INSTITUTO NACIONAL DE SALUD PÚBLICA

Centro Regional de Investigación en Salud Pública

IDENTIFICACIÓN DE HAPLOTIPOS DE *Trypanosoma cruzi* LINAJE I EN AISLADOS DE MÉXICO Y GUATEMALA.

T E S I S

QUE PARA OBTENER EL **GRADO** DE:

MAESTRO EN CIENCIAS DE LA SALUD

P R E S E N T A:

FEDERICO ALONSO ZUMAYA ESTRADA

Director de Tesis:

DRA. TERESA LÓPEZ ORDÓÑEZ

CENTRO REGIONAL DE INVESTIGACIÓN EN SALUD PÚBLICA
INSTITUTO NACIONAL DE SALUD PÚBLICA

Asesores:

DR. CARLOS IBARRA CERDEÑA
DR. CARLOS ALBERTO FLORES LÓPEZ

TAPACHULA, CHIAPAS; MÉXICO. AGOSTO 2012

DEDICATORIA

Al ejemplo de mis padres Olivia Estrada Manzano y Héctor U. Zumaya Ramírez así como al amor de mis hermanos Zyanya M. Zumaya Estrada y Carlos Raúl Zumaya Estrada.

AGRADECIMIENTO

Muy en especial a la Dra. Teresa López Ordóñez por su incondicional apoyo y enseñanza en mi desarrollo académico y personal.

Al Dr. Martin Llewellyn y la Dra. Louisa Messenger por compartir su amistad y conocimiento para la realización de mi Tesis.

Al siempre buen consejo del Dr. Jose Alejandro Martínez Ibarra.

Al Consejo Nacional de Ciencia y Tecnología por las becas otorgadas para la realización de mi postgrado y estancia en el extranjero.

Index

Abstract

Background

Trypanosoma cruzi, the agent of Chagas disease, is currently recognized as a complex of six lineages or Discrete Typing Units (DTU): TcI-TcVI. Recent findings have identified a divergent group within TcI - $TcI_{DOM.}$ TcI_{DOM} is strongly related to sylvatic strains in North/Central America, largely absent from wild mammals and vectors, and associated with a significant proportion of human TcI infections in South America. Our aim was to develop hypotheses around the origin of the TcI_{DOM} clade. Several are relevant; including an emergence in northern South America as a sister group of North American strains and dispersal among domestic transmission cycles, or an emergence in North America prior to dispersal back into South America into domestic cycles, possibly anthropically. To provide further insight we undertook high resolution nuclear and mitochondrial genotyping of multiple Central American strains (from areas of Mexico and Guatemala) and included them in a meta-analysis with other published data.

Results

Mitochondrial sequence and nuclear microsatellite data revealed a cline in genetic diversity across isolates grouped into three populations: South America, North/Central America and TcI_{DOM}. As such, greatest diversity was observed in South America ($\pi = 0.00712$, $A_r = 4.851$) and lowest in TcI_{DOM} ($\pi = 0.00071$, $A_r = 1.813$). Nuclear genetic clustering (genetic distance based) analyses suggest that TcI_{DOM} clusters as a monophyletic group within the North/ Central American clade.

Conclusions

Declining genetic diversity across the populations, and corresponding hierarchical clustering suggest that this important human genotype probably emerged in North / Central America before moving southwards. The implication of this finding in the context of early human dispersal is discussed.

Key words

Trypanosoma cruzi, maxicircle, microsatellite, Chagas disease, phylogeography, population genetics.

Introduction

Trypanosoma cruzi is the causative agent of Chagas disease. It infects approximately 10 million people in Latin America, while some 25 million more are at risk of acquiring the disease [WHO, 2010]. Parasite transmission to mammal hosts, including humans, is accomplished mainly via infectious contact with the feces of blood-sucking triatomine bugs. However, several non-vectorial routes are also recognized including blood transfusion, organ transplantation, orally and congenital transmission, via ingestion of meals contaminated with infected triatomine feces [Lewis et al, 2011; Alarcón *et al.*, 2010]. *T. cruzi* (family Trypanosomatidae; Euglenozoa: Kinetoplastida) is phylogenetically nested within several widely dispersed species of bat trypanosomes [Lima *et al.*, 2012]. A more divergent species of this genus is *Trypanosoma brucei*, the causative agent of sleeping sickness [Hamilton *et al.*, 2004]. The time of divergence between these two species is the subject of active debate. Nevertheless the more classical hypotheses support a dominant role for biogeographic processes and a split concurrent when Africa separated from the South American, Antarctica and Australasia continents, approximately 100 million years ago (MYA) [Stevens *et al.*, 1999a; Stevens *et al.*, 1999b]. However, the likely origin of *T. cruzi* from an ancestral bat trypanosome capable of long range dispersal, and reports of *T. cruzi-like* trypanosomes among African arboreal mammalian fauna [Hamilton *et al.*, 2009], suggest that *T. cruzi* (or its progenitor) may have arrived in South America later than previously thought [Hamilton *et al.*, 2012]. Whilst the evolutionary history of *T. cruzi* and South America's paleogeology may not be as intertwined as we previously thought, the parasite is undeniably ancient and indigenous [Flores-Lopez & Machado, 2011]. *T. cruzi* transmission is maintained via hundreds of mammal and triatomine species in different biomes throughout South America, Central America, Mexico and some southern states of the USA [Yeo *et al.*, 2005]. This wide distribution as well as host and vector diversity, can explain the well-established genetic diversity observed in *T. cruzi* [Miles *et al.*, 2009]. Biochemical and molecular markers support the existence of six lineages or Discrete Typing Units (DTU): TcI, TcII, TcIII, TcIV, TcV and TcVI, recently defined by international consensus [Zingales *et al.*, 2009; 2012]. Each DTU can be loosely associated with a particular ecological and/or geographical framework [Miles *et al.*, 2009]. Additionally a new *T. cruzi* phylogenetic clade has been described in Brazil and Panama, termed TcBat. This clade forms a monophyletic group, very related to TcI [Marcili *et al.*, 2009; Pinto *et al.*, 2012]. Among all the six described DTUs perhaps the most widely sampled, distributed and abundant DTU is TcI. Several studies have associated this DTU with a mostly sylvatic niche in South America [Araujo *et*

al., 1996; Zingales *et al.*, 2012]. However it has also been reported as the main DTU infecting humans north of the amazon basin [Gonzales *et al.*, 2010; Añez *et al.*, 2004]. Several molecular tools now identify substantial genetic diversity displayed within TcI [Llewellyn *et al.*, 2009; Cura *et al.*, 2010; Guhl & Ramirez, 2011]. Furthermore, recent studies support the presence of a divergent group within TcI associated with human infections from Venezuela to Northern Argentina and since it is essentially absent from wild mammals and vectors [Llewellyn *et al.*, 2009; Messenger *et al.*, 2012] henceforth termed in this study as TcI_{DOM}. The origin of this clade is not clear, although some data suggest this clade is closely related to the TcI clade circulating in North America [Llewellyn *et al.*, 2009; Messenger *et al.*, 2012]. In this manuscript, we set out to evaluate the genetic diversity of TcI in North and Central America, undertaking a comparison with the TcI diversity found in South America, with a particular emphasis in TcI_{DOM}. Our aim was to distinguish between two competing hypotheses: 1) TcI_{DOM} emerged in northern South America as a sister group of North American strains and dispersed among domestic transmission cycles 2) TcI_{DOM} emerged in North America and was subsequently dispersed back into South America into domestic cycles. To achieve this we undertook high resolution nuclear and mitochondrial genotyping of multiple Central and northern South America strains (Mexico, Guatemala and Venezuela) and included them in a meta-analysis with other published data [Messenger *et al.*, 2012; Ramirez *et al.*, 2012].

