGTAGTCGTCAGCGTGGA (Tm = 59,4 °C) Gpx-R: CCAATTTAACCAGCGGGATGC (Tm = 59,8 °C)	451 bp	410 bp	54 °C	K2	Ch. 16 CM026598.1: 35664–36,114	Ch. 11 CM026636.1: 388462–388,912
3-hydroxy-3-methylglutaryl-CoA reductase (HMCOAR)	Hmcor-F: TGTTGCCGGGCCAATACTT (Tm = 56,7 °C) Hmcor-R: TCCACGTCCATCAATCCAGT (Tm = 57,3 °C)	487 bp	448 bp	54 °C	K2	Ch. 15 CM026597.1: 112948–113,434	Ch. 13 CM026638.1: 967955–968,441
Aminopeptidase, putative (*AP*)	Lap-F: CTACTGCCTACAAGCGGCAA (Tm = 59,4 °C) Lap-R: GTTTTGCGGAAAAGCTGCGA (Tm = 57,3 °C)	540 bp	500 bp	54 °C	JC + G	Ch. 1 CM026583.1: 1335852–1,336,391	Ch. 2 CM026627.1: 1239269–1,239,808
Pyruvate dehydrogenase (*PDH*)	Pdh-F: GATTCGTTGATGCGTTGCGT (Tm = 57,3 °C) Pdh-R: TTGCCTGGATGGCATCATCA (Tm = 57,3 °C)	490 bp	450 bp	54 °C	T92	Ch. 22 CM026604.1: 35148–35,637	Ch. 19: CM026644.1: 58050–58,499
Pumilio/PUF RNA binding protein (*RB*)	RB19-F: ACGAACGGCACCCGTTTTAC (Tm = 59,4 °C) RB19-R: AGTTCGTAGCCTCTTGGCCTG (Tm = 61,8 °C)	472 bp	431 bp	54 °C	K2 + G	Ch. 29 CM026611.1: 22594–23,065	Ch. 23 CM026648.1: 756941–757,412
Rho-like GTP binding protein, putative (*RHO1*)	Rho1-F: CGTCGCATGTGCTACACCAG (Tm = 61,4 °C) Rho1-R: GCGTACTGTGGCTGTCTCTG (Tm = 61,4 °C)	464 bp	424 bp	54 °C	K2	Ch. 1 CM026583.1: 421444–421,907	Ch. 20 CM026645.1: 597452–597,915
Iron superoxide dismutase A (*SODA*)	Sod-F: ATGGATGTGCCCCTGTCTTC (Tm = 59,4 °C) Sod-R: CCGGCGGTTTTCAAAGTCCT (Tm = 59,4 °C)	491 bp	451 bp	54 °C	K2	Chomosome 19 CM026601.1: 477240–477,730	Ch. 21 CM026646.1: 410308–410,798
Serine/threonine-protein phosphatase PP1, putative (*STPP*)		495 bp	455 bp	54 °C	K2 + G + I	Ch. 10 CM026592.1: 82835–83,329	Ch. 7 CM026632.1: 95493–94,999

(*continued on next page*)

Table 2 (*continued*)

Genes	Primers: name, sequence and melting temperature (Tm)	Length of PCR fragment		Annealing temperature for PCR	Best model of substitution[(1)]	[(2)]Localization on genome Brazil clone A4	[(2)]Localization on genome Y clone C6
		with primers	without primers				
	Sttp2-F: CACATCACACGGACGCTCTA (Tm = 59,4 °C) Sttp2-R: CGTGAGGAGGACATACGCAG (Tm = 61,4 °C)						
Trypanothione reductase (*TR*)	Tr-F: CCACACTCTCCGTGAGGAAC (Tm = 61,4 °C) Tr-R: TTGTGCATGAGCGGGGTAAA (Tm = 57,3 °C)	510 bp	470 bp	54 °C	JC	Ch. 4 CM026586.1: 753029–753,538	Ch. 4 CM026629.1800239–800,748
Trans-spliced leader mini-exon region (*ME*)	Tryp-ME3-F: TTCTGTACTATATTGGTACGCGAAG (Tm = 59,7 °C) TccH-R: CCCCCCTCCCAGGCCACACTGGG (Tm = 73,1 °C)	Around 480 bp		55 °C	TN93	Repeats on Ch. 23 and 40	Repeats on Ch. 10, 21 and 25

[(1)]best models of multiple substitutions: TN93 = Tamura-Nei; HKY = Hasegawa-Kishino-Yano; K2 = Kimura 2-parameter; T92 = Tamura 3-parameter; JC = Jukes-Cantor; G = discrete Gamma distribution and I = fraction of sites are evolutionarily invariable. [(2)] NR = not relevant for mitochondrial genes; Ch. = Chromosome.

nucleotide strings, are merged into a unique sequence to which a "size" corresponding to the observed abundance is attributed; c) clustering that performs denoising of Illumina amplicons according to the UNOISE algorithm (Edgar, 2016) with the alpha parameter set to 2 and the minimum size for clusters set to 8 in the first; d) removing potential chimera sequences using the UCHIME2 algorithm (Edgar et al., 2011); e) distribution of the clusters previously obtained for each sample into the different markers by performing a global alignment on the reference sequences of each marker and retaining the clusters whose identity was >0.7 for a given marker.

Finally, for a given sample and a given marker, only the clusters whose size was >10% of the total size of clusters were retained. Note that for all amplicons, the merging step gave correct results except for the mini-exon amplicon for which the merging score was very low. We thus used only the 301 bp R1 reads for the mini-exon. Note the Illumina technology used here is precise and reliable, but only allows sequencing fragments of <550 bp.

2.4. Concatenations of the mitochondrial and nuclear gene sequences

To increase the phylogenetic information and tree resolution, we concatenated i) the three mitochondrial gene fragments, and ii) the presumed single-copy gene fragments, using the function "cbind" of the R base package. It is worth noting that in the second case (ii), genotypes can be homozygous or heterozygous, especially in the DTUs TcV and TcVI. Due to the widely accepted fact that DTUs TcV-VI are true hybrids between TcII (i.e. alleles of type TcII) and TcIII (i.e. alleles of type TcIII) (Zingales et al., 2012), we concatenated together type II alleles on the one hand and type III alleles on the other hand. The lower number of heterozygotes within TcI, TcII, or TcIII have not generated any problem with allele phasing. The missing sequences in the concatenations were avoided as much as possible. However, in some cases, we had to introduce gaps in the final alignments. These gaps were treated as missing data in the different phylogenetic programs. The percentage of missing character states with the mitochondrial concatenation was <1% and did not exceed 5% with the nuclear concatenation. The intergenic region of the nuclear mini-exon gene was treated apart, being a sequence composed of tandem repetitions in the genome of *T. cruzi*.

2.5. Phylogenetic analyses

Analyses of each gene fragment involved the same three steps: i) Alignment: after removing the primers, the retained clusters from our samples were aligned using Muscle (Edgar, 2004) with the most similar sequences identified using Blastn against the 38 *Trypanosoma cruzi* WGS assemblies recorded until 2022 in NCBI databases except assembly ASM274941v1 (strain Bug2148), mislabeled TcV while it is actually TcI, and assembly ASM274942v1 (strain Y) mislabeled TcII while it is actually TcI as previously reported (Majeau et al., 2021; Probst et al., 2021), resulting in 36 assemblies. The *T. cruzi* WGS sequences were extracted and used as reference sequences (see list in Table S1); ii) Model selection: to better describe the DNA substitution pattern under the probability methods, the most adequate model was chosen according to the lowest BIC (Bayesian Information Criterion) score among the list of models included in MEGA7 (Kumar et al., 2016). For all tree constructions, the best model of substitutions was applied to each gene partition (Table 2), and iii) Tree inference: a maximum likelihood tree was first constructed by MEGA7, using the selected model and a statistical phylogeny test based on bootstraps of 100 replicates (Felsenstein, 1985).

For concatenated sequences, as well as for mini-exon fragments, two probability approaches were used to generate (i) a tree based on Bayesian Markov chain Monte Carlo (MCMC) analysis, giving the posterior probabilities of clustering at each node, with 10^6 MCMC cycles, transition and transversion rates set to beta (1,1), rates of reversible rate matrix accepting all GTR sub-models as equiprobable, parameter *Pi* fixed, parameter *alpha* set to exponential (1), partition-specific rate multiplier set to fixed (1), parameter *tau* accepting all topologies equally probable and branch lengths set to unconstrained, using MrBayes v.3.2.7a (Ronquist et al., 2012); and (ii) a maximum likelihood (ML) tree tested by a bootstrap method with 1000 replicates, using IQ-TREE v.2.2.0 (Minh et al., 2020). In all resulting trees, only statistically relevant values (posterior probability or bootstrap values) of >70% were reported to improve the readability of the trees.

3. Results

3.1. Mitochondrial phylogeny from three concatenated gene fragments

The concatenation of the amplified fragments of the three mitochondrial genes (*COII*, *CytB*, and *ND1* in that order) for 21 samples (12 concatenated sequences from 12 different strains newly sequenced in this work and nine concatenated sequences from WGS assemblies corresponding to nine different strains) resulted in an alignment of 1274 nucleotide positions and 16 different sequences. Properties and statistics of the alignment are presented in Table 3, including the percentage of variable sites, which reached 16.5%, and mean genetic diversity which

Table 3
Statistics for the three aligments under study: concatenated mitochondrial and nuclear gene fragments, and intergenic region of the mini-exon gene.

	Mitochondrial concatenation*	Nuclear concatenation*	Intergenic region of the mini-exon gene*
No. of gene fragments (abreviations of genes)	3 (*COII, CytB, ND1*)	9 (*DHFR-TS, GPI, GPX, HMCOAR, AP, PDH, RB, RHO1, TR*)	1 (*ME*)
No. of haplotypes	16	40	85
No. of nucleotide positions	1274	3959	327
No. of missing nucleotides	0 (0.0%)	7730 (4.9%)	0 (0.0%)
No. of gaps	14 (0.1%)	242 (0.2%)	4619 (16.6%)
No. of conserved sites	1064 (83.5%)	3690 (93.2%)	63 (19.3%)
No. of variable sites	210 (16.5%)	264 (6.7%)	249 (76.1%)
No. of parsimony informative sites	197 (15.5%)	246 (6.2%)	216 (66.1%)
No. of singletons	13 (1.0%)	18 (0.5%)	30 (9.2%)
Mean genetic diversity	0.053	0.015	0.274

* Values in parentheses correspond to the percentage relative to the total number of nucleotides.

was 0.053. The midpoint-rooted tree from the 16 different haplotypes (Fig. 1) showed three unambiguous clades supported by significant statistical values, corresponding to the three previously proposed clusters, and named mtTcI, mtTcII, and mtTcIII (Barnabé et al., 2016). The mtTcI cluster included the three strains belonging to TcI sequenced in this study and four other strains from genome assemblies belonging to TcI. Within mtTcI, a sub-cluster composed of strains sequenced in this study (OPS21 and Tehcl2) and others from WGS assemblies (Dm28c JRcl4, B. M. Lopez) was also identified. The mtTcII cluster included all TcII strains studied (one sequenced in this study and three from assemblies). The mtTcIII cluster included strains of TcIII, TcIV, TcV, and TcVI; the two TcIV North strains isolated in the United States (DogT and 92,122) were separated in a distinct sub-cluster from the other strains. Concerning the hybrid strains, it is worth noting that all the TcV strains (Bug 2148, MN, and SC43) presented the same haplotype and that all the TcVI strains were clustered into a group supported by high bootstrap values, 94 and 70 for Bayesian and ML tree respectively. The new mitochondrial sequences obtained from our sampling have been deposited in Genbank with the following accession numbers: *COII*, OR513510 - OR513522; *Cytb*, OR513523 - OR513535 and *ND1*, OR513536 - OR513546.

3.2. Nuclear phylogeny from nine concatenated single-copy gene fragments

Of the eleven presumed single-copy genes, nine that did not exhibit more than two different sequences (haplotypes) corresponding to the two expected alleles per sample (referred to as the diploid state of *T. cruzi*) were retained and concatenated in the following order: *DHFR-TS, GPI*, *GPX*, *HMGCR*, *AP*, *PDK*, *RB*, *RHO1*, and *TR*. The other two genes, *STPP* and *SODA*, were eliminated as they presented more than two haplotypes for most samples, evidencing they are not single-copy genes. The concatenation was possible for 31 samples (all the 16 strains sequenced in this study, and 15 strains from WGS assemblies). It resulted in an alignment of 3959 nucleotide positions for 50 concatenates, which exhibited 40 different concatenated haplotypes with no >5% of missing data. As presented in Table 3, the percentage of variable sites (6. 7%) and mean genetic variability (0.015) were lower than those observed for mitochondrial genes. The ML tree rooted by *T. c. marinkellei* showed four statistically supported clades that we named nucTcI, nucTcII, nucTcIII, and nucTcIV (Fig. 2). As shown by the branch lengths of the phylogram, the inter-clade nuclear DNA divergence was high whereas the intra-clade divergence was very low. NucTcI clustered all the concatenated haplotypes (equated to alleles) from TcI strains without haplotypes from other DTUs. NucTcII clustered all the

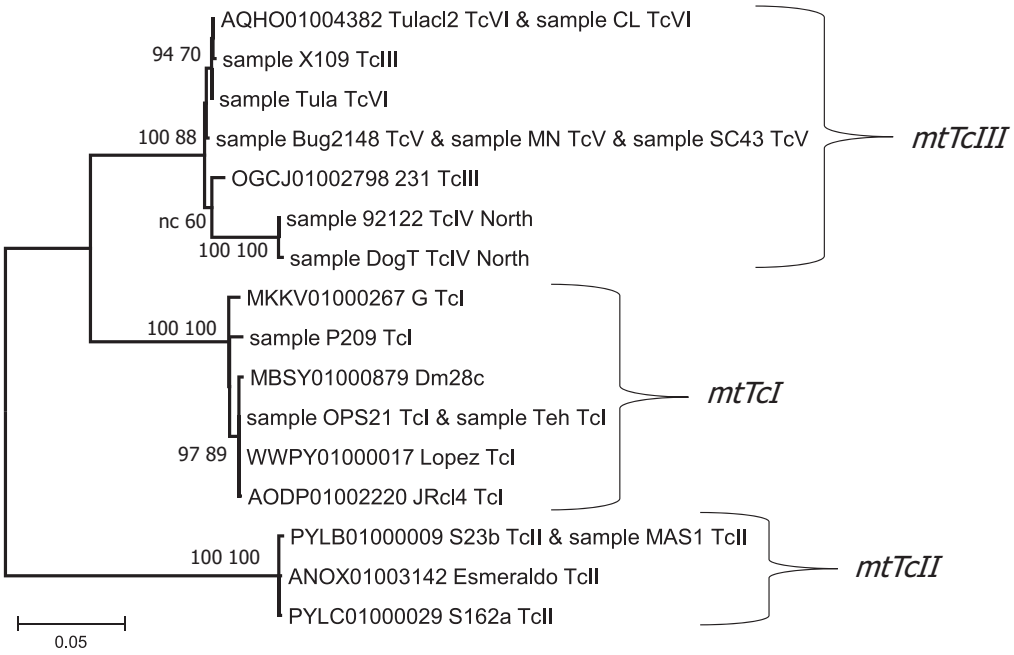


Fig. 1. Maximum likelihood (ML) tree of the 16 different haplotypes (on 1274 nucleotide positions, see alignment in File S2 (A)) of *T. cruzi* resulting from the concatenation of three mitochondrial gene fragments *COII*, *CytB*, and *ND1*, and coming from 21 DNA sequences in total: 12 obtained by NGS sequencing of our samples (leaf name beginning by "sample") and nine from BLAST hits against WGS assemblies of *Trypanosoma cruzi*. The numbers at the nodes are 1) the posterior probability of the Bayesian analysis, and 2) the bootstrap value for the ML tree; "nc" (not clustered) indicates that this particular cluster was absent in the analysis with the corresponding method.