Materials and Methods

T. cruzi isolates

A panel of 25 parasite DNA isolates from several vectors and hosts included human were derived from epimastigote culture. Of these, nine were from Mexico [Bucio *et al.*, 1999], twelve from Guatemala [Ramirez *et al.*, 2012] and four from Venezuela [Pennington *et al.*, 2009] as indicated in Table 1. The DTU characterization level was made via the amplification and sequencing of glucose-6-phosphate isomerase gene (GPI) as previously described by Lauthier *et al.*, (2012). To high resolution genotyping, nine maxicircle gene fragments (Mitochondrial multilocus sequence typing) and nineteen nuclear microsatellite loci (Multilocus microsatellite typing) were analyzed.

DTU level characterization (GPI)

Amplifications of the glucose-6-phosphate isomerase gene (GPI) were achieved for Polymerase Chain Reaction (PCR) in a final volume of 20 µl containing: 1 X reaction buffer (Bioline, UK), 50 mM $MgCl_2$ (Bioline, UK), 0.2 mM dNTPs (New England Biolabs, UK), 20 pmol of each primer, 0.2 U Taq DNA polymerase (Bioline, UK) and 10–100 ng of genomic DNA. PCR reactions were performed with an initial denaturation step of 5 minutes at 94°C, followed by 35 amplification cycles (94°C for 1 minute, 55 °C for 1 minute, 72°C for 1 minute) and a final elongation step at 72°C for five minutes. Primers are described in Table 2. Nucleotide sequence data was assembled manually in BioEdit v7.0.9.0 (Ibis Biosciences, USA) [Hall, 1999]. An unambiguous consensus sequence was produced for each isolate. Sequences were aligned using Clustal X for a comparative homology analysis [Thompson *et al.*, 1997]. All sequences were identical to the TcI references [Lauthier et al., 2012].

Table 1. Panel of *T. cruzi* DNA isolates.

Isolate code	Host/vector	Country	State	Latitude	Longitude	Date	Reference	Culture
38	*Triatoma dimidiata*	Guatemala	Jutiapa	14.29	-89.84	2000	This study	Pennington *et al.*, 2009
46	*Triatoma dimidiata*	Guatemala	Santa Rosa	14.18	-90.30	2001	This study	Pennington *et al.*, 2009
66	*Triatoma dimidiata*	Guatemala	Jalapa	14.63	-89.99	2001	This study	Pennington *et al.*, 2009
67	*Triatoma dimidiata*	Guatemala	Jutiapa	14.29	-89.84	2001	This study	Pennington *et al.*, 2009
70	*Triatoma dimidiata*	Guatemala	Jutiapa	14.29	-89.84	2001	This study	Pennington *et al.*, 2009
71	*Triatoma dimidiata*	Guatemala	Jalapa	14.63	-89.99	2001	This study	Pennington *et al.*, 2009
83	*Triatoma dimidiata*	Guatemala	Chiquimula	14.77	-89.46	2002	This study	Pennington *et al.*, 2009
95	*Triatoma dimidiata*	Guatemala	Chiquimula	14.77	-89.46	2002	This study	Pennington *et al.*, 2009
100	*Triatoma dimidiata*	Guatemala	Santa Rosa	14.18	-90.30	2002	This study	Pennington *et al.*, 2009
113	*Triatoma dimidiata*	Guatemala	Chiquimula	14.77	-89.46	2002	This study	Pennington *et al.*, 2009
116	*Triatoma dimidiata*	Guatemala	Baja Verapaz	15.08	-90.41	2002	This study	Pennington *et al.*, 2009
154	*Triatoma dimidiata*	Guatemala	Alta Verapaz	15.59	-90.15	2002	This study	Pennington *et al.*, 2009
ANITA II	*Triatoma dimidiata*	Mexico	Campeche	19.19	-90.30	2011	This study	Bucio *et al.*, 1999
CAM6	*Triatoma dimidiata*	Mexico	Campeche	19.19	-90.30	2011	This study	Bucio *et al.*, 1999
CRISTY	*Homo sapiens*	Mexico	San Luis Potosí	22.16	-100.99	2007	This study	Bucio *et al.*, 1999
MICH1	*Triatoma dimidiata*	Mexico	Michoacan	19.57	-101.71	2011	This study	Bucio *et al.*, 1999
NINOA	*Homo sapiens*	Mexico	Oaxaca	17.05	-96.71	1994	This study	Bucio *et al.*, 1999
PLI	*Dipetalogaster maxima*	Mexico	Baja California Sur	26.04	-111.67	2001	This study	Bucio *et al.*, 1999
QROI	*Triatoma barberi*	Mexico	Queretaro	20.59	-100.39	1986	This study	Bucio *et al.*, 1999
TQI	*Triatoma pallidipennis*	Mexico	Morelos	18.95	-99.22	1991	This study	Bucio *et al.*, 1999
XAL1	*Triatoma dimidiata*	Mexico	Veracruz	19.17	-96.13	2003	This study	Bucio *et al.*, 1999
10462P2C3	*Homo sapiens*	Venezuela	Miranda	10.27	-66.48	Unknown	This study	Ramirez *et al.*, 2012
10462P2C7	*Homo sapiens*	Venezuela	Miranda	10.08	-66.45	Unknown	This study	Ramirez *et al.*, 2012
10968P1C1	*Homo sapiens*	Venezuela	Sucre	10.41	-63.30	Unknown	This study	Ramirez *et al.*, 2012
ANT3P1C6	*Homo sapiens (oral)*	Venezuela	DC	10.50	-66.95	Unknown	This study	Ramirez *et al.*, 2012

Mitochondrial multilocus sequence typing (mtMLST)

Nine maxicircle gene fragments were amplified for this study: *ND1* (NADH dehydrogenase subunit 1), *COII* (cytochrome c oxidase subunit II), *MURF1* (Maxicircle Unidentified Reading Frame 1, two fragments), *CYT b* (cytochrome b), *12S rRNA*, *9S rRNA*, and *ND5* (NADH dehydrogenase subunit 5, two fragments) as previously described by Messenger *el al.,* (2012). Primers and annealing temperatures are in Table 2. Amplifications for all targets were achieved in a final volume of 20 µl containing: 1 X NH_4 reaction buffer (Bioline, UK), 1.5 mM $MgCl_2$ (Bioline, UK), 0.2 mM dNTPs (New England Biolabs, UK), 10 pmol of each primer, 1 U Taq DNA polymerase (Bioline, UK) and 10–100 ng of genomic DNA. PCR reactions were performed with an initial denaturation step of 3 minutes at 94°C, followed by 30 amplification cycles (94°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds) and a final elongation step at 72°C for ten minutes. Maxicircle and nuclear PCR products were purified using QIAquick PCR extraction kits (Qiagen, UK) according to the manufacturer's protocol. PCR products were visualized in 1.5% agarose gels and were purified by isopropanol precipitation. Bi-directional sequencing was performed for both nuclear and maxicircle targets using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, UK) according to the manufacturer's protocol. PCR products were sequenced using their respective PCR primers described in Table 2. Nucleotide sequences for all nine gene fragments are available from GenBank under the accession numbers listed in Table 2.

Table 2. GPI and Maxicircle primers.

Gene fragment	Primer name	Primer sequence	Annealing Temp.(°C)	Amplicon size (bp)*	Primer reference	Accesion number
12S rRNA	12S-Fwd	GTTTATTAAATGCGTTTGTCTAAGAA	50	299	Messenger *et al.,* 2012	JX431210 - JX431234
	12S-Rvs	GCCCCAATCAAACATACAA				
9S rRNA	9S-Fwd	TGCAATTCGTTAGTTGGGTTA	50	302	Messenger *et al.,* 2012	JX431185 - JX431209
	9S-Rvs	TCCACACCCATTAAATAGCACT				
Cytochrome b *(CYT b)*	Sp18-Fwr	GACAGGATTGAGAAGCGAGAGAG	50	717	Messenger *et al.,* 2012	JX431260 - JX431284
	Sp18-Rvs	CAAACCTATCACAAAAAGCATCTG				
Maxicircle unidentified reading frame 1 *(MURF1* Fragment A)	MurfA-Fwr	AAGGCRATGGGRATAGWRCCTATAC	50	482	Messenger *et al.,* 2012	JX431060 - JX431084
	MurfA-Rvs	TGGAACAATTRTATATCAGATTRGGA				
Maxicircle unidentified reading frame 1 *(MURF1* Fragment B)	MurfB-Fwr	ACMCCCATCCATTCTTCR	50	423	Messenger *et al.,* 2012	JX431085 - JX431109
	MurfB-Rvs	CCTTTGATYTATTGTGATTAACRKT				
NADH dehydrogenase subunit 1 *(ND1)*	ND1-Fwr	GCACTTTCTGAAATAATCGAAAA	50	400	Messenger *et al.,* 2012	JX431110 - JX431134
	ND1-Rvs	TTAATCTTATCAGGATTTGTTAGCC				
Cytochrome c oxidase subunit II *(COII)*	COII-Fwr	GTTATTATCTTTTGTTTGTTTTGTGTG	50	560	Messenger *et al.,* 2012	JX431235 - JX431259
	COII-Rvs	AACAATTGGCATAAATCCATGT				
NADH dehydrogenase subunit 5 *(ND5* Fragment A)	ND5a-Fwr	TATGRYTAACYTTTTCATGYTCRG	50	503	Messenger *et al.,* 2012	JX431135 - JX431159
	ND5a-Rvs	GTCCTTCCATYGCATCYGG				
NADH dehydrogenase subunit 5 (*ND5* Fragment B)	ND5b-Fwr	ARAGTACACAGTTTGGRYTRCAYA	50	444	Messenger *et al.,* 2012	JX431160 - JX431184
	ND5b-Rvs	CTTGCYAARATACAACCACAA				
Glucose-6-phosphate isomerase *(GPI)*	Gpi-L	CGCCATGTTGTGAATATTGG	55	405	Lauthier *et al.,* 2012	
	Gpi-R	GGCGGACCACAATGAGTATC				

Multilocus microsatellite typing (MLMT)

Nineteen microsatellite loci previously described by Llewellyn *et al.*, (2009), selected based on their level of TcI intra-lineage resolution, were analyzed. Primers and binding sites are listed in Table 3. The following reaction cycle was implemented across all loci: a denaturation step of 4 minutes at 95°C, then 30 amplification cycles (95°C for 20 seconds, 57°C for 20 seconds, 72°C for 20 seconds) and a final elongation step of 72°C for 20 minute. With a final volume of 10 µl, containing 1 X ThermoPol Reaction Buffer (New England Biolabs (NEB), UK), 4 mM $MgCl_2$, 34 µM dNTPs; 0.75 pmols of each primer, 1 U of Taq DNA polymerase (NEB, UK) and 1 ng of genomic DNA were added. Five fluorescent dyes were used to label forward primers: 6-FAM & TET (Proligo, Germany), NED, PET & VIC (Applied Biosystems, UK). Allele sizes were determined using an automated capillary sequencer (AB3730, Applied Biosystems, UK), and carried out automatically in GeneMapper software, was also checked manually for errors and inconsistencies associated with abnormal stutter patterns and other artifacts, and manually checked for errors and typed ‘‘blind’’ to control for user bias.

Table 3. Microsatellite loci and primers.

Primer code	Position	Repeat type	Foward/Reverse primers (5´-3´)
10101 (CA)a	88534-88549	(CA)n	GTCGCCATCATGTACAAACG
			CTGTTGGCGAATGGTCATAA
10101 (TA)	46598-46638	(TA)n	AACCCGCGCAGATACATTAG
			TTCATTTGCAGCAACACACA
10101 (TC)	49478-49495	(TC)n	CGTACGACGTGGACACAAAC
			ACAAGTGGGTGAGCCAAAAG
10187 (CA) (TA)	39618-39874	(CA)n(TA)n	CATGTCATTAAGTGGCCACG
			GCACATGTTGGTTGTTGGAA
10187 (GA)	71097-71226	(GA)n	GTCACACCACTAGCGATGACA
			ACTGCACAATACCCCCTTTG
10187 (TA)	44002-44057	(TA)n	AGAAAAAGGTTTACAACGAGCG
			CGATGGAGAACGTGAAACAA
10187 (TTA)	32430-32629	(TTA)n	GAGAGAGATTCGGAAACTAATAGC
			CATGTCCCTTCCTCCGTAAA
11283 (TA)b	12639-12819	(TA)n	AACATCCTCCACCTCACAGG
			TTTGAATGCGAGGTGGTACA
11283 TCG	88680-88708	(TCG)n	ACCACCAGGAGGACATGAAG
			TGTACACGGAACAGCGAAG
6529 (CA)a	92287-92316	(CA)n	TGTGAAATGATTTGACCCGA
			AGAGTCACGCCGCAAAGTAT
6529 (TA)b	75669-75701	(TA)n	TGAAGGAGATTCTCTGCGGT
			CTCTCATCTTTTGTTGTGTCCG
6855 (TA) (GA)	5369-5576	(TA)n(GA)n	TGTGATCAACGCGCATAAAT
			TTCCATTGCCTCGTTTTAGA
6925 (CT)	88658-88832	(CT)n	CATCAAGGAAAAACGGAGGA
			CGGTACCACCTCAAGGAAAG
7093 (TA)c	16233-16257	(TA)n	CGTGTGCACAGGAGAGAAAA
			CGTTTGGAGGAGGATTGAGA
7093 (TC)	69979-70016	(TC)n	CCAACATTCAACAAGGGAAA
			GCATGAATATTGCCGGATCT
7093 (TCC)	25751-25779	(TCC)n	AGACGTTCATATTCGCAGCC
			AGCCACATCCACATTTCCTC
mclf10 *	Unknown	(CA)nA(CA)n	GCGTAGCGATTCATTTCC
			ATCCGCTACCACTATCCAC
TcUn2	Unknown	Unknown	AACAAAATCTAGCGTCTACCATCC
			GGTGTTGGCGTGTATGATTG
TcUn3	Unknown	Unknown	CTTAAAGAGATACAAGAGGGAAGG
			CTGTTATTTCAATAACACGGGG

Published in Oliveira *et al.*, (1998)

Data analysis (Meta-analysis)

To evaluate the TcI genetic diversity of the Mexican, Guatemalan and Venezuela isolates with a particular emphasis in TcI_{DOM}, we included them in a meta-analysis with other published data of 47 isolates from South America countries as indicated in Table 4 [Messenger *et al.*, 2012; Ramirez *et al.*, 2012].

Phylogenetic analysis of maxicircle genes

Nucleotide sequence data were assembled manually in BioEdit v7.0.9.0 (Ibis Biosciences, USA) [Hall, 1999]. Sequences were aligned using Clustal X and manually checked for any obvious misalignments [Thompson *et al.*, 1997]. All maxicircle sequences were concatenated according to their structural arrangement (12S rRNA, 9S rRNA, Cyt b, MURF1, ND1, COII and ND5) in the correct coding direction. For two genes (MURF1 and ND5) two fragments were selected from each coding region to examine intra-gene variation. Seven sequences from additional DTUs (TcIII-TcIV) previously characterized by Messenger *et al.*, (2012) were included as out-groups (Table S1). Phylogenies were inferred using Maximum-Likelihood implemented in PhyML (4 substitution rate categories) [Guindon *et al.*, 2010]. The best-fit model of nucleotide substitution was selected from 88 models and its significance evaluated according to the Akaike Information Criterion (AIC) in jMODELTEST 1.0 [Posada, 2008]. The model selected to the data file was the General Time-Reversible Model (GTR), with gamma distributed rate variation among sites (G), and proportion of invariable sites (I). Maximum-Likelihood bootstrap support (BS) for clade topologies was estimated following the generation of 1000 pseudo-replicate datasets. Bayesian posterior probabilities analysis (B.P.P) was performed using MrBAYES v3.1 [Ronquist & Huelsenbeck, 2003]. Five independent analyses were run using a random starting tree with three heated chains and one cold chain over 10 million generations with sampling every 10 simulations (25% burn-in).