Fig. 2. Maximum likelihood (ML) tree of the 40 different haplotypes (3959 nucleotide positions, see alignment in File S2 (B))) resulting from, the concatenation of nine single-copy nuclear gene fragments (*DHFR-TS*, *GPI*, *GPX*, *HMGAR*, *AP*, *PDK*, *RB*, *RHO1*, and *TR)* for 31 strains: 16 presently sequenced by NGS sequencing (leaf name beginning by "sample") and from BLAST hits against 15 WGS assemblies of *Trypanosoma cruzi*. The numbers at the nodes are 1) the posterior probability of the Bayesian analysis, and 2) the bootstrap value for the ML tree. The terms h1, 2, 3, …n correspond to the different haplotypes for a specific sample, if any.

concatenated haplotypes from TcII strains and part of TcV and TcVI strains; within this nucTcII cluster, a sub-cluster composed only by haplotypes from TcVI and supported by significant statistical values (ranging from 96 to 99) was identified. NucTcIII clustered all the concatenated haplotypes from TcIII strains and part of TcV and TcVI strains, but interestingly, there was a clear-cut differentiation between the haplotypes from TcIII strains (231 and X109), TcV strains (SC43, Mn, and Bug2148, reduced to a single haplotype), and the six haplotypes from TcVI strains (Tulacl2, CL, and TCC). NucTcIV clustered the two concatenated haplotypes from TcIV North strains both isolated in the USA. Note that TcIV strains appeared in a sister cluster of TcIII strains in the mitochondrial tree whereas it was a sister group of nucTcI + nucTcII in the nuclear tree. The new nuclear sequences obtained from our sampling have been deposited in Genbank with the following accession

numbers: *DHFR-TS*, OR513547 - OR513569; *GPI*, OR513570 - OR513587; *GPX*, OR513588 - OR513612; *HMGCR*, OR513613 - OR513636; *AP*, OR513637 - OR513662; *PDK*, OR513663 - OR513684; *RB*, OR513685 - OR513704; *RHO1*, OR513705 - OR513729; *TR*, OR513730 - OR513751; *SODA*, OR532281 - OR532309 and *STPP*, OR532310 - OR532374. and mini-exon repeats, OR532240 - OR532266.

3.3. Construction of a phylogenetic tree using the intergenic region of the mini-exon gene

After processing the NGS reads with the usual pipeline, which begins by the merging of the two paired-end reads (R1 and R2), clusters of sequences corresponding to the mini-exon gene were obtained for only seven strains out of 16. Consequently, the analysis was performed again using only the 301 pb R1 reads, which were obtained for 13 strains out of 16. Applying the selection criterion of 10% (i.e., retaining only clusters whose abundance is >10% of the total abundance for a given strain and a given locus), up to five clusters per strain were retained, resulting in a total of 27 clusters. The BLASTn query (with options query cover >90% and identity >90%) of nine sequences of mini-exon from TcI, TcII, TcIII, TcV, and TcVI strains representatives of the whole diversity of *T. cruzi* (278 bp each, see File S1) against the 36 *T. cruzi* WGS assemblies from NCBI, resulted in 2132 sequences in total ranging from zero for 12 WGS projects (no significant similarity for these parameters) up to 712 for the PRFA01project (strain Dm28c, TcI). For the construction of the phylogenetic tree, this set of sequences from WGS assemblies was reduced in number by clustering the sequences into centroid sequences. A total of 50 centroid sequences was obtained (the number of centroids by project ranging from one to six, see Table S1). Additionally, to increase the number of sequences from TcIII and TcIV strains, which are drastically under-represented in the WGS, seven mini-exon sequences from the nucleotide NCBI database including sequences from TcIII and TcIV were added. The alignment of 327 nucleotide positions with a total of 85 sequences was processed for the construction of the ML tree (Fig. 3). Among the properties and statistics of the alignment (Table 3), it is important to note the low proportion of conserved sites (below 20%), associated with a mean genetic diversity (0. 274) much higher than that of the other two kinds of genes studied. The ML tree showed a strong structuring into six relevant clades we named MeA to MeF: (i) cluster MeA grouped all the 22 sequences from 11 TcI strains and two other sequences from TcII (Esmeraldo) and TcVI (Tula cl2) strains; (ii) cluster MeB grouped four sequences from two TcIII strains; (iii) cluster MeC included the three TcIV North sequences from USA while the other sequence of TcIV (strain CANIII from Brazil) was not included in any cluster; (iv) cluster MeD included six sequences from four TcV strains, seven sequences from three TcVI strains, and one from a TcIII strain; (v) cluster MeE included eight sequences from seven TcII strains, two from two TcVI strains, and one from a TcIII strain, this other TcIII sequence was identified in cluster MeD (strain Ikiakarora); (vi) cluster MeF included 12 sequences from six TcII strains, seven sequences from two TcV strains and eight sequences from four TcVI strains. The distribution of the mini-exon sequences obtained from the six DTUs is also presented in Fig. 4 according to the MeA to MeF clusters. It clearly shows that several sequences belonging to different DTUs can be grouped together in the same mini-exon cluster. Conversely, a single DTU (e.g., TcII, TcIII, TcIV, TcV, or TcVI) can be characterized by sequences belonging to different mini-exon clusters. Additionally, some interesting features of the tree are worth noting: (i) one sequence from the TcVI strain CL Brener (MKQG01004719. 1 CL q4 TcVI) was found out of any cluster, while another one belonged to MeE and three others to MeF; (ii) only the sequences from TcI were clustered into a same cluster (MeA) whereas the sequences from other the DTUs were scattered into different clusters; (iii) most of the sequences from TcII strains belonged to MeE and MeF except for the strain Esmeraldo, for which one haplotype belonged to MeA while another one belonged to MeE, both sequences being obtained from the same WGS assembly genome project (ANOX01); (iv) an unexpected distribution was observed for the TcIII sequences, indeed sequences from strain Ikiakarora were found in two distinct clusters (MeD and MeE) while those of M5631 and X109 were found in the very distant MeB cluster; (v) while most of the hybrid strains (TcV and TcVI) were scattered in two clusters (MeE and MeF), surprisingly the strain Tula had a sequence clustered in MeA with all the TcI sequences (see Figs. 3 and 4). The new mini-exon intergenic region sequences, obtained from our sampling, have been deposited in Genbank: OR532240 - OR532266.