Phylogenetic analysis of microsatellite loci

Pair-wise distances (D_{AS}) between microsatellite genotypes for individual samples were calculated in MICROSAT v1.5d [Minch *et al.*, 1999] under the infinite-alleles model (IAM). To accommodate multiallelic genotypes, a script was written in Microsoft Visual Basic to generate random multiple diploid re-samplings of each multilocus profile. A final pair-wise distance matrix was derived from the mean of each re-sampled dataset and used to construct a phylogenetic tree in PHYLIP v3.67 [Felsenstein, 1989] using a Neighbour-Joining (NJ) clustering algorithm. A further test of the robustness

of tree topology was carried out by majority-rule consensus analysis in PHYLIP v3.67 of 10,000 bootstrap trees collated by combining 100 bootstraps, made in MICROSAT v1.2, each drawn from 100 respective randomly re-sampled datasets.

Table 4. Panel of TcI isolates data previously published.

Isolate code	Host/vector	Country	State	Latitude	Longitude	Date	Reference
AAA1cl5	*Rhodnius prolixus*	Colombia	Casanare	4.15	-71.20	2010	Ramirez *et al.*, 2012
AAA7cl2	*Rhodnius prolixus*	Colombia	Casanare	5.10	-71.60	2010	Ramirez *et al.*, 2012
AAB3cl3	*Rhodnius prolixus*	Colombia	Casanare	4.15	-71.20	2010	Ramirez *et al.*, 2012
AAC1cl3	*Rhodnius prolixus*	Colombia	Casanare	5.10	-71.60	2010	Ramirez *et al.*, 2012
AACf1cl4	*Canis familiaris*	Colombia	Casanare	5.10	-71.60	2010	Ramirez *et al.*, 2012
AAD6cl6	*Rhodnius prolixus*	Colombia	Casanare	5.10	-71.60	2010	Ramirez *et al.*, 2012
CACQcl7	*Homo sapiens*	Colombia	Santander	6.96	-73.42	2009	Ramirez *et al.*, 2012
CACQcl8	*Homo sapiens*	Colombia	Santander	6.64	-73.65	2009	Ramirez *et al.*, 2012
DYRcl16	*Homo sapiens*	Colombia	Boyacá	5.64	-72.90	2007	Ramirez *et al.*, 2012
EBcl11	*Homo sapiens*	Colombia	Boyacá	5.13	-73.12	2007	Ramirez *et al.*, 2012
FECcl10	*Homo sapiens*	Colombia	Boyacá	5.92	-73.50	2001	Ramirez *et al.*, 2012
Td3cl11	*Triatoma dimidiata*	Colombia	Boyacá	6.27	-71.20	2000	Ramirez *et al.*, 2012
X-1084cl10	*Rhodnius prolixus*	Colombia	Boyacá	4.96	-73.63	2010	Ramirez *et al.*, 2012
X-236cl9	*Rhodnius prolixus*	Colombia	Boyacá	4.96	-73.63	2010	Ramirez *et al.*, 2012
YAS1cl3	*Alouatta spp*	Colombia	Casanare	5.30	-72.40	2010	Ramirez *et al.*, 2012
PALDA4	*Didelphis albiventris*	Argentina	Chaco	-27.13	-61.46	2001	Messenger *et al.*, 2012
PALDA21	*Didelphis albiventris*	Argentina	Chaco	-27.13	-61.46	2001	Messenger *et al.*, 2012
PALDA5	*Didelphis albiventris*	Argentina	Chaco	-27.13	-61.46	2001	Messenger *et al.*, 2012
PALDAV2	*Triatoma infestans*	Argentina	Chaco	-27.13	-61.46	2001	Messenger *et al.*, 2012
PALDA20	*Didelphis albiventris*	Argentina	Chaco	-27.13	-61.46	2001	Messenger *et al.*, 2012
COTMA38	*Akodon boliviensis*	Bolivia	Cotopachi	-17.43	-66.27	2004	Messenger *et al.*, 2012
P234	*Homo sapiens*	Bolivia	Cochabamba	-17.38	-66.16	1985	Messenger *et al.*, 2012
P238	*Homo sapiens*	Bolivia	Cochabamba	-17.38	-66.16	1985	Messenger *et al.*, 2012
P268	*Homo sapiens*	Bolivia	Cochabamba	-17.38	-66.16	1987	Messenger *et al.*, 2012
SJM22	*Didelphis marsupialis*	Bolivia	Beni	-14.81	-64.60	2004	Messenger *et al.*, 2012
SJM34	*Didelphis marsupialis*	Bolivia	Beni	-14.81	-64.60	2004	Messenger *et al.*, 2012
SJM37	*Didelphis marsupialis*	Bolivia	Beni	-14.81	-64.60	2004	Messenger *et al.*, 2012
SJM39	*Didelphis marsupialis*	Bolivia	Beni	-14.81	-64.60	2004	Messenger *et al.*, 2012
SJM41	*Philander opossum*	Bolivia	Beni	-14.81	-64.60	2004	Messenger *et al.*, 2012
SJMC12	*Philander opossum*	Bolivia	Beni	-14.81	-64.60	2004	Messenger *et al.*, 2012
XE5167	*Didelphis marsupialis*	Brasil	Para	-1.71	-48.88	1999	Messenger *et al.*, 2012
IM4810	*Didelphis marsupialis*	Brasil	Manaus	-3.07	-60.16	2002	Messenger *et al.*, 2012
B2085	*Didelphis marsupialis*	Brasil	Belem	-1.36	-48.36	1991	Messenger *et al.*, 2012
XE2929	*Didelphis marsupialis*	Brasil	Pará	-5.83	-48.03	1988	Messenger *et al.*, 2012
DAVIScl1	*Triatoma dimidiata*	Honduras	Tegucigalpa	14.08	-87.20	1983	Messenger *et al.*, 2012
9209802P	*Didelphis marsupialis*	USA	Georgia	32.43	-83.31	1992	Messenger *et al.*, 2012
9307	*Didelphis marsupialis*	USA	Georgia	32.43	-83.31	1993	Messenger *et al.*, 2012
ARMA	*Dasypus novemcinctus*	USA	Lousiana	30.50	-91.00	Unknown	Messenger *et al.*, 2012
USA	*Didelphis marsupialis*	USA	Lousiana	30.50	-91.00	Unknown	Messenger *et al.*, 2012
9354	*Homo sapiens*	Venezuela	Sucre	10.46	-63.61	1999	Messenger *et al.*, 2012
11541	*Homo sapiens*	Venezuela	Merida	8.59	-71.23	2003	Messenger *et al.*, 2012
11713	*Homo sapiens*	Venezuela	Lara	10.23	-69.87	2003	Messenger *et al.*, 2012
11804	*Homo sapiens*	Venezuela	Portuguesa	9.08	-69.10	2003	Messenger *et al.*, 2012
M13	*Didelphis marsupialis*	Venezuela	Barinas	7.50	-71.23	2004	Messenger *et al.*, 2012
M16	*Didelphis marsupialis*	Venezuela	Barinas	7.50	-71.23	2004	Messenger *et al.*, 2012
M18	*Didelphis marsupialis*	Venezuela	Barinas	7.50	-71.23	2004	Messenger *et al.*, 2012
M7	*Didelphis marsupialis*	Venezuela	Barinas	7.50	-71.23	2004	Messenger *et al.*, 2012

Results

Maxicircle genes (mtMLST)

Across the 3,449 bp final concatenated alignment, a total of 374 variable sites were found. No evidence of heterozygosity ('split peaks') was observed, a pattern that is only observed in mixed infections. The mitochondrial phylogeny supports the presence of three main clades: NORTH-CENT, SOUTH and TcI_{DOM}. NORTH-CENT includes isolates from USA, México, Guatemala and Honduras; SOUTH corresponds to South America with isolates from Argentina, Bolivia, Colombia, Venezuela and Brazil and TcI_{DOM} with exclusively human isolates from Colombia and Venezuela, already know to belong to a restricted diversity genotype: TcIa, previously described by Herrera *et al.,* (2007) and VENDom by Llewellyn *et al.,* (2009). (Figure 1). The TcI_{DOM} isolates formed a monophyletic clade [60% M.L BS/0.98 B.P.P] nested within the NORTH-CENT. Whilst the NORTH-CENT clade is nested within the paraphyletic clade: SOUTH [78% M.L BS/1.0 B.P.P.] (Figure 1). Interestingly the most of Mexican and Guatemalan isolates were more basal than TcI_{DOM} clade and formed a paraphyletic clade that also appear to be more genetically diverse than the TcI_{DOM} isolates (Figure 1). All isolates belonging to other DTUs (TcIII and TcIV) cluster within a same clade irrespective of their geographic origin (Figure 1).

Nuclear microsatellite loci (MLMT)

A Neighbour-Joining tree was constructed based on Pair-wise distances (D_{AS}) between individual isolates (Figure 2). Two major clades were identified: NORTH-CENT [65% / 88.7] and SOUTH [61 / 98.9]. These clades correspond to the split between North/Central America and South America isolates. The TcI_{DOM} clade [44% / 100] remained monophyletic clustered within the NORTH-CENT although with poor BS support (Figure 1). In agreement with the mitochondrial data the Mexican and Guatemalan isolates also appear to be paraphyletic groups and more basal than the TcI_{DOM} isolates (Figure 2).

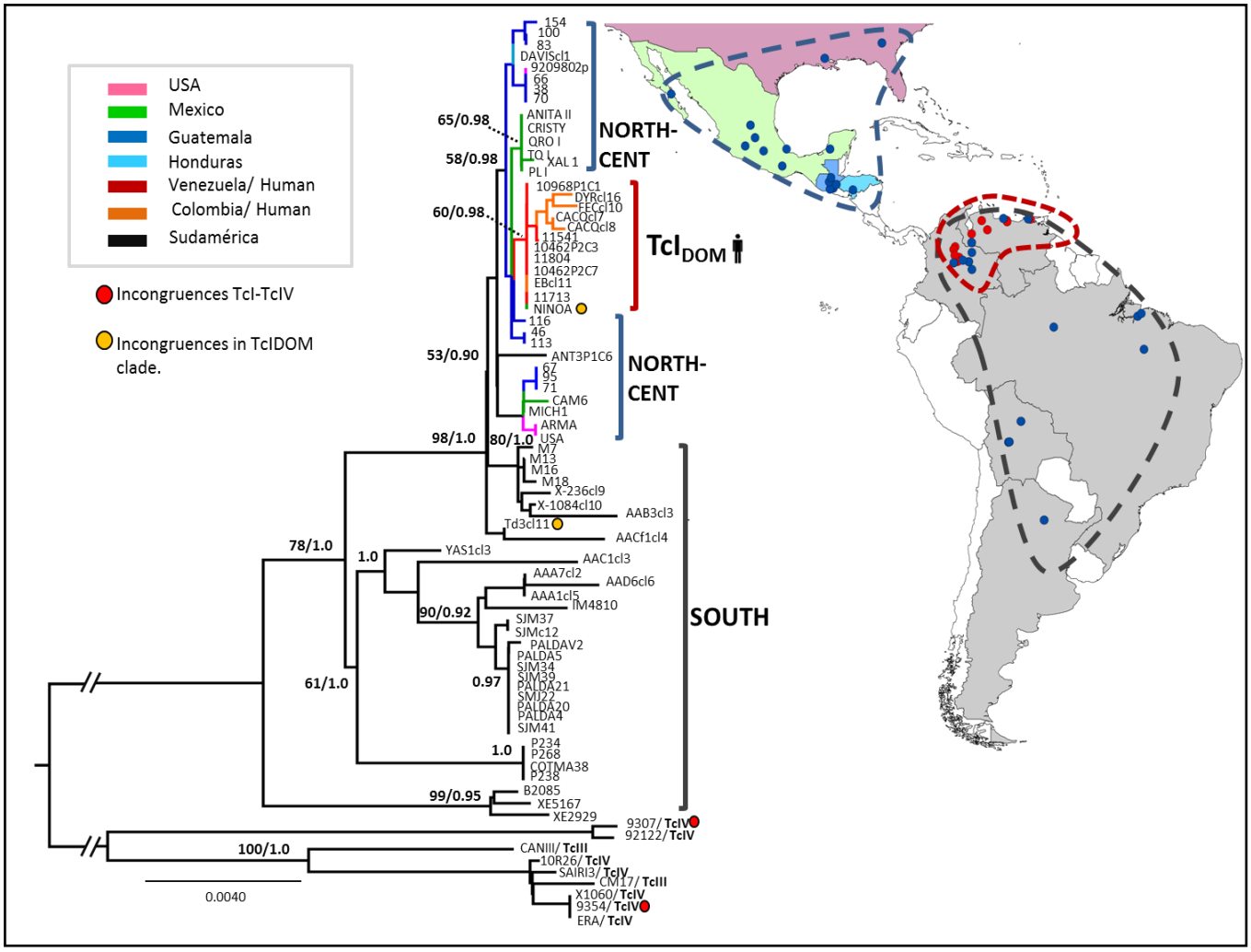

Figure 1. Maximum-Likelihood (M.L) tree from concatenated maxicircle sequences showing the respective TcI populations structure across the Americas.

Geographic origin of individual isolates is shown on the map by small blue circles for NORTH-CENT and SOUTH clade whilst red circles represent TcI_{DOM} clade. Stipple large circle cluster each populations isolated from the same geographical area, NORTH-CENT (blue), SOUTH (gray), and TcI_{DOM} (Red). Branch colors indicate strain geographical source and color arrows the broad population. Branch labels are M.L. Bootstrap Support (BS) values and Bayesian Posterior Probabilities (BPP) respectively. The maxicircle topology is rooted with additional DTU strains (out-group). Comparison between mitochondrial and nuclear (figure 2) topologies revealed incongruence isolates across these phylogenies. Yellow circles represent sequences from previously characterized TcI strains: 9307 and 9354 that were divergent from all other TcI strains. These strains were most closely related to additional DTUs clade (out-group). Red circles represent sequences from Td3cl11 and NINOA with different phylogenetic positions between mitochondrial and nuclear (figure 2).

Nucleotide diversity and Allelic richness

Estimates of Nucleotide diversity (π) and Allelic richness (A_r) were calculated in DnaSP v.5 [Librado & Rozas, 2009] and FSTAT 2.9.3.2 [Goudet, 1995] respectively for the three main clades (NORTH-CENT, SOUTH and TcI_{DOM}) after correcting for sample size effects (Table 5 and Figure 3). The highest values for both estimates were observed in South America (π = 0.00712, Ar = 4.851) and lowest in TcI_{DOM} (π = 0.00071, Ar = 1.813).