4. Discussion

4.1. Mitochondrial variation of T. cruzi and DTU identification

Mitochondrial DNA (mtDNA) has long been used as a reference phylogenetic marker because of its properties such as its uniparental transmission, absence of recombination, quasi-neutrality of its coding genes, and probably a molecular clock-like regular rate (Dong et al., 2021), although some of these properties may be questionable (Galtier et al., 2009). The high variability of mtDNA joined to the ease of its PCR amplification due to its multiple identical copies in the cell generally allow both to construct relevant phylogenetic trees and to characterize clusters at the sub-specific level. Besides a lot of other markers reviewed by others (Messenger et al., 2015; Izeta-Alberdi et al., 2016), mtDNA was also widely used with *T. cruzi*, even recently to characterize Panamanian strains of *T. cruzi* (Calzada et al., 2022). In the present work, NGS applied to three mitochondrial genes allowed to explore the existence of heteroplasmy (mitochondrial polymorphism in a single strain) within some strains, bringing a plus compared to traditional sequencing (Messenger et al., 2012). The phylogenic tree including the six *T. cruzi* DTUs showed three significant clusters (mtTcI, mtTcII, and mtTcIII, Fig. 1), as previously found and labeled (Machado and Ayala, 2001; Barnabé et al., 2016), in contrast to the six clusters proposed for the nuclear phylogeny. Indeed, as previously stated, TcV and TcVI DTUs arise from at least one recent hybridization events involving TcII and TcIII, the latter being considered as ancestral lineages, together with TcI (de Freitas et al., 2006). The uniparental transmission of mtDNA has systematically been observed from TcIII DTU, without any known explanation for this pattern (Berna et al., 2021; Garcia et al., 2017; Gaunt et al., 2003; Lewis et al., 2011), and we confirm here this pattern, with the cluster mtTcIII containing TcIII, TcIV, TcV and TcVI strains. Because the mtTcIII cluster grouped strains belonging to different DTUs, we also looked for a potential sub-structuring. Some sub-structuring was observed: i) the two TcIV North strains from the USA were grouped together in a significant sub-cluster as previously reported by Flores-Lopez et al. (2022) who proposes TcIV North (or TcIV-USA) as a separated lineage, a result consistent compared with other data (Lima-Cordon et al., 2021) which clearly showed an independent clade named $mtTcIII_{NA\text{-}CA}$ in Central and North America, distinct from the mtTcIII observed in South America, ii) the TcVI strains and the Paraguayan strain X109 (TcIII) were also clustered together, while iii) the TcV strains did not present variability and the unique corresponding sequence was apart from the two other detected sub-clusters. Moreover, the separate position of the two other TcIII strains (231 and X109) suggests that the TcIII strains might also be subdivided. These results suggest that the concatenation of more mitochondrial fragments for a higher number of strains, particularly from North and Central America underrepresented here, belonging to the different DTUs, would be wise to better identify sub-structuring in the TcIII cluster, and maybe discriminate the different DTUs in this cluster. Note that mitochondrial introgression has been described in several *T. cruzi* strains (observation of genetic incongruence between the mitochondrial and nuclear genomes). Although rare, this can generate a misidentification of DTUs in some *T. cruzi* strains when mitochondrial sequencing is used alone (Barnabé and Brenière, 2012; Messenger et al., 2012) and the simultaneous sequencing of at least one nuclear gene is necessary to identify potential states of introgression

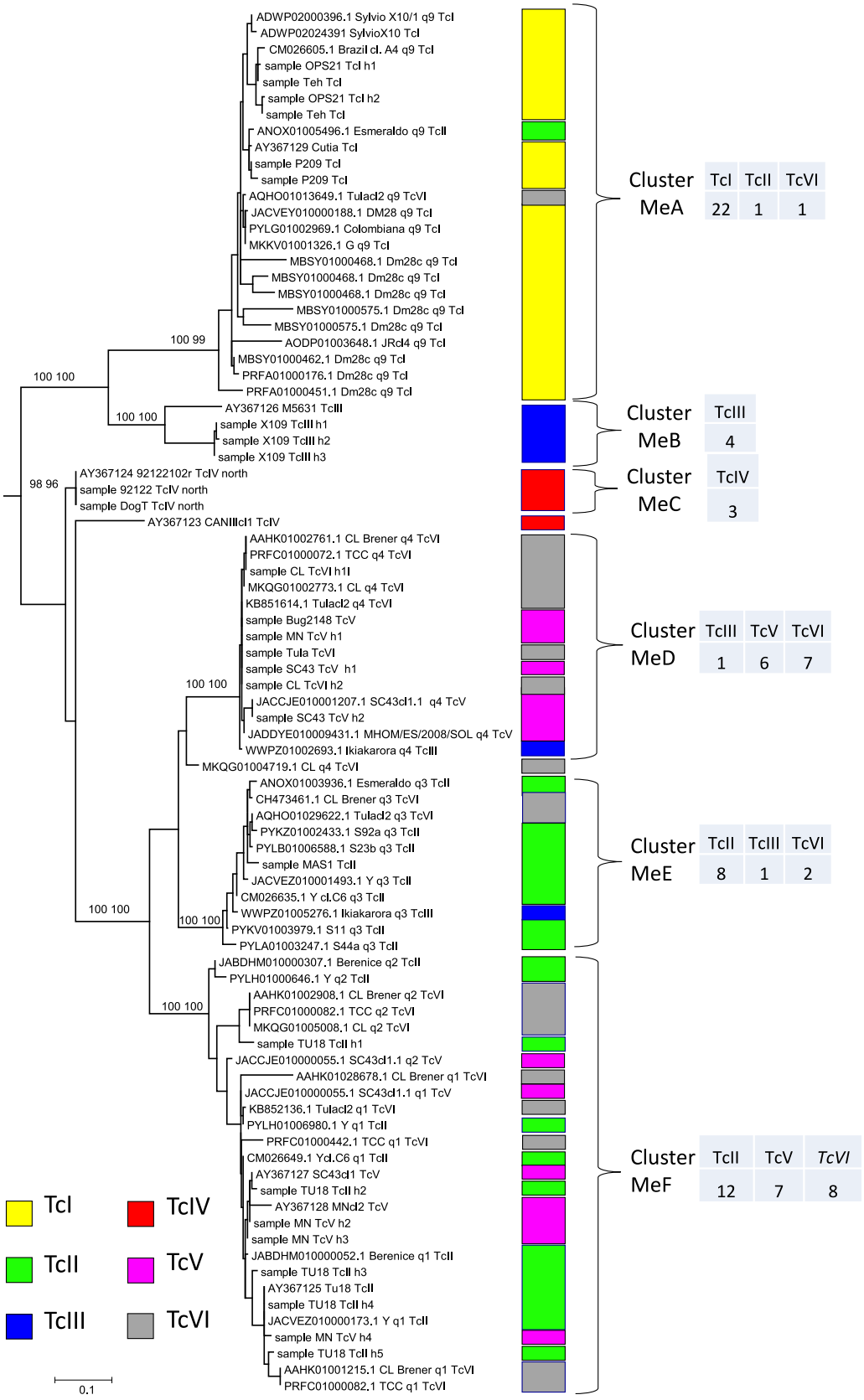
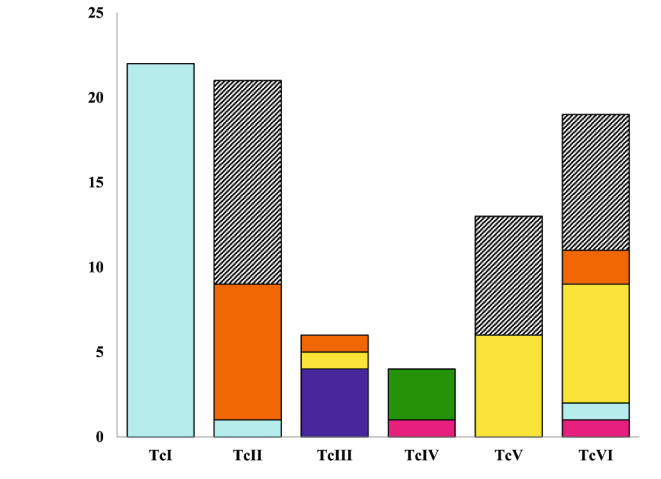


Fig. 3. Maximum likelihood (ML) tree involving 85 mini-exon intergenic sequences coming from our NGS sequencing (leaf names beginning by "sample"), from BLAST hits against the 36 WGS of *Trypanosoma cruzi*, and seven informative mini-exon sequences from the nucleotide NCBI database, including sequences from TcIII and TcIV which are drastically under-represented in the WGS. See alignment in File S2 (C). The numbers at the nodes are 1) the posterior probability of the Bayesian analysis, and 2) the bootstrap value for the ML tree. The terms q1, 2, 3,..0.9 in sequence names refer to the best query hits obtained with BLASTn using the corresponding nine mini-exon reference sequences (File S1). The terms h1, 2, 3, …n correspond to the different haplotypes for a specific sample, if any.