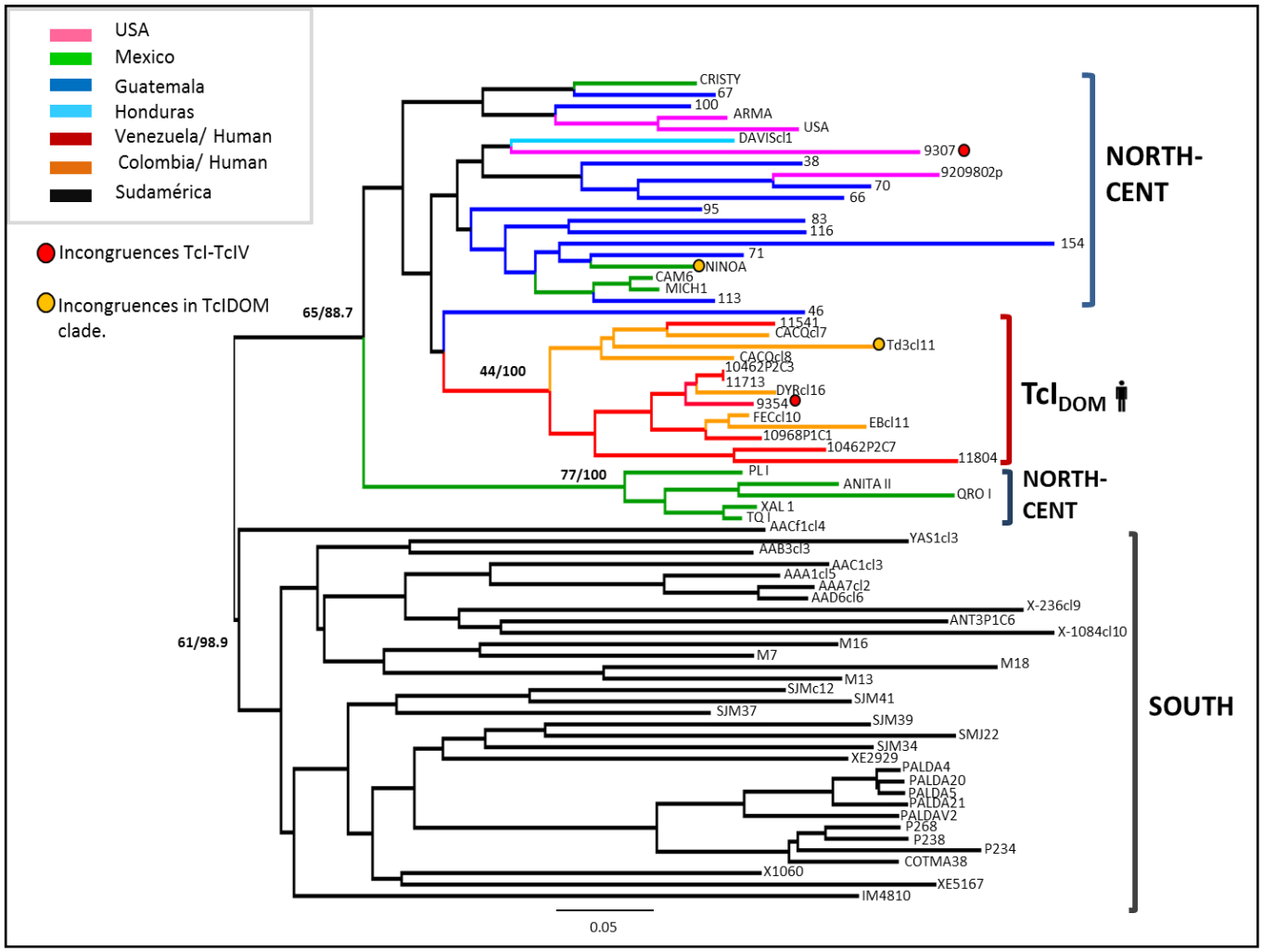

Figure 2. Unrooted Neighbour-Joining tree based on D_{AS} values from 19 microsatellite loci.

Branch colors indicate strain geographical origin and color arrows the broad clades. Branch labels are D_{AS}-based bootstrap values calculated over 1000 random diploid re-samplings to accommodate multi-allelic loci, these are shown >40%. A second values were calculated to test the topology across 10,000 bootstrap trees from 100 diploid datasets. These are shown on the major clades.

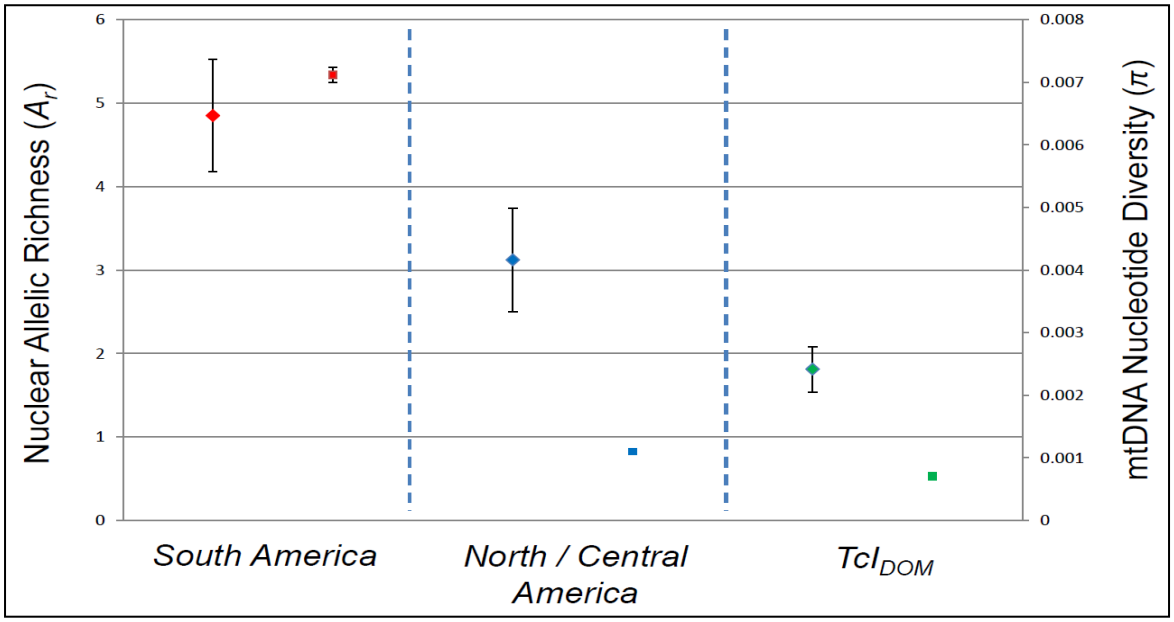

Figure 3. Nucleotide diversity and allelic richness comparisons across North and South American TcI populations.

Left hand data points (diamond) indicate allelic richness ± standard error over loci. Right hand data points (square) Indicate nucleotide diversity (π) ± standard error over pair-wise comparisons.

Table 5. Nucleotide diversity and Allelic richness.

Population	N	Nucleotide diversity (π) ±SE	Allelic richness (Ar) ±SE
NORTH-CENT	26	0.0011± 0.0000204	3.123 ± 0.62
SOUTH	33	0.00712 ± 0.00012	4.851 ± 0.67
TcIDOM	13	0.00071 ± 0.000049	1.813 ± 0.272

Population sample size corrected, Isolates from other DTUs (TcIII and TcIV) were not include in the analysis.
N = Number of isolates in population.
MNA = Allelic richness as a mean over loci ± standard error, calculated in FSTAT (Goudet, 1995).

AMOVA and Average pairwise differences.

Analyses of molecular variance (AMOVA) and Average pairwise differences were calculated in Arlequin v3.0 [Excoffier *et al.*, 2007] after getting the haplotype diversity in DnaSP v.5 [Librado & Rozas, 2009] for the three main clades. The AMOVA estimates did demonstrate significant structuring among and within the clades (FST = 0.218, P < 0.0000001, 10100 random permutations). Moreover, the average pairwise differences among clades were predominantly significant between SOUTH versus NORTH-CENT and TcI_{DOM}. Additionally, NORTH-CENT versus TcI_{DOM} did not show significant differences (Table S2 and S3).