	TcI	TcII	TcIII	TcIV	TcV	TcVI
MeA	22	1	0	0	0	1
MeB	0	0	4	0	0	0
MeC	0	0	0	3	0	0
MeD	0	0	1	0	6	7
MeE	0	8	1	0	0	2
MeF	0	12	0	0	7	8
Unclustered	0	0	0	1	0	1

Fig. 4. Distribution of the 85 haplotypes of mini-exon intergenic sequences through the six clusters (A to F) obtained on the corresponding clustering tree (see Fig. 3) and the six DTUs of *Trypanosoma cruzi* (TcI to TcVI). Note that only haplotypes from TcI and TcIV were present in a single mini-exon cluster, (MeA and MeC, respectively); all haplotypes from other DTUs were present in two or more mini-exon clusters.

(Rodrigues et al., 2017).

4.2. Nuclear variation based on nuclear single-copy genes of T. cruzi, and DTU identification

The nuclear genome is an almost inexhaustible source of diverse genetic markers widely used for all kinds of organisms including *T. cruzi* (Izeta-Alberdi et al., 2016). We agree with the following sentence previously stated in (Eberle et al., 2020): "We argue that nearly universal single-copy nuclear protein-coding genes deliver the desired characteristics and could be used to reliably delimit and identify animal species", and we think that this assertion can be extended to protists. The MLST (Multilocus Sequence Typing) method was already proposed in 1998 to identify clonal bacterial pathogens (Maiden et al., 1998). This high-resolution genetic typing was successful in numerous applications: molecular epidemiology, taxonomy, phylogenetics, and population structure and dynamics (Perez-Losada et al., 2013). With *T. cruzi*, the relevance of using MLST, even with diploid organisms, has been demonstrated (Yeo et al., 2011) and MLST has been used successfully in several studies (Lauthier et al., 2012; Ramirez et al., 2013; Tomasini et al., 2013; Diosque et al., 2014; Tomasini et al., 2014). In fact, a team working on *T. cruzi* has created a program (MLSTest) that specifically manages MLST data from diploid organisms (Tomasini et al., 2013). Unfortunately, unlike the mitochondrial genome, the concatenation of MLST nuclear fragments is impossible in the case of highly recombinant organisms. Nevertheless, concatenation is possible in certain organisms with predominant clonal evolution as suggested for *T. cruzi* (Tibayrenc and Ayala, 2015). Moreover, as stated in the previous section, the association of nuclear and mitochondrial typing should be mandatory to detect scarce but possible events of mitochondrial introgression (Messenger et al., 2012). Here, the concatenation of nine fragments of *T. cruzi* single-copy genes resulted in a tree exhibiting four haplotype clusters, namely NucTcI, NucTcII, NucTcIII, and NucTcIV, which actually corresponds to TcIV-USA described by Flores-Lopez et al. (2022) in the absence of TcIV from other regions. These clusters included the strains belonging to their respective DTUs. The hybrid strains belonging to TcV and TcVI presented alleles from their parental DTUs TcII and TcIII (Lewis et al., 2011), one allele clustering in NucTcII and the other clustering in NucTcIII cluster; the identification of the hybrid strains being therefore unequivocal. Interestingly, some sub-clusters appeared in NucTcII and NucTcIII: in NucTcII, the alleles of the three TcVI strains under study formed together a significant sub-cluster, and in NucTcIII, the alleles from TcIII strains were significantly separated from those of TcV and TcVI, suggesting a divergent evolution between hybrids and parental lines which could be confirmed using longer concatenations on more strains in future investigations.

4.3. Variation of the intergenic region of the mini-exon gene and DTU identification

The use of mini-exon for DTU identification became popular from the publication reporting a multiplex PCR method able to separate *Trypanosoma rangeli*, and three clades of *T. cruzi* (Fernandes et al., 2001) by a mere electrophoresis of the amplicons that we validated as others (Aliaga et al., 2011): *T. rangeli*, ~100 bp; TcI, ~200 bp, TcII-V-VI, ~250 bp and finally TcIII-IV ~150 bp. The ease of amplification of this marker, including in biological samples, in part thanks to its repeated nature in the genome of *T. cruzi*, has been very useful and numerous studies dedicated to *T. cruzi* characterization have associated this multiplex PCR with other markers, either nuclear (often 18S or 24S rRNA genes) or mitochondrial (often COII or CytB) with relevance. When the use of DNA sequencing has become widespread, and more recently with the development of NGS technologies, the use of the sequences of the intergenic region of the mini-exon gene as the only reference marker of DTU has appeared in various studies (see for example Herrera et al., 2007; Falla et al., 2009; Herrera et al., 2009; Lisboa et al., 2009; Garcia et al., 2017; Herrera et al., 2019a: Herrera et al., 2019b; Majeau et al., 2019; Villanueva-Lizama et al., 2019; Dumonteil et al., 2018; Dumonteil et al., 2020; Maiguashca Sanchez et al., 2020; Pronovost et al., 2020; Murillo-Solano et al., 2021; Polonio et al., 2021). The current study, however, shows that the intergenic region of the mini-exon gene presents some pitfalls and difficulties for *T. cruzi* phylogenetic construction and DTU identification (cf. Figs. 3-4). The first highlighted difficulty, when used with Illumina paired-end reads, is the impossibility of obtaining a correct merging between the R1 and R2 sequences because of poly-T and microsatellite motifs present in this kind of sequences (O'Connor et al., 2007; Herrera et al., 2013). This restricted us to focus only on the R1 sequences of 301 bp, instead of using the entire amplified fragment (around 500 bp), thus resulting in reduced genetic information. Another difficulty was the high variability of these sequences due also to the presence of poly-T and microsatellite regions, and to frequent insertions and deletions (Grisard et al., 1999); this complexity led to a questionable alignment of sequences between DTUs. In the present study, the number of conserved nucleotide positions reached only 19.3%. Moreover, the constituent elements and sequence repeats of the intergenic region probably present their own pattern of evolution that is likely to complicate the phylogenetic reconstruction. A third difficulty is the variation of the number of haplotypes within the same strain (up to five in the current study), as well as their high variability. Moreover, the BLASTn results of mini-exon queries against each one of the 36 WGS assemblies gave unexpected and puzzling results. Indeed, with identical restrictive parameters in terms of identity and minimum query cover, BLAST hits ranged from 0 for many WGS to 712 for the TcI strain Dm28c. Although some variation in the number of copies is expected between *T. cruzi* genomes with such a high level of plasticity and chromosomal copy number variation (Reis-Cunha et al., 2015), these enormous observed variations may be better explained by the ongoing difficulty of

taking into account repeated elements in NGS sequencing (Treangen and Salzberg, 2011). Moreover, we can suspect that copy number variations generated by differential duplications and losses of mini-exon copies among different strains involve the comparison of paralogous sequences and yield to incongruences between the phylogenic trees estimated from mini-exon sequences and from other nuclear and/or mitochondrial genomic regions.