Discussion

TcI dispersion into Central and North America

This study represents an attempt to document the within-DTU diversity in TcI isolates from Central, North and South America. By analyzing mitochondrial nucleotide sequences and nuclear microsatellite data we were able to resolve the presence of three clades corresponding with the previously classified populations: NORTH-CENT, SOUTH and TcI_{DOM}. According both phylogenies the South America isolates were the most genetically diverse (Figure 3). For other hand, the reduced genetic diversity among North-Central American isolates is powerful evidence in support of others who suggest that TcI originated in the south [Llewellyn *et al.*, 2009; Bernabe *et al.*, 2001]. Moreover, molecular clock analysis of *T. cruzi* major DTUs across 32 nuclear and mitochondrial loci, places the origin of TcI between 0.25-1.75 MYA [Flores-Lopez & Machado, 2011]. This date closely coincides with the formation of the Isthmus of Panama 2-4 MYA, suggest a divergence of the TcI clade before its dispersion from South America to Central and North America via the Great American Interchange and may correspond to the northerly migration of didelphid marsupials [Marshall & Sempere, 1993]. Meanwhile, it has been observed that TcI has unusual high parasitemias in opossums (*Didelphis spp*.) and has the capacity to complete its developmental cycle within this host, without the requirement of going through the digestive tract of an insect vector [Legey *et al.*, 2003; Carreira & Jansen, 2001]. This may predispose *D. marsupialis* to play an important role in *T. cruzi* transmission and dispersion. Additionally, TcI shows a close relation with insect vectors of the tribe Rhodniini which suggests that TcI could have an ancestral association with both of these taxa [Yeo *et al*, 2005; Gaunt & Miles, 2000]. Nevertheless, is unlikely that these taxa alone were responsible for the wide dispersal between geographic foci [Zingales *et al.*, 2012].

Origin of TcI_{DOM}

Recent findings by Llewellyn *et al.* (2009) and Messenger *et al.* (2012) reported a close resemblance between domestic TcI populations from the northern region of South America (VEN_{Dom}) and sylvatic populations from Central and North America by the use of nuclear and mitochondrial markers. Indeed, genotyping by analysis of the intergenic region of spliced-leader (SL-IR) suggests a distribution that now extends as far south as the Argentine Chaco where multiple sequences have been identified from human and domestic vector sources [Cura *et al.*, 2010]. Our study had the objective of unraveling the evolutionary origin of the domestic TcI isolates from Colombia and Venezuela (termed TcI_{DOM}).

Through mitochondrial and nuclear phylogenies we were able to place the TcI_{DOM} isolates as a well-defined clade nested within the NORTH-CENT. So that the explanation for this pattern observed among our data is that TcI_{DOM} strains are derived from Central and North America TcI strains. Furthermore, the robust genetic clustering of TcI_{DOM}; by comparison to the extensively sampled and genetically diverse parasite population from the sylvatic environment serves to make this domestic genotype representative and important. Moreover, estimation of allelic richness and nucleotide diversity across the three groups (SOUTH, NORTH-CENT and TcI_{DOM}) show TcI_{DOM} as the clade with smallest genetic diversity and allelic richness followed by the NORTH-CENT (Figure 3). These findings are bolstered by low genetic diversity identified among new mini-exon sequence data derived from North and Central American TcI isolates [O´Connor *et al.*, 2007]. This fits the pattern that might be expected during the colonization of a new host species / transmission cycle. Meanwhile, humans are considered to be relatively recent *T. cruzi* hosts. It is believed that the first contact between *T. cruzi* and humans is related to the entry of the first Americans to the continent [Pena *et al.*, 2009]. Furthermore, recent findings by Ramirez *et al.*, (2012) date the emergence of TcI_{DOM} around 23 000 ± 12 000 years before present. This date corresponds with the firsts migrations of humans in the Americas via the Bering Land Bridge around 30 000 – 13 000 years ago [Goebel *et al.*, 2008]. The human colonization of South America is thought to have occurred via the Isthmus of Panama rapidly along the Pacific coastline before lateral migration. Different reports of settlements in the late Pleistocene in South-America and estimations of human arrival and settlement in South-America fluctuate around 12,500 years ago [Goebel *et al.*, 2008]. Crucially, our data, which show that TcI_{DOM} is nested among North and Central American strains, suggest that this widespread domestic *T. cruzi* genotype actually made first contact with early colonizing Amerindians who were responsible for its southerly migration and dispersal from North / Central America. Nonetheless, static human population densities sufficient to support a sustained domestic cycle are presumably vital. However, such early settler populations were probably small, dynamic, and inherently unsuitable to sustain such a genotype. Many questions, therefore, remain around its emergence.

Nuclear-Mitochondrial Incongruence

Comparison of the mitochondrial and nuclear phylogenies revealed two individual isolates having unambiguously different phylogenetic positions: 9307 and 9354 (Figure 1 and 2). The maxicircle sequences from these TcI isolates were divergent from all other TcI strains. However, the mitochondrial phylogenetic tree was the most strongly supported, due to the presence of a larger amount of informative sites in comparison to the microsatellite tree. This could possibly be the case when mitochondrial introgression was reported to play a role in multiple genetic exchange events in *T. cruzi*, and could also explain the phylogenetic incongruences observed among some taxa [Messenger *et al.*, 2012].

References

1. Alarcón de Noya B, Díaz-Bello Z, Colmenares C, Ruiz-Guevara R, Mauriello L, Zavala-Jaspe R, Suarez JA *et al*: **Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela**. J Infect Dis 2010, **19**:1308-15.
2. Añez N, Crisante G, da Silva FM, Rojas A, Carrasco H, Umezawa ES, Stolf AM, Ramírez JL, Teixeira MM: **Predominance of lineage I among *Trypanosoma cruzi* isolates from Venezuelan patients with different clinical profiles of acute Chagas' disease.** Trop. Med Int Health 2004, **9**:1319-26
3. Araujo Carreira JC, Jansen AM, Deane MP, Lenzi HL: **Histopathological study of experimental and natural infections by *Trypanosoma cruzi* in *Didelphis marsupialis*.** Mem Inst Oswaldo Cruz 1996, **91**:609-618.
4. Barnabe C, Yaeger R, Pung O, Tibayrenc M: ***Trypanosoma cruzi:* a considerable phylogenetic divergence indicates that the agent of Chagas disease is indigenous to the native fauna of the United States**. *Exp Parasitol* 2001, **99**(2):73-79.
5. Bucio MI, Cabrera M, Segura EL, Zenteno E, Salazar-Schettino M: **Identification of immunodominant antigens in Mexican strains of *Trypanosoma cruzi.*** Immunol Invest 1999, **28**:257-68.
6. Carreira JC, Jansen AM, de Nazareth Meirelles M, Costa e Silva F, Lenzi HL: ***Trypanosoma cruzi* in the scent glands of *Didelphis marsupialis*: the kinetics of colonization.** Exp Parasitol 2001, **97**:129-40.
7. Cura CI, Mejía-Jaramillo AM, Duffy T, Burgos JM, Rodriguero M, Cardinal MV, Kjos S, *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**:1599-607.
8. Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): **an integrated software package for population genetics data analysis.** Evol Bioinform Online 2007, **1**:47-50.
9. Felsenstein J: **PHYLIP – Phylogeny Inference Package (Version 3.2).** Cladistics 1989, **5**:164–166.

10. Flores-López CA, Machado CA: **Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact.** PLoS Negl Trop Dis 2011, **5**:e1272.
11. Gaunt MW, Miles MA: **The ecotopes and evolution of triatomino bugs (Triatominae) and their associated trypanosomes.** Mem. Inst. Oswaldo Cruz 2000, **95**:557–565.
12. Goebel T, Waters MR, O'Rourke DH: **The late Pleistocene dispersal of modern humans in the Americas.** Science 2008, 319:1497-502.
13. Gonzalez CI, Ortiz S, Solari A: **Colombian *Trypanosoma cruzi* major genotypes circulating in patients: minicirle homologies by cross-hybridization analysis.** Int. J. Parasitol 2010, **40**:1685-92.
14. Goudet J: **FSTAT (Version 1.2): A Computer Program to Calculate FStatistics**. J Hered 1995, **86:**485–486.
15. Guhl F, Ramírez JD: ***Trypanosoma cruzi* I diversity: towards the need of genetic subdivision?** Acta Trop 2011, **119**:1-4.
16. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, *et al:* **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** Syst Biol 2010, **59**:307–321.
17. Hall TA: **Bioedit: a user-friendly biological sequence alignment edit and analysis program for Windows 95/98/NT.** Nucleic Acids Symp Ser 1999, **41**:95–98.
18. Hamilton PB, Adams ER, Njiokou F, Gibson WC, Cuny G, Herder S: **Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals.** Infect Genet Evol 2009, **9**:81-6.
19. Hamilton PB, Stevens JR, Gaunt MW, Gidley J, Gibson WC: **Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA.** Int J Parasitol 2004, **34**:1393-404.
20. Hamilton PB, Teixeira MM, Stevens JR. **The evolution of *Trypanosoma cruzi*: the 'bat seeding' hypothesis.** Trends Parasitol 2012, **28**:136-41.
21. Herrera C., Bargues M.D., Fajardo A., Montilla M., Triana O., Vallejo G.A *et al:* **Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia**. Infect. Genet. Evol 2007, **7**: 535–539.