Beyond the pitfalls already reported, the current study rigorously conducted on a sample of strains representative of the different DTUs clearly highlights the existence of sequence clusters, but the corresponding clades do not fully correspond to the DTU classification currently adopted. With the notable exception of the sequences from TcI strains and those from TcIV North strains from the USA, which are grouped in the clusters MeA and MeC, respectively. The sequences of the other DTU strains are unexpectedly scattered into different mini-exon clusters. Indeed, sequences from TcII and TcIII can be found in three different clusters, and those from TcV and TcVI in two and four clusters, respectively. Clearly, such dispersion of sequences among clusters does not allow a non-risky identification of the majority of *T. cruzi* DTUs using mini-exon sequences, as previously suggested (Lewis et al., 2009; Maiguashca Sanchez et al., 2020). Fortunately or by chance, the mini-exons sequencing was mainly used for DTU determination in geographical areas where TcI and/or TcIV were the only ones present, or explicitly for typing TcI (Herrera et al., 2007; O'Connor et al., 2007; Falla et al., 2009; Herrera et al., 2009; Zuriaga et al., 2012; Garcia et al., 2017; Herrera et al., 2019b) or to differentiate between TcI from the other DTUs (Dumonteil et al., 2018). Other uses involving DTUs other than TcI and TcIV seem to us risky (for example, see Herrera et al., 2019a; Pronovost et al., 2020), as well as the proposition to use the mini-exon sequences as universal markers for any DTU (Majeau et al., 2019).

5. Concluding remarks

The ability to distinguish the genetic groups/clusters that compose a species with great biological variability, such as *T. cruzi*, is essential. Indeed, these clusters, which reflect genetic divergence during evolution, may group together strains with similar biological and ecological properties, different from the strains of the other clusters, and have a possible impact on the epidemiology and evolution of the disease. *T. cruzi* is a like-diploid species with high genome plasticity and evolutionary complex history including hybridization events, aneuploidy, ability to recombine but preponderantly clonal characteristics (Lewis et al., 2011; Tibayrenc and Ayala, 2015). In this context, the choice of the best evolutionary marker to understand genetic relationships between strains is particularly tricky. Originally, the six *T. cruzi* DTUs, including two hybrids, have been proposed based on the analysis of two nuclear markers, RAPD and MLEE (Barnabé et al., 2000; Brisse et al., 2000; Tibayrenc, 2003). Subsequently, nuclear gene sequencing did not invalidate this proposition: on the contrary, it deepened the knowledge of the genetic origin of the hybrid strains (Messenger et al., 2015; Matos et al., 2022). With the availability of NGS data, the population structure of *T. cruzi* is well confirmed by the nuclear genes analysis. Indeed, single-copy genes have simple and known transmission rules that can be largely applied to *T. cruzi*, while this is not the case for the intergenic sequences belonging to highly repeated genes such as the mini-exon gene. The significant incongruence between the trees constructed with the sequences of the nuclear genes and those of the intergenic region of the mini-exon gene suggests that the sequencing of the intergenic region of the mini-exon gene to unequivocally characterize *T. cruzi* strains should be avoided. Mitochondrial gene sequencing remains valid for detecting introgression phenomena and heteroplasmy, but remains a limited tool to distinguish between TcIII, TcV, TcVI, and to some extent TcIV. Indeed, a sub-structuring of the mitochondrial groups cannot be excluded, and the current growing practice of NGS could lead to new interest of this marker.

Funding

This work received financial support from CONACYT (National Council of Science and Technology, Mexico) Basic Science (Project ID: CB2015–258752) and National Problems (Project ID: PN2015–893) Programs attributed to EW, and from the French National Research Institute for Sustainable Development (IRD, France).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2023.105504.

References

Albajar-Vinas, P., Jannin, J., 2011. The hidden Chagas disease burden in Europe. Euro Surveill. 16 (38).
Aliaga, C., Brenière, S.F., Barnabé, C., 2011. Further interest of miniexon multiplex PCR for a rapid typing of *Trypanosoma cruzi* DTU groups. Infect. Genet. Evol. 11 (5), 1155–1158.
Barnabé, C., Brenière, S.F., 2012. Scarce events of mitochondrial introgression in *Trypanosoma cruzi*: new case with a Bolivian strain. Infect. Genet. Evol. 12 (8), 1879–1883.
Barnabé, C., Brisse, S., Tibayrenc, M., 2000. Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas disease: a multilocus enzyme electrophoresis approach. Parasitology 120 (Pt 5), 513–526.
Barnabé, C., Mobarec, H.I., Jurado, M.R., Cortez, J.A., Brenière, S.F., 2016. Reconsideration of the seven discrete typing units within the species *Trypanosoma cruzi*, a new proposal of three reliable mitochondrial clades. Infect. Genet. Evol. 39, 176–186.
Berna, L., Greif, G., Pita, S., Faral-Tello, P., Diaz-Viraque, F., Souza, R.C.M., Vallejo, G.A., Alvarez-Valin, F., Robello, C., 2021. Maxicircle architecture and evolutionary insights into *Trypanosoma cruzi* complex. PLoS Negl. Trop. Dis. 15 (8) e0009719.
Breniere, S.F., Waleckx, E., Barnabé, C., 2016. Over six thousand *Trypanosoma cruzi* strains classified into discrete typing units (DTUs): attempt at an inventory. PLoS Negl. Trop. Dis. 10, e0004792.
Brisse, S., Barnabé, C., Tibayrenc, M., 2000. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. Int. J. Parasitol. 30 (1), 35–44.
Calzada, J.E., Samudio, F., de Junca, C., Pineda, V., Burleigh, B.A., Saldana, A., 2022. Genetic diversity of *trypanosoma cruzi* in panama inferred by multi-locus sequence typing of mitochondrial genes. Microorganisms 10 (2).
de Freitas, J.M., Augusto-Pinto, L., Pimenta, J.R., Bastos-Rodrigues, L., Goncalves, V.F., Teixeira, S.M., Chiari, E., Junqueira, A.C., Fernandes, O., Macedo, A.M., Machado, C. R., Pena, S.D., 2006. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. PLoS Pathog. 2 (3), e24.
Diosque, P., Tomasini, N., Lauthier, J.J., Messenger, L.A., Monje Rumi, M.M., Ragone, P. G., Alberti-D'Amato, A.M., Perez Brandan, C., Barnabé, C., Tibayrenc, M., Lewis, M. D., Llewellyn, M.S., Miles, M.A., Yeo, M., 2014. Optimized multilocus sequence typing (MLST) scheme for *Trypanosoma cruzi*. PLoS Negl. Trop. Dis. 8 (8) e3117.
Dong, Z.K., Wang, Y.Z., Li, C., Li, L.L., Men, X.Y., 2021. Mitochondrial DNA as a molecular marker in insect ecology: current status and future prospects. Ann. Entomol. Soc. Am. 114 (4), 470–476.
Dumonteil, E., Ramirez-Sierra, M.J., Perez-Carrillo, S., Teh-Poot, C., Herrera, C., Gourbiere, S., Waleckx, E., 2018. Detailed ecological associations of triatomines revealed by metabarcoding and next-generation sequencing: implications for triatomine behavior and *Trypanosoma cruzi* transmission cycles. Sci. Rep. 8 (1), 4140.
Dumonteil, E., Elmayan, A., Majeau, A., Tu, W., Duhon, B., Marx, P., Wolfson, W., Balsamo, G., Herrera, C., 2020. Genetic diversity of *Trypanosoma cruzi* parasites infecting dogs in southern Louisiana sheds light on parasite transmission cycles and serological diagnostic performance. PLoS Negl. Trop. Dis. 14 (12) e0008932.
Eberle, J., Ahrens, D., Mayer, C., Niehuis, O., Misof, B., 2020. A plea for standardized nuclear markers in metazoan DNA taxonomy. Trends Ecol. Evol. 35 (4), 336–345.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26 (19), 2460–2461.

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv, 081257.
Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27 (16), 2194–2200.
Espinosa-Alvarez, O., Ortiz, P.A., Lima, L., Costa-Martins, A.G., Serrano, M.G., Herder, S., Buck, G.A., Camargo, E.P., Hamilton, P.B., Stevens, J.R., Teixeira, M.M.G., 2018. *Trypanosoma rangeli* is phylogenetically closer to Old World trypanosomes than to *Trypanosoma cruzi*. Int. J. Parasitol. 48 (7), 569–584.
Falla, A., Herrera, C., Fajardo, A., Montilla, M., Vallejo, G.A., Guhl, F., 2009. Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. Acta Trop. 110 (1), 15–21.
Felsenstein, J., 1985. Confidence-limits on phylogenies - an approach using the bootstrap. Evolution 39 (4), 783–791.
Fernandes, O., Santos, S.S., Cupolillo, E., Mendonca, B., Derre, R., Junqueira, A.C., Santos, L.C., Sturm, N.R., Naiff, R.D., Barret, T.V., Campbell, D.A., Coura, J.R., 2001. A mini-exon multiplex polymerase chain reaction to distinguish the major groups of *Trypanosoma cruzi* and *T. rangeli* in the Brazilian Amazon. Trans. R. Soc. Trop. Med. Hyg. 95 (1), 97–99.
Flores-Lopez, C.A., Mitchell, E.A., Reisenman, C.E., Sarkar, S., Williamson, P.C., Machado, C.A., 2022. Phylogenetic diversity of two common *Trypanosoma cruzi* lineages in the southwestern United States. Infect. Genet. Evol. 99, 105251.
Galtier, N., Nabholz, B., Glemin, S., Hurst, G.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. 18 (22), 4541–4550.
Garcia, M.N., Burroughs, H., Gorchakov, R., Gunter, S.M., Dumonteil, E., Murray, K.O., Herrera, C.P., 2017. Molecular identification and genotyping of *Trypanosoma cruzi* DNA in autochthonous Chagas disease patients from Texas, USA. Infect. Genet. Evol. 49, 151–156.
Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S. S., Veazey, P., Miles, G.A., Acosta, N., de Arias, A.R., Miles, M.A., 2003. Mechanism of genetic exchange in American trypanosomes. Nature 421 (6926), 936–939.
Grisard, E.C., Campbell, D.A., Romanha, A.J., 1999. Mini-exon gene sequence polymorphism among *Trypanosoma rangeli* strains isolated from distinct geographical regions. Parasitology 118 (Pt 4), 375–382.
Hernández-Andrade, A., Moo-Millan, J., Cigarroa-Toledo, N., Ramos-Ligonio, A., Herrera, C., Bucheton, B., Bart, J.-M., Jamonneau, V., Bañuls, A.-L., Paupy, C., Roiz, D., Sereno, D., Ibarra-Cerdeña, C.N., Machaín-Williams, C., García-Rejón, J., Gourbière, S., Barnabé, C., Telleria, J., Oury, B., Brenière, F., Simard, F., Rosado, M., Solano, P., Dumonteil, E., Waleckx, E., 2020. Metabarcoding: A Powerful Yet Still Underestimated Approach for the Comprehensive Study of Vector-Borne Pathogen Transmission Cycles and their Dynamics. Vector-Borne Diseases - Recent Developments in Epidemiology and Control. IntechOpen.
Herrera, C., Bargues, M.D., Fajardo, A., Montilla, M., Triana, O., Vallejo, G.A., Guhl, F., 2007. Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. Infect. Genet. Evol. 7 (4), 535–539.
Herrera, C., Guhl, F., Falla, A., Fajardo, A., Montilla, M., Vallejo, G. Adolfo, Bargues, M. D., 2009. Genetic variability and phylogenetic relationships within *Trypanosoma cruzi* I isolated in colombia based on miniexon gene sequences. J. Parasitol. Res. 2009, 1–9.
Herrera, C.P., Barnabé, C., Brenière, S.F., 2013. Complex evolutionary pathways of the intergenic region of the mini-exon gene in *Trypanosoma cruzi* TcI: a possible ancient origin in the Gran Chaco and lack of strict genetic structuration. Infect. Genet. Evol. 16, 27–37.
Herrera, C., Truyens, C., Dumonteil, E., Alger, J., Sosa-Estani, S., Cafferata, M.L., Gibbons, L., Ciganda, A., Matute, M.L., Zuniga, C., Carlier, Y., Buekens, P., 2019a. Phylogenetic analysis of *Trypanosoma cruzi* from pregnant women and newborns from Argentina, Honduras, and Mexico suggests an association of parasite haplotypes with congenital transmission of the parasite. J. Mol. Diagn. 21 (6), 1095–1105.
Herrera, C., Majeau, A., Didier, P., Falkenstein, K.P., Dumonteil, E., 2019b. *Trypanosoma cruzi* diversity in naturally infected nonhuman primates in Louisiana assessed by deep sequencing of the mini-exon gene. Trans. R. Soc. Trop. Med. Hyg. 113 (5), 281–286.
Izeta-Alberdi, A., Ibarra-Cerdena, C.N., Moo-Llanes, D.A., Ramsey, J.M., 2016. Geographical, landscape and host associations *of Trypanosoma cruzi* DTUs and lineages. Parasit. Vectors 9 (1), 631.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7. 0 for bigger datasets. Mol. Biol. Evol. 33 (7), 1870–1874.
Lauthier, J.J., Tomasini, N., Barnabé, C., Rumi, M.M., D'Amato, A.M., Ragone, P.G., Yeo, M., Lewis, M.D., Llewellyn, M.S., Basombrio, M.A., Miles, M.A., Tibayrenc, M., Diosque, P., 2012. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. 12 (2), 350–358.
Lewis, M.D., Ma, J., Yeo, M., Carrasco, H.J., Llewellyn, M.S., Miles, M.A., 2009. Genotyping of *Trypanosoma cruzi*: systematic selection of assays allowing rapid and accurate discrimination of all known lineages. Am J. Trop. Med. Hyg. 81 (6), 1041–1049.
Lewis, M.D., Llewellyn, M.S., Yeo, M., Acosta, N., Gaunt, M.W., Miles, M.A., 2011. Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl. Trop. Dis. 5 (10), e1363.
Lima-Cordon, R.A., Cahan, S.H., McCann, C., Dorn, P.L., Justi, S.A., Rodas, A., Monroy, M.C., Stevens, L., 2021. Insights from a comprehensive study of *Trypanosoma cruzi*: a new mitochondrial clade restricted to north and Central America and genetic structure of TcI in the region. PLoS Negl. Trop. Dis. 15, e0010043.
Lisboa, C.V., Xavier, S.C., Herrera, H.M., Jansen, A.M., 2009. The ecology of the *Trypanosoma cruzi* transmission cycle: Dispersion of zymodeme 3 (Z3) in wild hosts from Brazilian biomes. Vet. Parasitol. 165 (1–2), 19–24.
Lopez-Dominguez, J., Lopez-Monteon, A., Ochoa-Martinez, P., Dumonteil, E., Barnabé, C., Waleckx, E., Hernandez-Giles, R.G., Ramos-Ligonio, A., 2022. Molecular characterization of four Mexican isolates of *Trypanosoma cruzi* and their profile susceptibility to Nifurtimox. Acta Parasitol. 67, 1584–1593.
Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. U. S. A. 98, 7396–7401.
Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95 (6), 3140–3145.
Maiguashca Sanchez, J., Sueto, S.O.B., Schwabl, P., Grijalva, M.J., Llewellyn, M.S., Costales, J.A., 2020. Remarkable genetic diversity of *Trypanosoma cruzi* and *Trypanosoma rangeli* in two localities of southern Ecuador identified via deep sequencing of mini-exon gene amplicons. Parasit. Vectors 13 (1), 252.
Majeau, A., Herrera, C., Dumonteil, E., 2019. An improved approach to *Trypanosoma cruzi* molecular genotyping by next-generation sequencing of the mini-exon gene. Methods Mol. Biol. 1955, 47–60.
Majeau, A., Murphy, L., Herrera, C., Dumonteil, E., 2021. Assessing *Trypanosoma cruzi* parasite diversity through comparative genomics: implications for disease epidemiology and diagnostics. Pathogens 10.
Matos, G.M., Lewis, M.D., Talavera-Lopez, C., Yeo, M.T., Grisard, E.C., Messenger, L.A., Miles, M., Andersson, B., 2022. Microevolution of *Trypanosoma cruzi* reveals hybridization and clonal mechanisms driving rapid genome diversification. Elife 11.
Messenger, L.A., Llewellyn, M.S., Bhattacharyya, T., Franzen, O., Lewis, M.D., Ramirez, J.D., Carrasco, H.J., Andersson, B., Miles, M.A., 2012. Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and next generation sequencing. PLoS Negl. Trop. Dis. 6 (4) e1584.
Messenger, L.A., Miles, M.A., Bern, C., 2015. Between a bug and a hard place: *Trypanosoma cruzi* genetic diversity and the clinical outcomes of Chagas disease. Expert Rev. Anti-Infect. Ther. 13 (8), 995–1029.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37 (5), 1530–1534.
Monteiro, W.M., Margioto Teston, A.P., Gruendling, A.P., dos Reis, D., Gomes, M.L., de Araujo, S.M., Bahia, M.T., Magalhaes, L.K., de Oliveira Guerra, J.A., Silveira, H., Toledo, M.J., Barbosa, M. Vale, 2013. *Trypanosoma cruzi* I and IV stocks from Brazilian Amazon are divergent in terms of biological and medical properties in mice. PLoS Negl. Trop. Dis. 7 (2) e2069.
Murillo-Solano, C., Ramos-Ligonio, A., Lopez-Monteon, A., Guzman-Gomez, D., Torres-Montero, J., Herrera, C., Dumonteil, E., 2021. Diversity of *Trypanosoma cruzi* parasites infecting *Triatoma dimidiata* in Central Veracruz, Mexico, and their One Health ecological interactions. Infect. Genet. Evol. 95, 105050.
O'Connor, O., Bosseno, M.F., Barnabé, C., Douzery, E.J., Brenière, S.F., 2007. Genetic clustering of *Trypanosoma cruzi* I lineage evidenced by intergenic miniexon gene sequencing. Infect. Genet. Evol. 7 (5), 587–593.
Orantes, L.C., Monroy, C., Dorn, P.L., Stevens, L., Rizzo, D.M., Morrissey, L., Hanley, J.P., Rodas, A.G., Richards, B., Wallin, K.F., Helms Cahan, S., 2018. Uncovering vector, parasite, blood meal and microbiome patterns from mixed-DNA specimens of the Chagas disease vector *Triatoma dimidiata*. PLoS Negl. Trop. Dis. 12, e0006730.
Perez-Losada, M., Cabezas, P., Castro-Nallar, E., Crandall, K.A., 2013. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. Infect. Genet. Evol. 16, 38–53.
Pinto, C.M., Kalko, E.K., Cottontail, I., Wellinghausen, N., Cottontail, V.M., 2012. TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. Infect. Genet. Evol. 12 (6), 1328–1332.
Polonio, R., Lopez-Dominguez, J., Herrera, C., Dumonteil, E., 2021. Molecular ecology of *Triatoma dimidiata* in southern Belize reveals risk for human infection and the local differentiation of *Trypanosoma cruzi* parasites. Int. J. Infect. Dis. 108, 320–329.
Probst, C.M., Melo, M., Pavoni, D.P., Toledo, M.J.O., Galdino, T.S., Brandao, A.A., Britto, C., Krieger, M.A., 2021. A new Trypanosoma cruzi genotyping method enables high resolution evolutionary analyses. Mem. Inst. Oswaldo Cruz 116 e200538.
Pronovost, H., Peterson, A.C., Chavez, B.G., Blum, M.J., Dumonteil, E., Herrera, C.P., 2020. Deep sequencing reveals multiclonality and new discrete typing units of *Trypanosoma cruzi* in rodents from the southern United States. J. Microbiol. Immunol. Infect. 53 (4), 622–633.
Ragone, P.G., Perez Brandan, C., Padilla, A.M., Monje Rumi, M., Lauthier, J.J., Alberti D'Amato, A.M., Tomasini, N., Cimino, R.O., Romero, N.M., Portelli, M., Nasser, J.R., Basombrio, M.A., Diosque, P., 2012. Biological behavior of different *Trypanosoma cruzi* isolates circulating in an endemic area for Chagas disease in the Gran Chaco region of Argentina. Acta Trop. 123 (3), 196–201.
Ramirez, J.D., Tapia-Calle, G., Guhl, F., 2013. Genetic structure of *Trypanosoma cruzi* in Colombia revealed by a High-throughput Nuclear Multilocus Sequence Typing (nMLST) approach. BMC Genet. 14, 96.
Reis-Cunha, J.L., Rodrigues-Luiz, G.F., Valdivia, H.O., Baptista, R.P., Mendes, T.A., de Morais, G.L., Guedes, R., Macedo, A.M., Bern, C., Gilman, R.H., Lopez, C.T., Andersson, B., Vasconcelos, A.T., Bartholomeu, D.C., 2015. Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains. BMC Genomics 16, 499.