22. Lauthier JJ, Tomasini N, Barnabé C, Rumi MM, D'Amato AM, Ragone PG, Yeo M. *et al:* **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 2012, **12**:350-8.
23. Legey AP, Pinho AP, Xavier SC, Marchevsky R, Carreira JC, Leon LL, Jansen AM: ***Trypanosoma cruzi* in marsupial didelphids (*Philander frenata* and *Didelhis marsupialis*): differences in the humoral immune response in natural and experimental infections.** Rev Soc Bras Med Trop 2003, **36**:241-8.
24. Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW, Miles MA: **Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids.** PLoS Negl Trop Dis 2011, **5**:e1363.
25. Librado P, Rozas J. DnaSP v5: **A software for comprehensive analysis of DNA polymorphism data.** Bioinformatics 2009, **25**:1451–1452.
26. Lima L, Silva FM, Neves L, Attias M, Takata CS, Campaner M, de Souza W, Hamilton PB, Teixeira MM: **Evolutionary Insights from Bat Trypanosomes: Morphological, Developmental and Phylogenetic Evidence of a New Species, *Trypanosoma* (Schizotrypanum) *erneyi sp*. nov., in African Bats Closely Related to *Trypanosoma* (Schizotrypanum) *cruzi* and Allied Species.** Protist 2012, in press.
27. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, Vargas J, Torrico F, Diosque P, Valente V, Valente SA, Gaunt MW: **Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection.** PLoS Pathog 2009, **5**:e1000410.
28. Marcili A, Lima L, Cavazzana M, Junqueira AC, Veludo HH, Maia Da Silva F, Campaner M, *et al*: **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.** Parasitology 2009, **136**:641-55.
29. Marshall L, Sempere T: **Evolution of the Neotropical Cenozoic land mammal fauna in its geochronologic, stratigraphic, and tectonic context.** In: Goldblatt P, ed. Biological relationships between Africa and South America. New Haven (Connecticut): Yale University Press; 1993:329–392.

30. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzén O, Lewis MD, Ramírez JD, Carrasco HJ, Andersson B, Miles MA: **Multiple Mitochondrial Introgression Events and Heteroplasmy in *Trypanosoma cruzi* Revealed by Maxicircle MLST and Next Generation Sequencing.** PLoS Negl Trop Dis 2012, **6**:e1584.
31. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, Fitzpatrick S, Gaunt MW, Mauricio IL: **The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on Leishmania: looking back and to the future.** Parasitology 2009, **136**:1509-28.
32. Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza L: **MICROSAT v1.5d: A computer programme for calculating various statistics on microsatellite allele data.** Stanford, CA: Department of Genetics, Stanford University 1999.
33. O'Connor O, Bosseno MF, Barnabé C, Douzery EJ, Brenière SF: **Genetic clustering of *Trypanosoma cruzi* I lineage evidenced by intergenic miniexon gene sequencing.** Infect Genet Evol 2007, **7**:587-93.
34. Pena SD, Machado CR, Macedo AM: ***Trypanosoma cruzi*: ancestral genomes and population structure.** Mem Inst Oswaldo Cruz 2009, **104**:108-14.
35. Pennington PM, Paiz C, Grajeda LM, Cordón-Rosales C. Short report: **concurrent detection of *Trypanosoma cruzi* lineages I and II in domestic *Triatoma dimidiata* from Guatemala.** Am J Trop Med Hyg. 2009; 80(2):239-41.
36. Pinto CM, Kalko KV, Cottontail F, Wellinghausen N, Cottontail V: **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 2012. In press.
37. Posada D: **jModelTest: phylogenetic model averaging.** Mol Biol Evol 2008, **25**: 1253–1256.
38. Ramirez JD, Guhl F, Messenger LA, Montilla M, Cucunuba Z, Lewis MD, Miles MA, Llewellyn MS: **Contemporary cryptic sexuality in *Trypanosoma cruzi*.** Mol Ecol 2012, Jul 9. doi: 10.1111/j.1365-294X.2012.05699.x. [Epub ahead of print].
39. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** Bioinformatics 2003; **19**:1572–1574.
40. Stevens JR, Gibson W: **The molecular evolution of trypanosomes.** Parasitol Today 1999b, **15**:432-7.

41. Stevens JR, Noyes HA, Dover GA, Gibson WC: **The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*.** Parasitology 1999a, **118**:107-16.
42. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** Nucleic Acids Res 1997, **25**:4876–4882.
43. World Health Organization: **"Chagas disease (American trypanosomiasis)"** Fact sheet N°340, June 2010 [http://www.who.int/mediacentre/factsheets/fs340/.
44. Yeo M, Acosta N, Llewellyn M, Sánchez H, Adamson S, Miles GA, López E, González N, Patterson JS, Gaunt MW, de Arias AR, Miles MA: **Origins of Chagas disease: Didelphis species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids.** Int J Parasitol 2005, **35**:225-33.
45. Zingales B, Andrade SG, Briones MR, Campbell DA, Chiari E, Fernandes O, Guhl F, Lages-Silva E, Macedo AM, Machado CR, Miles MA, Romanha AJ, Sturm NR, Tibayrenc M, Schijman AG: **Second Satellite Meeting. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI.** Mem Inst Oswaldo Cruz 2009, **104**:1051-4.
46. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MM, Schijman AG *et al:* **The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications.** Infect Genet Evol 2012, **12**:240-53.

Supplementary files

Table S1. Additional DTUs.

Strain code	Strain	Host/vector	DTU	Country	State	Reference
92122	92122102R	*Procyon lotor*	TcIV	USA	Georgia	Messenger *et al.*, 2012
CANIII	CANII cl1	*Homo sapiens*	TcIV	Brasil	Belem	Messenger *et al.*, 2012
CM17	CM17	*Dasypus spp.*	TcIII	Colombia	Carimaga	Messenger *et al.*, 2012
X1060	X10610 cl5	*Homo sapiens*	TcIV	Venezuela	Guárico	Messenger *et al.*, 2012
ERA	ERA cl2	*Homo sapiens*	TcIV	Venezuela	Anzoátegui	Messenger *et al.*, 2012
10R26	10R26	*Aotus spp.*	TcIV	Bolivia	Santa Cruz	Messenger *et al.*, 2012
SAIRI3	Saimiri3 cl1	*Saimiri sciureus*	TcIV	Venezuela	Venezuela	Messenger *et al.*, 2012

Table S2. Analysis of Molecular Variance.

AMOVA					Populations specific FST indices		
Source of variation	**d.f**	**Sum of squares**	**Variance components**	**Percent of variation**	**Pop**	**Name**	**FST**
Among populations	2	208.759	4.00819 va	21.82	1	SOUTH	0.22015
Within populations	69	991.018	14.36258 vb	78.18	2	NORTH-CENT	0.22302
Total	71	1199.778	18.37077		3	TcIDOM	0.20371

Fixation Index (FST): 0.21818

Table S3. Population average pairwise differences.

Poputations	SOUTH	NORTH-CENT	TcIDOM
SOUTH			
NORTH-CENT	9.4****		
TcIDOM	10.28****	0.43 °	

**= 0.000001 (NORTH-CENT vs. SOUTH)
*= 0.02 (TcI_{DOM} vs. SOUTH)
°= 0.11 (TcI_{DOM} vs. NORTH-CENT**)**