Rodrigues, M.S., Morelli, K.A., Jansen, A.M., 2017. Cytochrome c oxidase subunit 1 gene as a DNA barcode for discriminating *Trypanosoma cruzi* DTUs and closely related species. Parasit. Vectors 10 (1), 488.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahe, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4 e2584.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3. 2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542.

Stevens, J.R., Teixeira, M.M., Bingle, L.E., Gibson, W.C., 1999. The taxonomic position and evolutionary relationships of *Trypanosoma rangeli*. Int. J. Parasitol. 29 (5), 749–757.

Tibayrenc, M., 1998. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. Int. J. Parasitol. 28 (1), 85–104.

Tibayrenc, M., 2003. Genetic subdivisions within *Trypanosoma cruzi* (Discrete Typing Units) and their relevance for molecular epidemiology and experimental evolution. Kinetoplastid Biol. Dis. 2 (1), 12.

Tibayrenc, M., Ayala, F.J., 2015. The population genetics of *Trypanosoma cruzi* revisited in the light of the predominant clonal evolution model. Acta Trop. 151, 156–165.

Tomasini, N., Lauthier, J.J., Llewellyn, M.S., Diosque, P., 2013. MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. Infect. Genet. Evol. 20, 188–196.

Tomasini, N., Lauthier, J.J., Monje Rumi, M.M., Ragone, P.G., Alberti D'Amato, A.M., Brandan, C.P., Basombrio, M.A., Diosque, P., 2014. Preponderant clonal evolution of *Trypanosoma cruzi* I from Argentinean Chaco revealed by Multilocus Sequence Typing (MLST). Infect. Genet. Evol. 27, 348–354.

Treangen, T.J., Salzberg, S.L., 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. 13 (1), 36–46.

Villanueva-Lizama, L., Teh-Poot, C., Majeau, A., Herrera, C., Dumonteil, E., 2019. 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. 219 (12), 1980–1988.

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., 2011. Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in *Trypanosoma cruzi*. PLoS Negl. Trop. Dis. 5 (6) e1049.

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., Andrade, S.G., Sturm, N.R., 2012. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect. Genet. Evol. 12 (2), 240–253.

Zuriaga, M.A., Blandon-Naranjo, M., Valerio-Campos, I., Salas, R., Zeledon, R., Bargues, M.D., 2012. Molecular characterization of *Trypanosoma cruzi* and infection rate of the vector *Triatoma dimidiata* in Costa Rica. Inf. Res. 111 (4), 1615–1620.